

Article

Short-Term PV Power Forecasting Using a Regression-Based Ensemble Method

Andi A. H. Lateko ^{1,2}, Hong-Tzer Yang ¹ and Chao-Ming Huang ^{3,*}

¹ Department of Electrical Engineering, National Cheng Kung University, Tainan 701, Taiwan; halik@unismuh.ac.id (A.A.H.L.); htyang@mail.ncku.edu.tw (H.-T.Y.)

² Department of Electrical Engineering, Muhammadiyah University of Makassar, Makassar 90221, Indonesia

³ Department of Electrical Engineering, Kun Shan University, Tainan 710, Taiwan

* Correspondence: cmhuang@mail.ksu.edu.tw; Tel.: +886-6-205-0518

Abstract: One of the most critical aspects of integrating renewable energy sources into the smart grid is photovoltaic (PV) power generation forecasting. This ensemble forecasting technique combines several forecasting models to increase the forecasting accuracy of the individual models. This study proposes a regression-based ensemble method for day-ahead PV power forecasting. The general framework consists of three steps: model training, creating the optimal set of weights, and testing the model. In step 1, a Random forest (RF) with different parameters is used for a single forecasting method. Five RF models (RF₁, RF₂, RF₃, RF₄, and RF₅) and a support vector machine (SVM) for classification are established. The hyperparameters for the regression-based method involve learners (linear regression (LR) or support vector regression (SVR)), regularization (least absolute shrinkage and selection operator (LASSO) or Ridge), and a penalty coefficient for regularization (λ). Bayesian optimization is performed to find the optimal value of these three hyperparameters based on the minimum function. The optimal set of weights is obtained in step 2 and each set of weights contains five weight coefficients and a bias. In the final step, the weather forecasting data for the target day is used as input for the five RF models and the average daily weather forecasting data is also used as input for the SVM classification model. The SVM output selects the weather conditions, and the corresponding set of weight coefficients from step 2 is combined with the output from each RF model to obtain the final forecasting results. The stacking recurrent neural network (RNN) is used as a benchmark ensemble method for comparison. Historical PV power data for a PV site in Zhangbin Industrial Area, Taiwan, with a 2000 kWp capacity is used to test the methodology. The results for the single best RF model, the stacking RNN, and the proposed method are compared in terms of the mean relative error (MRE), the mean absolute error (MAE), and the coefficient of determination (R^2) to verify the proposed method. The results for the MRE show that the proposed method outperforms the best RF method by 20% and the benchmark method by 2%.

Keywords: PV power forecasting; ensemble method; Random forest; linear regression; support vector machine; clustering method



Citation: Lateko, A.A.H.; Yang, H.-T.; Huang, C.-M. Short-Term PV Power Forecasting Using a Regression-Based Ensemble Method. *Energies* **2022**, *15*, 4171. <https://doi.org/10.3390/en15114171>

Academic Editor: Frede Blaabjerg

Received: 3 May 2022

Accepted: 2 June 2022

Published: 6 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Forecasting photovoltaic (PV) power generation is a vital element in the planning and operation of an electric power grid. Renewable energy resources are rapidly integrated into smart grids [1–3]. The variability and uncertainty of PV power output and availability must be considered in the complex decision-making processes required to balance supply and demand for the power system. A solar generator at the ground level is affected by cloud cover, atmospheric aerosol levels, and other atmospheric parameters, so solar power is intermittent and variable [4]. Meteorological features, such as solar irradiance, air temperature, relative humidity, and wind speed, directly or indirectly affect the power generated by a PV [5]. The intermittent nature of power generation from solar PV systems means

that maximizing power output and connecting to the utility grid is difficult. Forecasting is critical to the efficient use of solar power for grid operations.

PV power forecasting involves very short-term, short-term, medium-term, or long-term forecasting horizons. Very short-term forecasting has prediction periods from 1 min to several hours and is useful for the electricity market, power smoothing, and real-time electricity dispatch. Short-term forecasting is widely used in the electricity market to ensure economic load dispatch, and the time horizons range from one day to a week. Medium-term forecasting ranges from one week to a month and is used for maintenance planning. Long-term forecasting forecasts for a month to a year ahead and is used to determine plans for long-term power generation, transmission, distribution, and solar energy rationing [6].

A previous study [7] used a univariate data-driven approach to increase the accuracy of very short-term solar power forecasting. The forecasting horizon is 15 min ahead, and the real solar power dataset is the only input for the model. The performance indices are MAE, mean relative error (MRE), and RMSE. Another study [8] performed medium- and long-term PV power forecasting using LSTM. The MAE and RMSE are used as evaluation metrics.

Current methods to forecast PV power generation are categorized as physical, statistical, or machine-learning, as well as hybrid methods that integrate two or more methods [9,10]. A physical approach generates PV forecasts using solar and PV models and a statistical approach uses past data to train models. A physical model uses satellite images and numerical weather predictions (NWP) to predict PV power generation [11,12].

A previous study [13] used a statistical approach to forecast photovoltaic power generation using autoregressive moving average (ARMA) models. These models are simple and give good forecasting results for one-step-ahead predictions using a resolution of one hour, but forecast errors increase proportionally with forecast times. Support vector machines (SVM), artificial neural networks (ANN), deep neural networks (DNN), random forest (RF), and metaheuristic methods are used in machine learning [14–18]. A previous study [19] used the Random forest to forecast solar power using principal component analysis (PCA)-K-means clustering together with the differential evolution grey-wolf algorithm. However, the calculation time increases if the algorithm is used to optimize parameters because the number of iterative operations increases.

The most commonly used machine-learning technique is deep learning. Deep learning uses neural networks with more than three layers. A previous study [20] compared various deep learning neural networks for short-term PV power forecasting: long short-term memory (LSTM), bidirectional LSTM (Bi-LSTM), a gated recurrent unit (GRU), bidirectional GRU (Bi-GRU), a convolutional neural network (CNN), and other hybrid configurations such as CNN-LSTM and CNN-GRU. Another study [21] used a hybrid implementation of physical models and an ANN. Hybrid models give accurate forecasts for photovoltaic production but there are significant forecasting errors due to inaccurate weather forecasts.

A metaheuristic method is used to optimize the hyperparameters for a forecasting model. One study [22] used a metaheuristic method called differential evolution and a particle swarm optimization (DEPSO) algorithm to optimize the forecasting model for short-term PV power output forecasting. Another study [23] used a metaheuristic method called a CNN-salp swarm algorithm (SSA) with a deep learning method. To allow predictions for different weather types, five CNN regression models were created and the hyperparameters were optimized using a salp swarm algorithm (SSA). Another study [24] used an LSTM with four hidden layers and Bayesian optimization to select the best combination of features. The simulation gives an accurate forecast in sunny and cloudy weather in terms of the mean squared error (MSE), the mean absolute error (MAE), the coefficient of determination (R^2), and the root mean squared error (RMSE).

Clustering and classification improve forecasting accuracy. One study [25] used K-means clustering to define the different types of sky for each hour using different levels of irradiance and weather features, such as solar irradiance, temperature, wind speed, and relative humidity, as inputs. This produces a 33–44.6% improvement in accuracy compared

to the benchmark method. Another study [26] compared classification methods, such as K-nearest neighbor (KNN) and SVM models, in terms of performance. The results show that an SVM performs well on a small sample scale.

An ensemble learning method is used to increase the accuracy of PV power forecasting, which involves combining multiple models to make predictions. One study [27] developed a stacked generalization ensemble model for short-term PV power generation forecasting. This uses base learners such as extreme learning machines (ELM), extremely randomized trees, K-nearest neighbor (KNN), and the Mondrian forest model. A deep belief network is used as a meta learner to generate the final outputs from meteorological features such as global horizontal irradiance (GHI), diffuse horizontal irradiance (DHI), relative humidity, wind direction, and temperature. The MAE, RMSE, mean absolute percentage error (MAPE), and R^2 values are used as evaluation criteria. The proposed model gives a MAPE that is 2.30% more than that for the benchmark and a single model.

One study [28] used a seasonal time series model to develop a regression-based ensemble forecasting combination. Seasonal time series models use a seasonal autoregressive integrated moving average (SARIMA), exponential smoothing (ETS), multilayer perceptron (MLP), seasonal trend decomposition, a TBATS model, and a theta model. Eight ensemble forecasting combination methods were used to combine the forecasting results. The normalized root mean squared error (nRMSE), normalized mean bias error (nMBE), forecast skill, and Kolmogorov–Smirnov test integral (KSI) are used to calculate the accuracy. Sometimes the best individual model is more accurate than the ensemble model.

Another study [29] used the bagging ensemble method with Random forest (RF) and extra trees (ET) to predict hourly PV generation, and an SVR was used as the benchmark model. The inputs for the model are solar radiation, air temperature, relative humidity, wind speed, and the previous hourly value for PV output. The RMSE and MAE values are used for error validation. ET outperforms RF and SVR, with an MAE of 1.0851 kWh. However, this study did not involve different weather conditions, such as sunny, cloudy, or rainy.

One study [30] blended forecasting results from multiple feedforward neural network (FNN) predictors using the RF model. Meteorological measurements, such as solar irradiance, ambient temperature, and wind speed, were used as model inputs. The method for this study outperforms six benchmark models in terms of persistence, SVR, linear regression (LR), RF, gradient boosting (GB), and extreme GB (XGBoost) by 40%, but the method only uses one-hour-ahead forecasts for very short-term PV power forecasting.

Table 1 shows the previous researches on PV power forecasting using the ensemble method. Many studies show that the use of the ensemble method can increase the accuracy of the single forecasting method. In fact, the weight coefficients in every weather condition, such as sunny, cloudy, or rainy, are different from one another. The suitable weights must be implemented in the proper weather conditions to increase the accuracy. This study proposed an ensemble-based model for short-term PV power forecasting to increase the accuracy of the short-term PV power output predictions. The proposed model incorporates five RF models for five weather types: sunny, light-cloudy, cloudy, heavy-cloudy, and rainy. It also uses regression-based methods such as linear regression (LR) and support vector regression (SVR) and uses LASSO and Ridge regularization for weighting to combine the forecasting results. A previous study implemented a stacked generalization ensemble method for short-term PV power forecasting using an RNN meta learner [31]. The stacking RNN method is used as a benchmark for the ensemble forecasting method for this study. The goal of this study is to improve the accuracy and performance of the individual forecasting models for day-ahead PV power forecasting by implementing a regression-based ensemble method. This study makes the following significant contributions to this field of study:

- A new PV forecasting structure that incorporates K-means clustering, RF models, and the regression-based method with LASSO and Ridge regularizations is used to increase forecasting accuracy.

- A regression-based ensemble learning with Bayesian optimization is used with LASSO and Ridge regularization to calculate the five optimal sets of weight coefficients, which allows us to determine which predictors in the model are significant.
- The regression-based method is easier to implement and has fewer hyperparameters compared to the stacking RNN method. The results show that the proposed regression-based method outperforms the benchmark stacking RNN by 2%.

Table 1. The previous study of ensemble PV power forecasting.

Single Methods	Ensemble Method	Ref.	Error Validation	Forecasting Horizon	Resolution	Best Result
FNNs	RF	[30]	nRMSE, nMAE	1-h	1-h	nMAE = 2.42%
ANN, DNN, SVR, LSTM, CNN	RNN	[31]	MRE, MAE, nRMSE, R ²	1-day	1-h	MRE = 4.29%
ARIMA, VAR, LSTM	CNN	[32]	MAE, MSE, RMSE	1-year	1-month	MAE = 16.70 MWh
GRU, XGBoost, MLP	Simple averaging	[33]	RMSE, MAE, MAPE	1-day	1-h	MAPE = 1.60%
RFs	Weighted averaging	[34]	nMBE, nMAE, nRMSE, forecast skill	1-day	1-h	nMAE = 4.06%
SVM, MLP, MARS	Weighted averaging	[35]	RMSE, MAE, MAPE	1-day	5-min	MAPE = 0.78%

The remainder of the paper is structured as follows. Section 2 briefly describes the proposed methodology and setup modeling. Section 3 explains the ensemble forecasting strategy. Section 4 details the proposed PV power forecasting simulation results, and Section 5 details the conclusions and future applications.

2. Modelling and Methodologies

2.1. The K-Means Model

The K-means clustering method is used for this study to divide the training set into clusters. K-Means clustering is a type of unsupervised machine-learning technique that is frequently used to divide a set of data into several subgroups. K-means is a traditional clustering method that is simple, fast, and robust [36] and produces groups that have similar characteristics that are significantly different from those other groups.

The K-Means clustering minimizes the sum of squared errors (SSE) as in [37]:

$$SSE = \sum_{j=1}^k \sum_{i=1}^n \|x_i - c_j\|^2 \quad (1)$$

where k represents the number of clusters, n represents the number of observations, x_i represents the i th observation, and c_j represents the centroid for cluster j .

To iteratively update the centroid of each cluster, Equation (2) is used:

$$c_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \quad (2)$$

where $|C_j|$ represents the total number of points in cluster j .

The following are the steps of the K-means clustering:

1. The initial centers of each group, K samples, are chosen at random to eliminate the dimensional effects. Each feature is normalized using the min-max method.
2. Samples are assigned to groups based on their Euclidean distance from the center of the group, and the group with the smallest Euclidean distance is chosen for each sample.

3. The centers of each group are recalculated using the sample data for each group, and the results are output if none of the centers are changed.
4. Steps 2 and 3 are repeated until convergence is achieved.

The elbow method is a common method for determining the optimal number of clusters. This method uses the concept of the within-cluster sum of squares (WCSS) value [38]. The total variance within a cluster is defined using the WCSS.

The elbow method uses the following steps to determine the optimal value of clusters:

1. K-means clustering is performed on a given dataset for various K values.
2. The WCSS value is calculated for each value of K .
3. A line is drawn between the calculated WCSS values and the number of clusters K .
4. When the point on the plot looks like an arm, it has the best value for K .

The elbow method is defined as:

$$WCSS = \sum_K \left(\sum_{X_i \in K} (Dist(X_i, C_K)) \right)^2 \quad (3)$$

where K denotes the number of clusters, X_i denotes the number of observations, and C_K denotes the cluster center.

2.2. The Random Forest Model

Random forest is a machine-learning approach that is used for classification and regression problems. A Random forest is an ensemble of decision trees. The Random forest output is the class that is chosen by the majority of trees that are used for classification problems, but for regression problems the mean or average prediction for an individual tree is used.

Figure 1 shows the structure of the Random forest model. Random forest uses the ensemble technique of bagging, which is also known as bootstrap aggregation. Bagging selects a sample at random from the original dataset, so rows are sampled to construct each model using the bootstrap samples from the original data. The bootstrap method is used for row sampling with replacement. The results are generated using each model, which is independently trained. A majority vote or mean decision is made when all models are combined. Aggregation involves combining all of the results and generating output. The RF model is robust to missing values and outliers and is less affected by noise. A detailed model of RF is shown in [39].

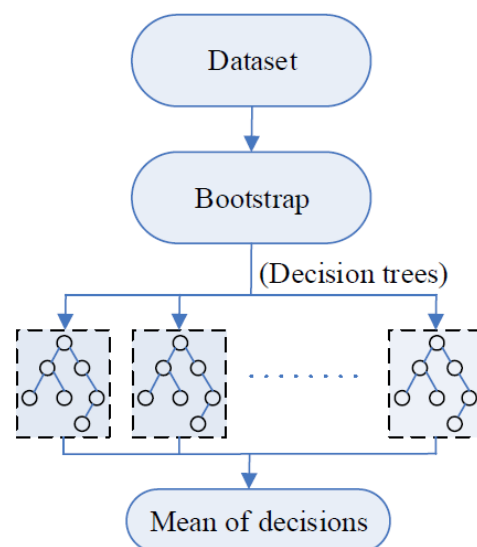


Figure 1. The Random forest models [40].

2.3. The Stacking RNN Ensemble Method

In this study, the stacking RNN was used as the benchmark method. Stacking RNN is the ensemble method based on stacked generalization by training the first-level learners and combining them using the second-level learner to obtain the final forecasting results. A more detailed explanation of the stacking RNN ensemble method can be found in [31].

2.4. The Ensemble Combination Strategy

2.4.1. The Linear Regression (LR) Model

This study uses a linear regression model [41] with ordinary least squares (OLS) to combine the forecasting results of five different RF forecasting models:

$$y = f(x) = w \cdot x_i + b \quad (4)$$

where x_i denotes the forecast result for model i , w denotes the weight coefficient, and b denotes the intercept or bias.

The best fit is determined by minimizing the sum of squared errors:

$$\min \sum_{i=1}^m (y_i - \hat{y}_i)^2 = \sum_{i=1}^m (y_i - (w \cdot x_i + b))^2 \quad (5)$$

The solution involves solving:

$$\hat{w} = (X^T X)^{-1} X^T Y \quad (6)$$

To avoid overfitting, a regularization term is used (to minimize the magnitude of w):

- LASSO regression:

LASSO stands for least absolute shrinkage and selection operator. LASSO regression performs L1 regularization by adding a penalty coefficient λ equal to the absolute value of the magnitude of the coefficients.

$$\min \sum_{i=1}^m (y_i - w \cdot x_i - b)^2 + \lambda \sum_{j=1}^n |w_j| \quad (7)$$

- Ridge regression:

Ridge regression performs L2 regularization by adding a penalty coefficient λ equal to the square of the magnitude of the coefficients.

$$\min \sum_{i=1}^m (y_i - w \cdot x_i - b)^2 + \lambda \sum_{j=1}^n |w_j^2| \quad (8)$$

where λ represents the penalty coefficient.

2.4.2. The Support Vector Regression (SVR) Model

Support vector regression (SVR) is used in this study to combine the forecasting results for five different RF models. Figure 2 shows the structure of the SVR.

The function of SVR is:

$$y = f(x_i) = w^T \varphi(x_i) + b \quad (9)$$

where $f(x)$ represents the forecast values, $\varphi(x)$ represents the kernel function (RBF function as a kernel function) for the inputs, and w and b represent the weighted coefficient and the bias, respectively.

A penalty function is used to calculate the values of coefficients w and b :

$$R(C) = \frac{1}{2} \|w\|^2 + C \cdot \frac{1}{n} \sum_{i=1}^n |y_i - f(x)|_\varepsilon \quad (10)$$

$$|y - f(x)|_\varepsilon = \begin{cases} 0, & |y - f(x)| \leq \varepsilon, \\ |y - f(x)| - \varepsilon, & \text{otherwise} \end{cases} \quad (11)$$

where $\|w\|^2$ represents the regularization term, C represents the penalty coefficient, and ε represents the maximum value for the tolerable error.

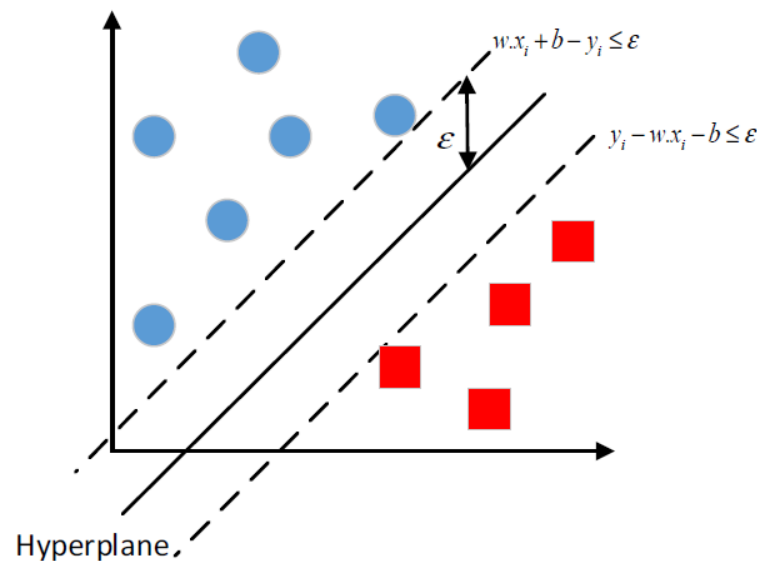


Figure 2. A support vector regression model [42].

2.4.3. Bayesian Optimization

In machine learning, hyperparameters need to be tuned to ensure the performance of the prediction model. The best results can be obtained by using the optimal hyperparameters. Hyperparameter optimization is used to optimize the model. Bayesian optimization is one of the global optimization algorithms that generates a probabilistic model of the function mapping from hyperparameter values to the target, which is then tested on a validation set. A detailed description of the Bayesian optimization algorithm can be found in [43].

2.5. Setup Modelling

2.5.1. Data Preprocessing

Data preprocessing involves data normalization, cleaning, repair, and data splitting. During the data preparation stage, data are normalized using min-max normalization. The min-max normalization is defined as:

$$\bar{x}_n = \frac{x_n - x_{\min}}{x_{\max} - x_{\min}} \quad (12)$$

where \bar{x}_n is the normalized data, x_n is the original data, and x_{\max} and x_{\min} are the maximum and minimum values of x_n .

After data normalization, data cleaning removes outliers and data repair replaces missing values using linear interpolation. The good data are then divided into data training and testing sets.

2.5.2. Datasets

The PV site for this study is located at Zhangbin Industrial Area in Taiwan, at a latitude of 24.12809° and longitude of 120.4281°. Zhangbin Industrial Area's PV site has a ground-mounted panel with a 2000 kWp capacity. The PV power output data for 2020 is used for this study.

Two types of datasets are used for this study: meteorological data that is obtained from Solcast and measurement data from the PV site. The meteorological data from Solcast is open access data that contains the real values for irradiance and weather with a 10 min resolution [44]. The meteorological data features for this study are solar irradiance (GHI), air temperature, precipitation, relative humidity, and wind speed. This data is averaged to a one-hour resolution to meet the requirements of this study. The measured real data for PV power output from the PV site's ground panel in Zhangbin Industrial Area, Taiwan, is also used as a data feature.

The Pearson correlation coefficient (PCC) and *t*-statistics are used to select appropriate data features. PCC is used to calculate the correlation value between each weather variable and the PV power output. The values are between -1 and 1 [45]. A value of $r = 1$ indicates a positive correlation, $r = 0$ indicates no correlation, and $r = -1$ indicates a negative correlation. The formula for PCC is:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (13)$$

where r is the Pearson correlation coefficient (PCC), $\bar{x} = \frac{1}{n} \sum_{i=1}^N x_i$ represents the mean of x , and $\bar{y} = \frac{1}{n} \sum_{i=1}^N y_i$ represents the mean of y .

The PCC, *t*-test, and *p*-value between weather features and PV power output are shown in Table 2. Precipitation and wind speed have a low correlation with PV power output, but solar irradiance, air temperature, and relative humidity have a high correlation. Even though precipitation and wind speed have a low correlation, the *t*-test results show that input variables with *p*-values less than 0.05 are still significant and can be used as input variables [24]. A previous study [5] also demonstrated that precipitation and wind speed indirectly affect PV power output. Therefore, solar irradiance (GHI), air temperature, precipitation, relative humidity, and wind speed are the weather variables used for this study.

Table 2. The statistical test between weather features and PV power output.

Weather Variables	Correlation Coefficient	<i>t</i> -Test	<i>p</i> -Values
Irradiance (W/m ²)	0.970	56.499	0
Temperature (°C)	0.364	−2.119	0.034
Precipitation (kg/m ²)	−0.012	2.189	0.029
Humidity (%)	−0.521	−2.451	0.014
Wind speed (m/s)	−0.056	7.975	2.038×10^{-15}

2.5.3. Evaluation Criteria

The mean relative error (MRE), the mean absolute error (MAE), and the coefficient of determination (R^2) are used as evaluation criteria to validate the error. The MRE is calculated by dividing the actual and forecasted values by the nominal capacity of the photovoltaic facility [46]. MAE represents the accuracy of the prediction [47]. R^2 is the coefficient of determination, which ranges from 0 to 1 [48]. The higher the value of R^2 , the more accurate the model. The formulas for MRE, MAE, and R^2 are:

$$\text{MRE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{N_p} \right| \times 100\% \quad (14)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \tag{15}$$

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \tag{16}$$

where y_i and \hat{y}_i are the forecast value and the true value for PV power output at the i th point, respectively, N is the number of prediction points, N_p is the PV site’s nominal power capacity, and \bar{y}_i is the average PV power output.

3. Ensemble Forecasting Strategy

The clustering method, classification techniques, RF models, and the regression-based ensemble model are used for the proposed PV power ensemble forecasting strategy. Figure 3 shows the overall structure of the ensemble PV power generation forecast. The general framework consists of three steps: model training, optimal set of weights creation, and model testing. In step 1, K-means clustering uses the daily average historical PV power output for k different weather conditions using the optimal number of k . The optimal number of k is calculated using the elbow method, which is five in this case. The five clusters are labeled as rainy, heavy-cloudy, cloudy, light-cloudy, and sunny. Then an RF model was trained on each cluster using the historical hourly weather data as input and PV power generation as output. In addition, an SVM classification model was also trained using the historical daily average weather data as input and the label defined by K-means clustering as output. There were five RF models (RF₁, RF₂, RF₃, RF₄, and RF₅) and an SVM classification model obtained in this step. Figure 4 shows the detailed process of step 1.

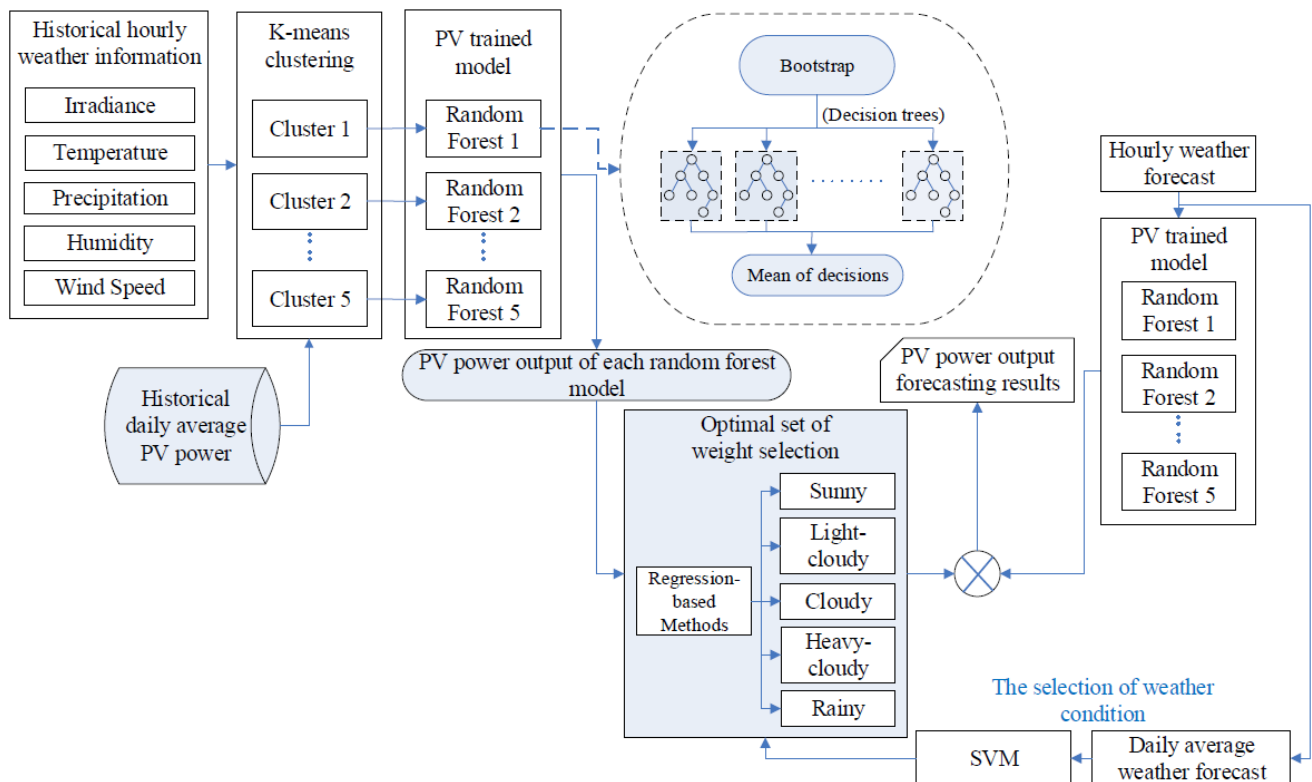


Figure 3. The general framework of a one-day ahead ensemble PV power forecasting strategy.

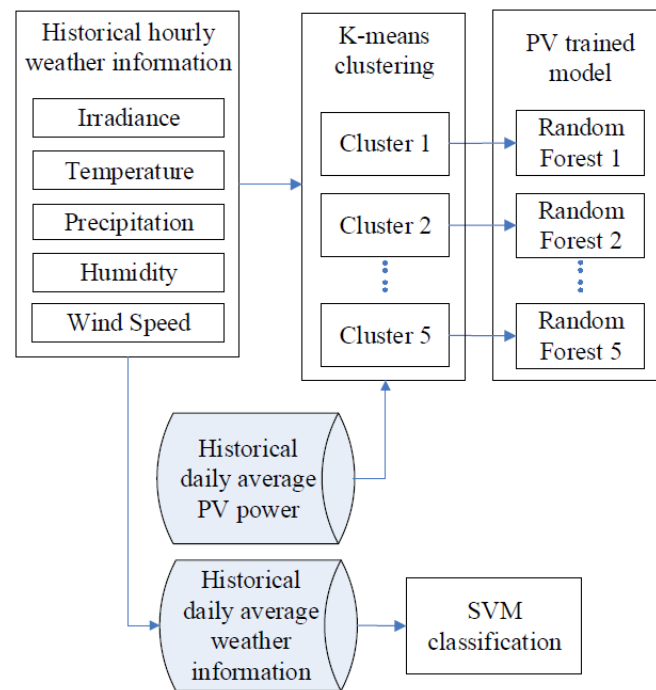


Figure 4. The detailed process of step 1.

In step 2, the dataset for each cluster was trained using RF models that we obtained from the first step. Then, the PV power output of each Random forest model is used as an input to the regression-based method to construct the set of weights. The hyperparameters for the regression-based method are learner (LR or SVR), regularization (LASSO or Ridge), and λ (penalty coefficient for regularization). Bayesian optimization is performed to find the optimal value of these three hyperparameters. The optimal set of weights is obtained in this step and each set of weights contains five weight coefficients and a bias. Different weather conditions have different sets of weights to ensure accurate forecasting results. Figure 5 shows the detailed process of step 2.

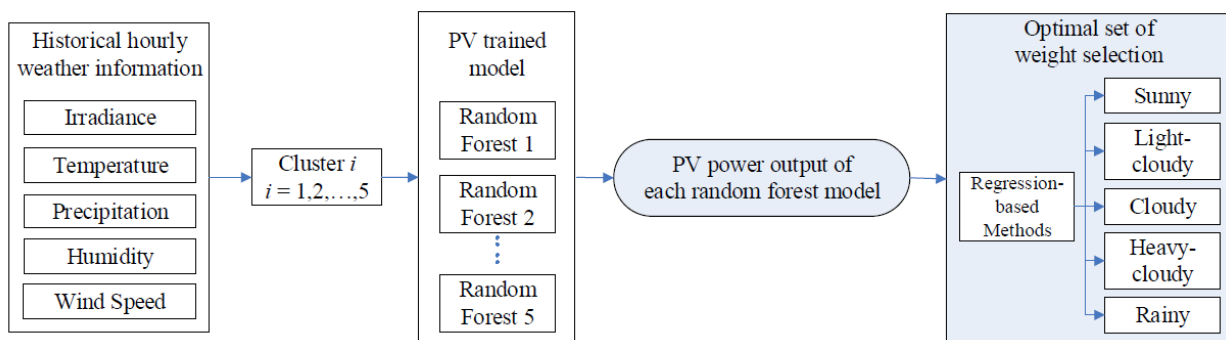


Figure 5. The detailed process of step 2.

In order to simulate weather forecasting inaccuracies, random errors of 10%, 20%, 30%, 40%, 50%, and 60% are applied to the actual weather value in load forecasting [49]. In step 3, we assumed that a random error of $\pm 20\%$ is added to the actual weather data to allow it to be used as forecasting data due to insufficient weather forecasting data. The weather forecasting data for the target day is used as input for five RF models. The average daily weather forecasting data is also used as input for the SVM classification model obtained in step 1. The SVM output selects the weather conditions, and the corresponding set of

weight coefficients from step 2 is combined with the output from each RF model to obtain the final forecasting results by using (17):

$$\hat{Y} = w_1\hat{y}_1 + w_2\hat{y}_2 + \dots + w_5\hat{y}_5 + b \quad (17)$$

where \hat{Y} is the final forecasting results, w_1, w_2, \dots, w_5 are the weight coefficients, $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_5$ are the forecasting results of RF models, and b is the bias. Figure 6 shows the detailed process of step 3.

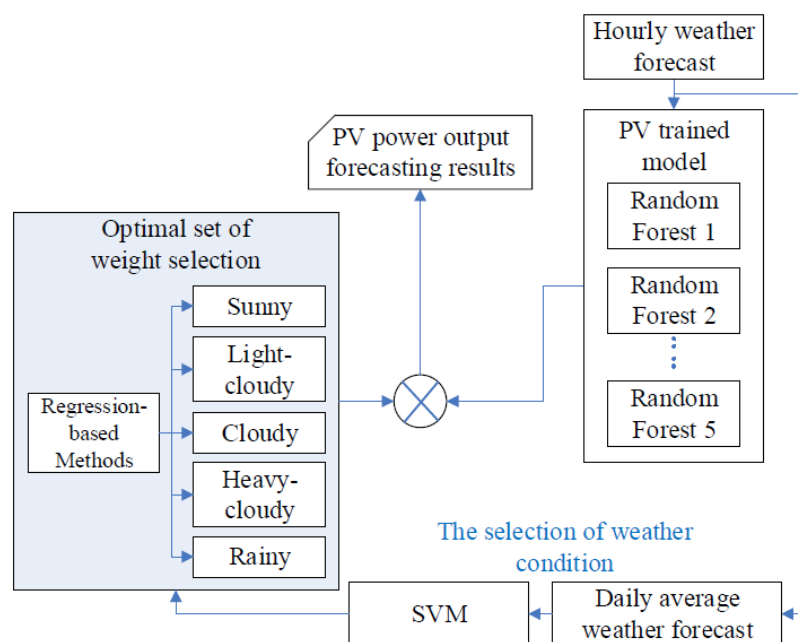


Figure 6. The detailed process of step 3.

4. PV Power Forecasting Simulation Results

The software package, MATLAB 2021b edition, with an Intel Core i7 CPU at 3.60 GHz and an 8 GB RAM computer, is used for the simulation. The stacking RNN ensemble method is used as a benchmark model to compare the results for one-day-ahead PV power forecasting. The stacking RNN ensemble method is proven to give accurate short-term PV power forecasts [31].

4.1. Test System

A case study used a 2000 kWp PV farm in Zhangbin Industrial Area, Taiwan, as a test system to determine the accuracy of PV power output forecasting. The actual irradiance and weather features from Solcast are used to train the model. There is a lack of weather prediction data, so a $\pm 20\%$ random error is generated in the actual data to simulate the weather forecast. The measured PV power generation for this study was obtained from the Zhangbin Industrial Area's PV site in Taiwan.

The test system includes historical data for PV power output and hourly average values for irradiance, temperature, precipitation, relative humidity, and wind speed. The data for 2020 is used as a dataset for the system. The data preprocessing that is described in Section 2.5.1 gives 300 days of good data that is used for the simulation. This study uses twelve points for each PV power output and corresponding weather variables on one day: the PV power output and weather variables from 06:00–17:00.

The collected data is classified into five weather conditions using K-means clustering, and the elbow method is used to determine the optimal number of clusters. The weather conditions are sunny, light-cloudy, cloudy, heavy-cloudy, and rainy.

The training and testing datasets for various weather conditions that are used to train and test the RF model and the ensemble method are shown in Table 3. From the 300-day dataset, 223 days (75%) are used to train a single RF model, and 77 days (25%) are used to test the model. The test results for the single RF model are used as datasets to train and test the ensemble learner. Sixty days (80%) are used to train the ensemble learner, and 17 days (20%) are used for testing. A total of 10 days of the ensemble learner's testing data are used to test the proposed method for each model, and the proposed method is tested using the ensemble learner's testing data for seven consecutive days from 14 May to 20 May 2020. Figure 7 shows a detailed illustration of the RF model and the data preparation for the ensemble learner.

Table 3. The number of days that are used for training and testing.

Weather Conditions	Random Forest Model		Ensemble Learner	
	Training	Testing	Training	Testing
Sunny	59	20	16	4
Light-cloudy	57	19	15	4
Cloudy	54	17	14	3
Heavy-cloudy	27	11	8	3
Rainy	26	10	7	3
Total	223	77	60	17

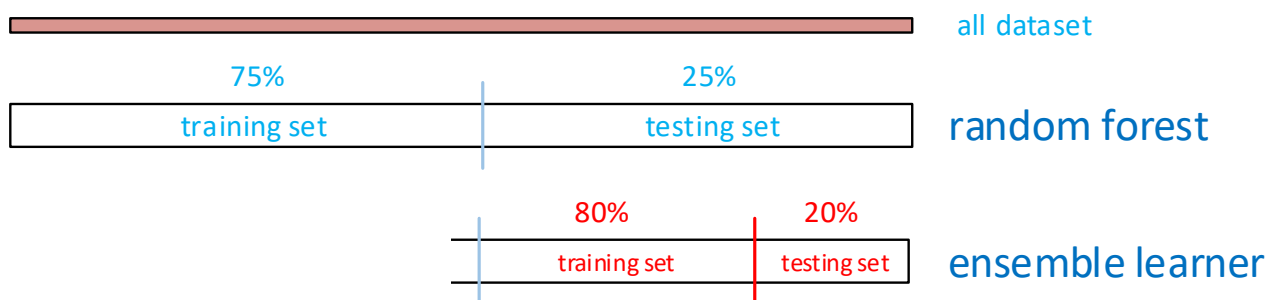


Figure 7. Details of the data preparation procedure.

4.2. Hyperparameters Setting for the RF and Ensemble Models

The hyperparameters for the RF model are the number of trees and the minimum leaf size. The search spaces are 100, 200, 500, and 1000 trees, with minimum leaf sizes of 1, 3, and 5. Table 4 shows the parameters for each single RF model, as determined by the experiment.

Table 4. Parameters for the single RF models.

Model	Parameters	Value
RF ₁	Number of trees	1000
	Min leaf size	3
RF ₂	Number of trees	1000
	Min leaf size	3
RF ₃	Number of trees	800
	Min leaf size	3
RF ₄	Number of trees	1000
	Min leaf size	3
RF ₅	Number of trees	200
	Min leaf size	3

The penalty coefficient (lambda), a learner, and regularization are used to tune the optimal hyperparameters for the proposed ensemble model. Lambda is a positive coefficient. The regression-based learners are linear regression using ordinary least square (OLS) and support vector regression (SVR), and the regularization methods are LASSO and Ridge regression.

The ensemble model uses Bayesian optimization to optimize the hyperparameters. “Bayesian optimization” is a global optimization problem [50]. The benchmark method is a stacking RNN, which has the same structure as that of a previous study [31]. Table 5 shows the parameters for the benchmark model and the optimal hyperparameters for the proposed ensemble model that are determined using the optimization process.

Table 5. Hyperparameters for the ensemble model.

Model	Parameters	Sunny	Light-Cloudy	Cloudy	Heavy-Cloudy	Rainy
Stacking RNN	Hidden layer	1	1	1	1	1
	Hidden neuron	5	7	7	5	4
	Input delay	2	2	2	2	2
	Learning rate	0.001	0.05	0.005	0.005	0.005
Proposed Method	Lambda (λ)	4.898×10^{-5}	5.145×10^{-6}	1.448×10^{-5}	1.191×10^{-4}	0.015
	Learner	LR	SVR	SVR	LR	SVR
	Regularization	LASSO	LASSO	LASSO	Ridge	Ridge

4.3. Short-Term PV Power Output Forecasting

K-means clustering is used to label the data. The elbow method is used to determine the optimal number of clusters (k). The plot with the best number of clusters is shaped like an arm. The elbow method gives the results that are shown in Figure 8. The optimal value for k is 3–5 clusters. Ensemble forecasting requires diverse models [51] so different individual models use different datasets [32], or the same dataset uses different parameters [52] so a maximum value must be assigned for k , which in this case is 5.

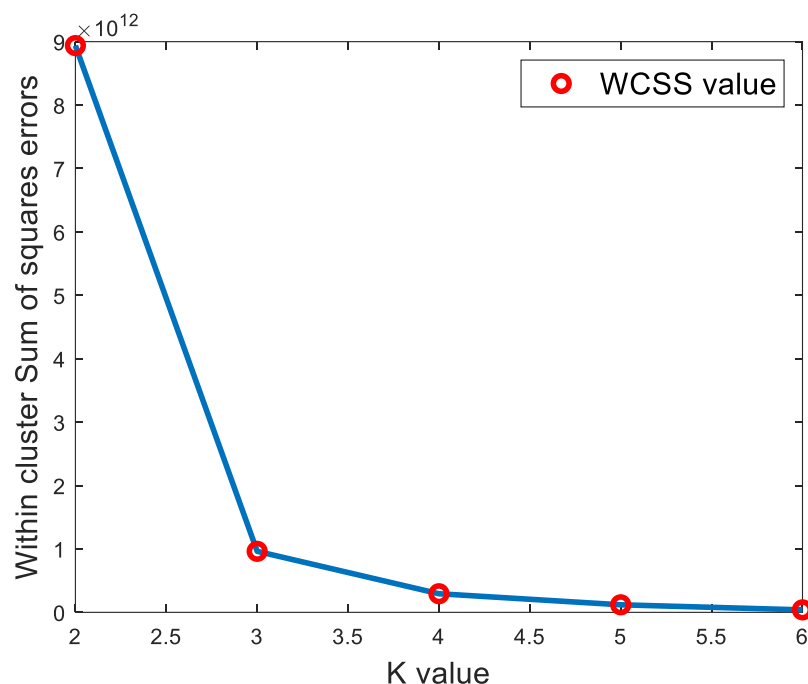


Figure 8. The Elbow method for optimizing clusters.

The results of a previous study [31] show that an ensemble of five models outperforms an ensemble of three models in terms of accuracy. The five weather conditions for this study are sunny, light-cloudy, cloudy, heavy-cloudy, and rainy. The RF models are trained using these five weather conditions, and five RF models are produced: RF₁, RF₂, RF₃, RF₄, and RF₅, for rainy, heavy-cloudy, cloudy, light-cloudy, and sunny, respectively. The dataset for each weather condition is trained using these five RF models in order to calculate a set of weights. Each set contains five weight coefficients and a bias. Regression-based ensemble learning with Bayesian optimization is then used to calculate five optimal sets of weight coefficients and a bias for each weather condition, and Equation (17) is used to calculate the final ensemble forecasting results for each weather condition.

The SVM classification gives the weather conditions for the target day, and an appropriate weight set is used. The SVM classification model receives weather forecasts as input and weather conditions as output. To simulate weather forecasting, a $\pm 20\%$ random error is added to the real weather value.

Figure 9 shows the results for the RF models and the proposed regression-based ensemble forecasting method for sunny weather conditions. The RF₅ model has the lowest MRE value of 7.91% compared to other RF models. The stacking RNN shows that the ensemble method gives more accurate PV power forecasting, with an MRE value of 4.49%. However, the proposed ensemble method provides the most accurate results, with an MRE of 3.49%.

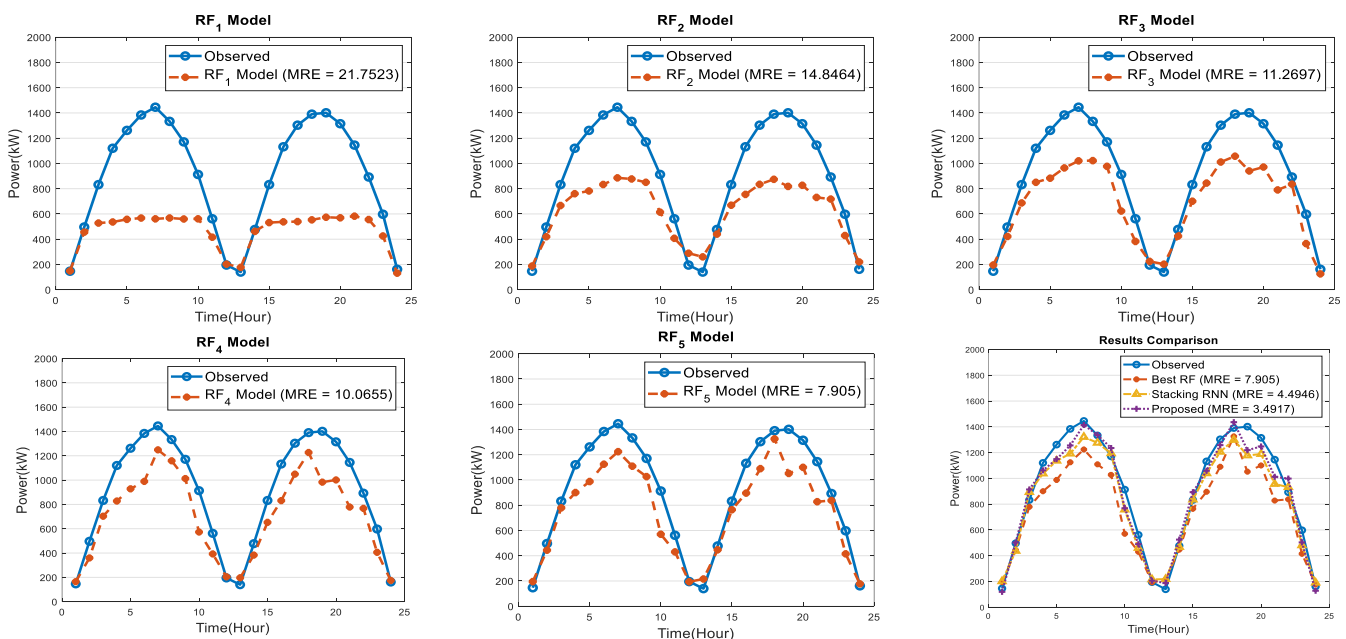


Figure 9. Results for one-day-ahead PV power forecasting for sunny weather conditions.

Figure 10 shows the results for the RF models and the proposed regression-based ensemble forecasting method for light-cloudy weather conditions. Compared to other RF models, the RF₅ model has the lowest MRE value of 5.83%. With an MRE value of 5.61%, the stacking RNN demonstrates that the ensemble model provides more accurate PV power forecasting. On the other hand, the proposed ensemble method produces the most accurate results with an MRE of 5.22%.

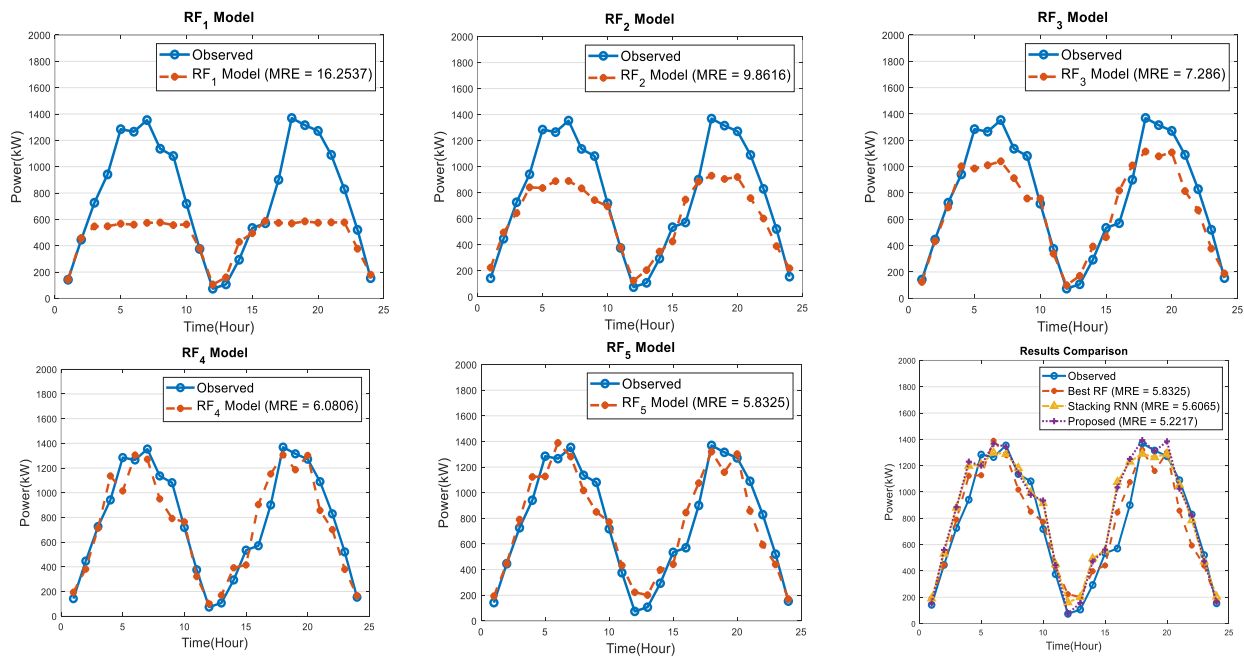


Figure 10. Results for one-day-ahead PV power forecasting for light-cloudy weather conditions.

The results for the RF models and the proposed regression-based ensemble forecasting method for cloudy conditions are shown in Figure 11. The RF₃ model has the lowest MRE value of 6.19% compared to the other RF models. In terms of forecasting PV power, the stacking RNN outperforms the best RF model, with an MRE value of 5.49%. However, the proposed ensemble method has the lowest MRE of 4.19%.

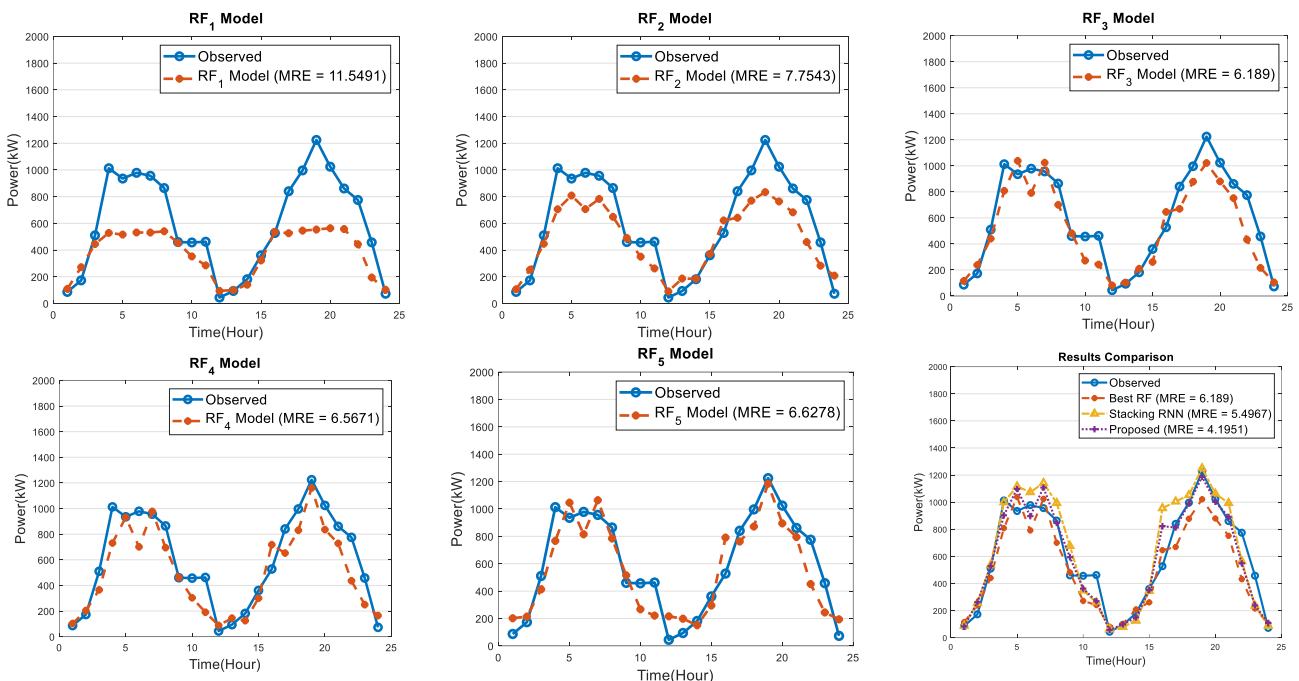


Figure 11. Results for one-day-ahead PV power forecasting for cloudy weather conditions.

The results for the RF models and the proposed regression-based ensemble forecasting method for heavy-cloudy weather conditions are shown in Figure 12. The RF₃ model has the lowest MRE value of 4.62% compared to the other RF models. With an MRE value of 4.4%, the stacking RNN outperforms the best RF model in forecasting PV power. However,

the proposed ensemble method has the lowest MRE of 3.93% compared to all other models.

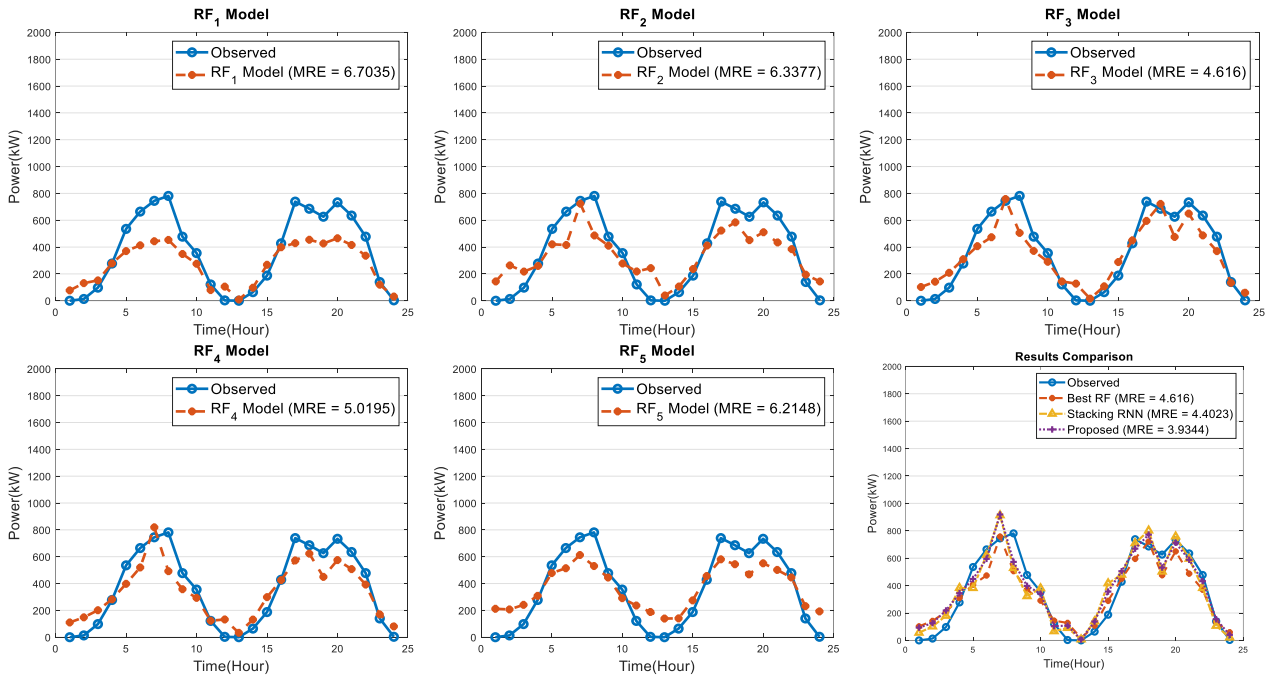


Figure 12. Results for one-day-ahead PV power forecasting for heavy-cloudy weather conditions.

Figure 13 shows the results for the RF models and the proposed regression-based ensemble forecasting method for rainy conditions. The RF₃ has the lowest MRE value of 1.87% compared to the other RF models, but the stacking RNN outperforms the best RF model with an MRE value of 1.76%. Nevertheless, the proposed ensemble method has the best MRE value of 1.59%.

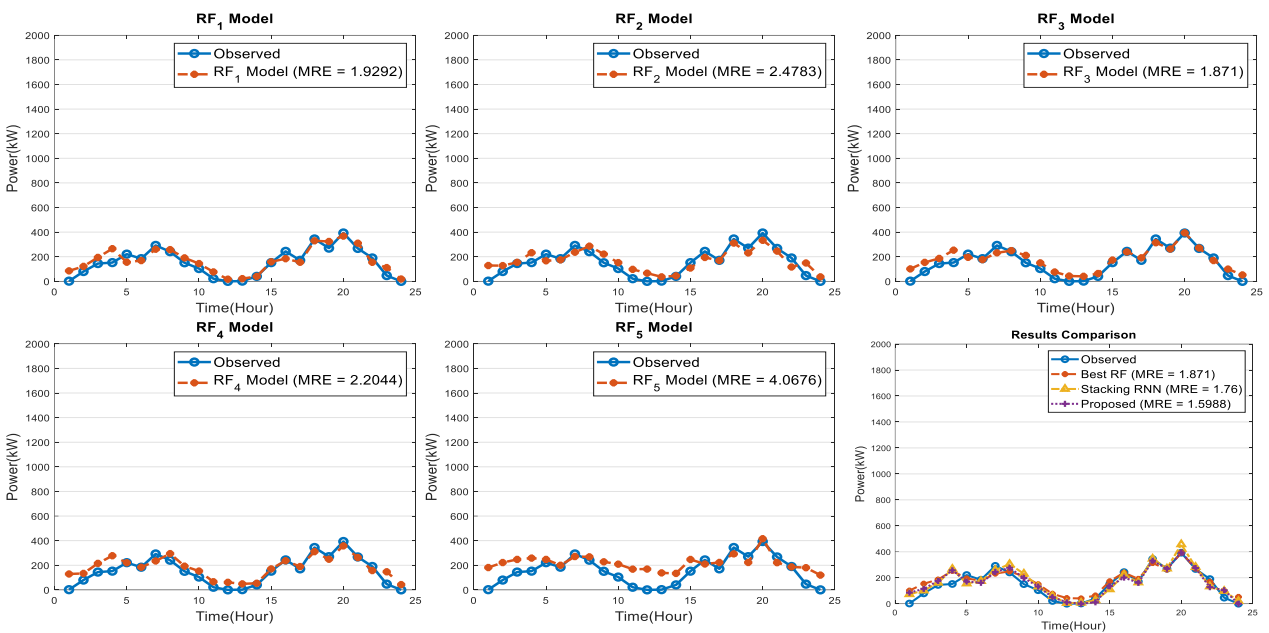


Figure 13. Results for one-day-ahead PV power forecasting for rainy weather conditions.

Table 6 compares the proposed method to the RF model and benchmark method in terms of one-day-ahead observations. The proposed method produces the lowest MRE and MAE values. The MRE for sunny weather conditions is 3.492%, and the MRE values for the stacking RNN and the best RF model are 4.495% and 7.905%, respectively. The proposed method gives an MAE of 69.833 kW, and the best RF and stacking RNN models give MAE values of 158.1 kW and 89.893 kW, respectively. The proposed method achieves a 5.222% MRE and a 104.434 kW MAE for light-cloudy weather conditions. The best RF and stacking RNN models give MRE values of 5.833% and 5.607%, respectively, and MAE values of 116.651 kW and 112.13 kW.

Table 6. Forecasting accuracy for all weather conditions.

Error Validation	Weather Type	Random Forest Model					Ensemble Model	
		RF ₁	RF ₂	RF ₃	RF ₄	RF ₅	Stacking RNN	Proposed Method
MRE (%)	Sunny	21.752	14.846	11.269	10.066	7.905	4.495	3.492
	Light-cloudy	16.254	9.862	7.286	6.081	5.833	5.607	5.222
	Cloudy	11.549	7.754	6.189	6.567	6.628	5.497	4.195
	Heavy-cloudy	6.704	6.338	4.616	5.019	6.215	4.402	3.934
	Rainy	1.929	2.478	1.871	2.204	4.068	1.760	1.599
MAE (kW)	Sunny	435.046	296.928	225.394	201.311	158.100	89.893	69.833
	Light-cloudy	325.073	197.232	145.719	121.612	116.651	112.130	104.434
	Cloudy	230.982	155.086	123.781	131.342	132.556	109.935	83.902
	Heavy-cloudy	134.071	126.755	92.321	100.389	124.296	88.047	78.688
	Rainy	38.584	49.566	37.420	44.087	81.351	35.199	31.976
R ²	Sunny	0	0.372	0.641	0.718	0.812	0.941	0.964
	Light-cloudy	0	0.662	0.827	0.874	0.898	0.866	0.868
	Cloudy	0.297	0.733	0.829	0.795	0.819	0.830	0.893
	Heavy-cloudy	0.649	0.717	0.845	0.825	0.763	0.854	0.891
	Rainy	0.830	0.738	0.829	0.755	0.268	0.835	0.865

For cloudy weather conditions, the proposed method gives an MRE of 4.195% and an MAE value of 83.902 kW, values that are lower than those for the stacking RNN model with a 5.497% MRE and a 109.935 kW MAE. The best RF model is less accurate than the stacking RNN model, which gives a 6.189% MRE and a 123.781 kW MAE. The proposed method is more accurate than the best RF and stacking RNN methods for heavy-cloudy conditions, with a 3.934% MRE and an MAE value of 78.688 kW. The proposed method gives a 1.599% MRE and a 31.976 kW MAE for rainy conditions, and the best RF method performs worse, with a 1.871% MRE and an MAE of 37.42 kW. The MRE value for the stacking RNN method is 1.76% and the MAE value is 35.199 kW.

The proposed method is more accurate than the best RF and benchmark methods in terms of the coefficient of determination (R²). It has a higher R² value for all weather conditions than the best RF and benchmark methods and the lowest R² value is for rainy conditions. The benchmark model stacking RNN is an ensemble method and is much more accurate than a single forecasting method such as the RF model. However, the proposed ensemble method is more accurate than the benchmark stacking RNN model, as demonstrated by the results for the sunny, light-cloudy, cloudy, heavy-cloudy, and rainy datasets.

The proposed regression-based ensemble method for short-term PV power forecasting performance was also tested using the data for seven consecutive days to represent a real industrial application. Figure 14 shows a 7-day comparison of the proposed method, and Figure 15 compares the MRE, MAE, and R² for the best RF model, the stacking RNN, and the proposed method. The best single RF model has an MRE of 5.611%, an MAE of 112.223 kW, and an R² of 0.903. The benchmark method with stacking RNN has an MRE of 4.457%, an MAE of 89.130 kW, and an R² of 0.93, so it is more accurate than the best RF

model. The proposed method has an MRE of 4.362%, an MAE of 87.242 kW, and an R^2 of 0.933, so it is the most accurate method. In terms of the MRE, the proposed method is 22% better than the best RF model and 2% better than the benchmark method.

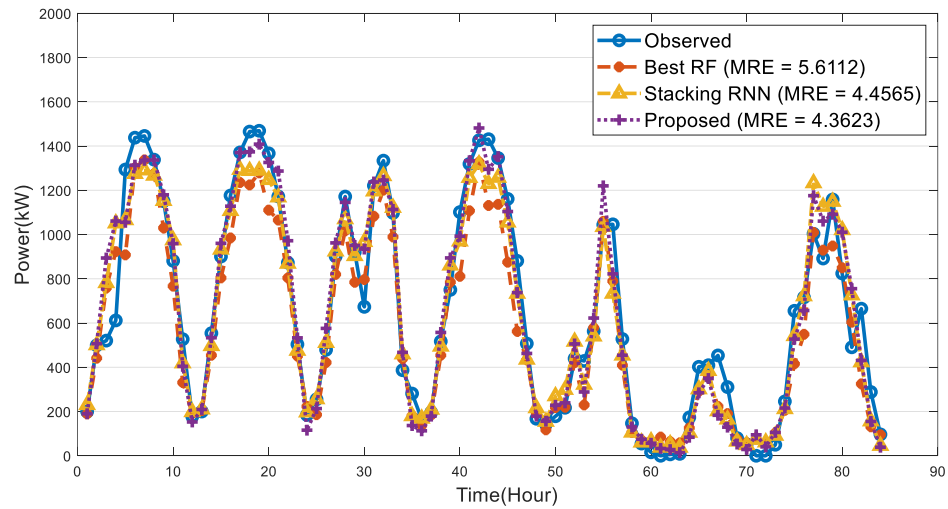


Figure 14. The comparison of the one-day-ahead PV power forecasting results for seven consecutive days (14 May 2020–20 May 2020).

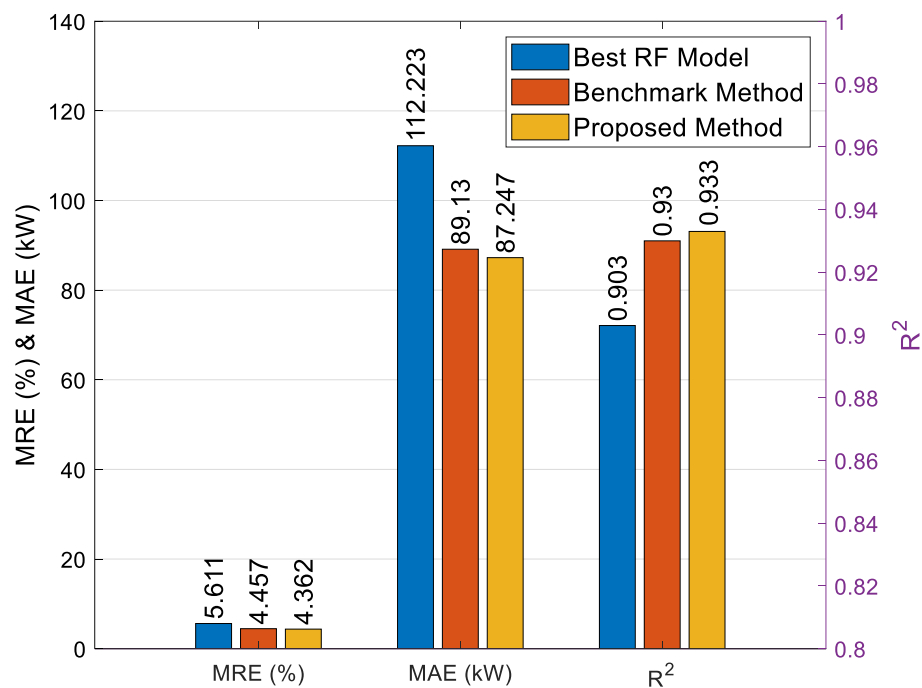


Figure 15. MRE, MAE, R^2 barplot for the best RF, stacking RNN and proposed method.

5. Conclusions

To increase the prediction accuracy for a one-day-ahead PV power forecasting strategy, this study has proposed a short-term PV power forecasting algorithm that uses the regression-based ensemble forecasting method. The ensemble model is constructed by combining individual forecasting models for the RF algorithm. K-means clustering and an SVM classification model are also used to increase the accuracy of the proposed method.

The combination strategy for this study uses linear regression (LR) and support vector regression (SVR), with LASSO and Ridge as regularization methods. The simulation results show that the proposed method is 20% more accurate than the best RF model. The

benchmark ensemble forecasting method for this study is a stacking RNN. The proposed method is 2% more accurate than the stacking RNN. The results of this study show that ensemble forecasting strategies, particularly the proposed method, are much more accurate than single forecasting models.

Future studies will involve the use of a metaheuristic optimization method to determine the optimal weighting coefficients and increase the accuracy of the proposed method. Dynamic ensembles will also replace static ensembles to increase the accuracy by recalculating the weight of the individual prediction model for each new input sample [53].

Author Contributions: This paper is the collaborative work of all authors. Conceptualization, A.A.H.L.; methodology, A.A.H.L., H.-T.Y. and C.-M.H.; validation, A.A.H.L.; writing—original draft preparation, A.A.H.L.; supervision, H.-T.Y. and C.-M.H.; funding acquisition, A.A.H.L., H.-T.Y. and C.-M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study is funded by the Ministry of Science and Technology, Taiwan, under the grant MOST 110-3116-F-006-001.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ANN	Artificial neural network
ARMA	Autoregressive moving average
ARIMA	Autoregressive integrated moving average
CNN	Convolutional neural network
DNN	Deep neural network
ET	Extra trees
ETS	Exponential smoothing
FNN	Feedforward neural network
GRU	Gated recurrent unit
KNN	K-nearest neighbors
LASSO	Least absolute shrinkage and selection operator
LR	Linear regression
LSTM	Long short-term memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MARS	Multivariate Adaptive regression spline
MLP	Multilayer perceptron
MRE	Mean relative error
MSE	Mean squared error
nMAE	Normalized mean absolute error
nMBE	Normalized mean bias error
nRMSE	Normalized root mean squared error
NWP	Numerical weather prediction
OLS	Ordinary least square
PCA	Principal component analysis
PCC	Pearson correlation coefficient
R ²	Coefficient of determination
RBF	Radial basis function
RF	Random forest
RMSE	Root mean squared error
RNN	Recurrent neural network
SARIMA	Seasonal autoregressive integrated moving average
SVM	Support vector machine
SVR	Support vector regression
VAR	Vector autoregressive
WCSS	Within cluster sum of squares
XGBoost	Extreme gradient boosting

References

1. Javaid, N.; Hafeez, G.; Iqbal, S.; Alrajeh, N.; Alabed, M.S.; Guizani, M. Energy efficient integration of renewable energy sources in the smart grid for demand side management. *IEEE Access* **2018**, *6*, 77077–77096. [[CrossRef](#)]
2. Shahid, A. Smart grid integration of renewable energy systems. In Proceedings of the 2018 7th International Conference on Renewable Energy Research and Applications (ICRERA), Paris, France, 14–17 October 2018.
3. Ullah, Z.; Asghar, R.; Khan, I.; Ullah, K.; Waseem, A.; Wahab, F.; Haider, A.; Ali, S.M.; Jan, K.U. Renewable energy resources penetration within smart grid: An overview. In Proceedings of the 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Istanbul, Turkey, 12–13 June 2020.
4. Wan, C.; Zhao, J.; Song, Y.; Xu, Z.; Lin, J.; Hu, Z. Photovoltaic and solar power forecasting for Smart Grid Energy Management. *CSEE J. Power Energy Syst.* **2015**, *1*, 38–46. [[CrossRef](#)]
5. Li, P.; Zhou, K.; Yang, S. Photovoltaic power forecasting: Models and methods. In Proceedings of the 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2) 2018, Beijing, China, 20–22 October 2018.
6. Ahmed, R.; Sreeram, V.; Mishra, Y.; Arif, M.D. 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. [[CrossRef](#)]
7. Rafati, A.; Joorabian, M.; Mashhour, E.; Shaker, H.R. High dimensional very short-term solar power forecasting based on a data-driven heuristic method. *Energy* **2021**, *219*, 119647. [[CrossRef](#)]
8. Son, N.; Jung, M. Analysis of meteorological factor multivariate models for medium- and long-term photovoltaic solar power forecasting using long short-term memory. *Appl. Sci.* **2020**, *11*, 316. [[CrossRef](#)]
9. Mellit, A.; Massi Pavan, A.; Ogliari, E.; Leva, S.; Lughi, V. Advanced methods for photovoltaic output power forecasting: A Review. *Appl. Sci.* **2020**, *10*, 487. [[CrossRef](#)]
10. Massaoudi, M.; Chihi, I.; Abu-Rub, H.; Refaat, S.S.; Oueslati, F.S. Convergence of photovoltaic power forecasting and Deep Learning: State-of-art review. *IEEE Access* **2021**, *9*, 136593–136615. [[CrossRef](#)]
11. Mayer, M.J.; Gróf, G. Extensive comparison of physical models for photovoltaic power forecasting. *Appl. Energy* **2021**, *283*, 116239. [[CrossRef](#)]
12. Wolff, B.; Kühnert, J.; Lorenz, E.; Kramer, O.; Heinemann, D. Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, Numerical Weather Prediction, and Cloud Motion Data. *Sol. Energy* **2016**, *135*, 197–208. [[CrossRef](#)]
13. Singh, B.; Pozo, D. A guide to solar power forecasting using arma models. In Proceedings of the 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe), Bucharest, Romania, 29 September–2 October 2019.
14. Preda, S.; Oprea, S.-V.; Băra, A.; Belciu, A. PV forecasting using support vector machine learning in a big data analytics context. *Symmetry* **2018**, *10*, 748. [[CrossRef](#)]
15. Zhu, H.; Li, X.; Sun, Q.; Nie, L.; Yao, J.; Zhao, G. A power prediction method for photovoltaic power plant based on wavelet decomposition and artificial neural networks. *Energies* **2015**, *9*, 11. [[CrossRef](#)]
16. Son, J.; Park, Y.; Lee, J.; Kim, H. Sensorless PV power forecasting in grid-connected buildings through deep learning. *Sensors* **2018**, *18*, 2529. [[CrossRef](#)] [[PubMed](#)]
17. Meng, M.; Song, C. Daily photovoltaic power generation forecasting model based on random forest algorithm for north China in winter. *Sustainability* **2020**, *12*, 2247. [[CrossRef](#)]
18. Akhter, M.N.; Mekhilef, S.; Mokhlis, H.; Mohamed Shah, N. Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. *IET Renew. Power Gener.* **2019**, *13*, 1009–1023. [[CrossRef](#)]
19. Liu, D.; Sun, K. Random Forest Solar Power Forecast based on classification optimization. *Energy* **2019**, *187*, 115940. [[CrossRef](#)]
20. Mellit, A.; Pavan, A.M.; Lughi, V. Deep Learning Neural Networks for short-term photovoltaic power forecasting. *Renew. Energy* **2021**, *172*, 276–288. [[CrossRef](#)]
21. Niccolai, A.; Dolara, A.; Ogliari, E. Hybrid PV power forecasting methods: A comparison of different approaches. *Energies* **2021**, *14*, 451. [[CrossRef](#)]
22. Seyedmahmoudian, M.; Jamei, E.; Thirunavukkarasu, G.; Soon, T.; Mortimer, M.; Horan, B.; Stojcevski, A.; Mekhilef, S. Short-term forecasting of the output power of a building-integrated photovoltaic system using a metaheuristic approach. *Energies* **2018**, *11*, 1260. [[CrossRef](#)]
23. Aprillia, H.; Yang, H.-T.; Huang, C.-M. Short-term photovoltaic power forecasting using a convolutional neural network–salp swarm algorithm. *Energies* **2020**, *13*, 1879. [[CrossRef](#)]
24. Yang, T.; Li, B.; Xun, Q. LSTM-attention-embedding model-based day-ahead prediction of photovoltaic power output using Bayesian optimization. *IEEE Access* **2019**, *7*, 171471–171484. [[CrossRef](#)]
25. Wang, F.; Zhen, Z.; Wang, B.; Mi, Z. Comparative study on KNN and SVM based weather classification models for day ahead short term solar PV power forecasting. *Appl. Sci.* **2017**, *8*, 28. [[CrossRef](#)]
26. Hossain, M.S.; Mahmood, H. Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast. *IEEE Access* **2020**, *8*, 172524–172533. [[CrossRef](#)]
27. Massaoudi, M.; Abu-Rub, H.; Refaat, S.S.; Trabelsi, M.; Chihi, I.; Oueslati, F.S. Enhanced deep belief network based on ensemble learning and tree-structured of parzen estimators: An Optimal Photovoltaic Power Forecasting Method. *IEEE Access* **2021**, *9*, 150330–150344. [[CrossRef](#)]

28. Yang, D.; Dong, Z. Operational Photovoltaics Power Forecasting using seasonal time series ensemble. *Sol. Energy* **2018**, *166*, 529–541. [[CrossRef](#)]
29. Ahmad, M.W.; Mourshed, M.; Rezgui, Y. Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression. *Energy* **2018**, *164*, 465–474. [[CrossRef](#)]
30. Wang, J.; Qian, Z.; Wang, J.; Pei, Y. Hour-ahead photovoltaic power forecasting using an analog plus neural network ensemble method. *Energies* **2020**, *13*, 3259. [[CrossRef](#)]
31. Lateko, A.A.H.; Yang, H.-T.; Huang, C.-M.; Aprillia, H.; Hsu, C.-Y.; Zhong, J.-L.; Phuong, N.H. Stacking Ensemble method with the RNN meta-learner for short-term PV power forecasting. *Energies* **2021**, *14*, 4733. [[CrossRef](#)]
32. Eom, H.; Son, Y.; Choi, S. Feature-Selective Ensemble Learning-based long-term regional PV generation forecasting. *IEEE Access* **2020**, *8*, 54620–54630. [[CrossRef](#)]
33. Zhu, R.; Guo, W.; Gong, X. Short-term photovoltaic power output prediction based on k-fold cross-validation and an ensemble model. *Energies* **2019**, *12*, 1220. [[CrossRef](#)]
34. Pan, C.; Tan, J. Day-ahead hourly forecasting of solar generation based on cluster analysis and Ensemble Model. *IEEE Access* **2019**, *7*, 112921–112930. [[CrossRef](#)]
35. Liu, L.; Zhan, M.; Bai, Y. A recursive ensemble model for forecasting the power output of photovoltaic systems. *Sol. Energy* **2019**, *189*, 291–298. [[CrossRef](#)]
36. Wang, Y.; Liao, W.; Chang, Y. Gated recurrent unit network-based short-term photovoltaic forecasting. *Energies* **2018**, *11*, 2163. [[CrossRef](#)]
37. Wu, Y.-K.; Lai, Y.-H.; Huang, C.-L.; Phuong, N.T.; Tan, W.-S. Artificial Intelligence Applications in estimating invisible solar power generation. *Energies* **2022**, *15*, 1312. [[CrossRef](#)]
38. Bajpai, A.; Duchon, M. A Hybrid approach of solar power forecasting using machine learning. In Proceedings of the 2019 3rd International Conference on Smart Grid and Smart Cities (ICSGSC), Berkeley, CA, USA, 25–28 June 2019.
39. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
40. Niu, D.; Wang, K.; Sun, L.; Wu, J.; Xu, X. Short-term photovoltaic power generation forecasting based on random forest feature selection and CEEMD: A case study. *Appl. Soft Comput.* **2020**, *93*, 106389. [[CrossRef](#)]
41. Kim, Y.; Hur, J. An ensemble forecasting model of wind power outputs based on improved statistical approaches. *Energies* **2020**, *13*, 1071. [[CrossRef](#)]
42. Tao, D.; Ma, Q.; Li, S.; Xie, Z.; Lin, D.; Li, S. Support vector regression for the relationships between ground motion parameters and macroseismic intensity in the SICHUAN–Yunnan region. *Appl. Sci.* **2020**, *10*, 3086. [[CrossRef](#)]
43. Marco, R.; Ahmad, S.S.; Ahmad, S. Bayesian hyperparameter optimization and Ensemble Learning for Machine Learning Models on software effort estimation. *Int. J. Adv. Comput. Sci. Appl.* **2022**, *13*, 419–429. [[CrossRef](#)]
44. Solcast API Toolkit. Available online: <https://toolkit.solcast.com.au/weather-sites/48bb-7a5e-a09e-227f/detail> (accessed on 25 March 2022).
45. Jebli, I.; Belouadha, F.-Z.; Kabbaj, M.I.; Tilioua, A. Prediction of solar energy guided by Pearson Correlation Using Machine Learning. *Energy* **2021**, *224*, 120109. [[CrossRef](#)]
46. Yang, H.-T.; Huang, C.-M.; Huang, Y.-C.; Pai, Y.-S. A weather-based hybrid method for 1-day ahead hourly forecasting of PV Power Output. *IEEE Trans. Sustain. Energy* **2014**, *5*, 917–926. [[CrossRef](#)]
47. Nhuchhen, D.R.; Abdul Salam, P. Estimation of higher heating value of biomass from proximate analysis: A new approach. *Fuel* **2012**, *99*, 55–63. [[CrossRef](#)]
48. Qian, X.; Lee, S.; Soto, A.-M.; Chen, G. Regression model to predict the higher heating value of poultry waste from proximate analysis. *Resources* **2018**, *7*, 39. [[CrossRef](#)]
49. Chen, Y.; Zhang, D. Theory-guided deep-learning for electrical load forecasting (TGDLF) via ensemble long short-term memory. *Adv. Appl. Energy* **2021**, *1*, 100004. [[CrossRef](#)]
50. Jin, X.-B.; Zheng, W.-Z.; Kong, J.-L.; Wang, X.-Y.; Bai, Y.-T.; Su, T.-L.; Lin, S. Deep-learning forecasting method for electric power load via attention-based encoder-decoder with bayesian optimization. *Energies* **2021**, *14*, 1596. [[CrossRef](#)]
51. Ren, Y.; Suganthan, P.N.; Srikanth, N. Ensemble methods for wind and solar power forecasting—A state-of-the-art review. *Renew. Sustain. Energy Rev.* **2015**, *50*, 82–91. [[CrossRef](#)]
52. Wang, L.; Mao, S.; Wilamowski, B.M.; Nelms, R.M. Ensemble learning for load forecasting. *IEEE Trans. Green Commun. Netw.* **2020**, *4*, 616–628. [[CrossRef](#)]
53. Chen, Z.; Koprinska, I. Ensemble methods for solar power forecasting. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020.