Wiley
International Journal of Energy Research
Volume 2024, Article ID 3098943, 26 pages
https://doi.org/10.1155/2024/3098943

WILEY

Research Article

Multiphase Solar Photovoltaic Prediction Model Based on Season, Hierarchical *k*-Means Clustering, GRA-PCC, SVM, and Neural Network

Mariz B. Arias [1,2] **and Sungwoo Bae** [1]

[1]*Department of Electrical Engineering, Hanyang University, Seoul 04763, Republic of Korea*
[2]*Department of Electrical Engineering, University of Santo Tomas, España, Manila 1015, Philippines*

Correspondence should be addressed to Sungwoo Bae; swbae@hanyang.ac.kr

Received 31 December 2023; Revised 3 April 2024; Accepted 4 May 2024

Academic Editor: Muharrem Hilmi Aksoy

Solar photovoltaic (PV) has accounted for the highest percentage of power generation capacity among other renewables. However, solar PV power generation is highly variable because of different factors; therefore, accurate forecasting is critical for reliable integration into the power system. This paper proposes a multiphase solar PV prediction model that includes grouping, clustering, linking, classifying, and predicting using historical solar PV power and weather data. Seasonal variation is considered in the grouping phase, followed by hybrid hierarchical *k*-means clustering to enhance data division in the clustering phase. A hybrid gray relational analysis-Pearson correlation coefficient identifies significant weather factors impacting solar PV power in the linking phase. The classification phase employs a support vector machine to establish the relationship between the clusters and the relevant weather factors. Lastly, a neural network (NN) is trained to predict solar PV power. The solar PV power profiles are presented to show the variability in season and time. The simulation results of the proposed model showed relatively accurate forecasting results, including MAE of 0.408 MW, MSE of 460.51 MW, RMSE of 0.679 MW, nRMSE of 4.345%, and MRE of 2.266%. These results represent that the uncertainties of the proposed model are 6 and 12 times lower than those of the conventional methods (i.e., conventional NN and ARMAX). These results assure that the proposed model can provide more accurate solar PV power profiles for reliable power system integration.

1. Introduction

According to the Renewables 2022 Global Status Report [1], there are already 3,146 GW of installed renewable power capacity as of 2021. In that year alone, 315 GW of new capacity was added, predominantly in solar photovoltaic (PV) and wind power, due to their cost-effectiveness and government support [1]. In addition, the levelized costs of solar PV and onshore wind power have become more economical than fossil fuels on average [1]. Half of the added capacities, amounting to 175 GW, were from solar PV, reaching a cumulative total capacity of 942 GW by the end of 2021 [1].

The increasing number of solar PV systems installed has raised concerns regarding their integration into power systems, leading to several issues related to power quality, system stability, and reliability [2, 3, 4, 5, 6, 7]. Comprehensive reviews in studies [3, 4, 5] highlight various power quality challenges associated with integrating solar PV systems into the utility grid. As solar PV penetration increases, these challenges are expected to escalate [3, 4, 5]. Another study [6] developed a framework assessing the impact of high solar PV penetration on grid voltage stability, considering the stochastic nature of solar PV power generation and load demand using Monte Carlo simulation. Additionally, studies [7] analyze grid-connected solar PV systems' reliability, availability, and maintainability to estimate the best probability density function for the system's failure rate. Moreover, solar PV system has also been increasing in the grid and microgrid to replace conventional synchronous generators. In study [8], a grid-connected inverter for PV-powered electric vehicle charging stations.

As highlighted in previous studies, integrating solar PV systems into the grid presents various challenges, primarily due to their intermittent nature, which depends on irradiance, temperature, and atmospheric conditions. The highly variable generation of solar PV power, influenced by solar irradiance and meteorological factors, can impact the power system network's balance between supply and demand. Enhancing the operation and integration of solar PV power systems requires accurate prediction, which is crucial for power system engineers to assess shortages or excesses in solar PV power generation. Reliable prediction results play a significant role in energy management systems. Consequently, different prediction models have been developed to forecast solar PV power and address these challenges.

Solar PV forecasting techniques have three main categories: physical, statistical, and hybrid approaches [9, 10, 11, 12]. The physical approach relies on a theoretical simulation model based on design and physical weather parameters, eliminating the need for historical data [9, 10, 11]. In contrast, the statistical approach includes data-driven models utilizing historical time series and real-time generated data [9, 11, 12]. Statistical methods include time series-based forecasting, such as exponential smoothing, autoregressive moving average model, and machine learning forecasting, such as artificial neural network (NN) and deep learning [12, 13, 14]. The hybrid approach combines two approaches, either physical or statistical, or two or more statistical methods [9, 10, 11, 12, 13, 14]. With the limitations of the individual techniques discussed above, hybrid methods were presented to enhance the strengths and improve the forecasting performance of the techniques. Recent studies have mainly focused on developing hybrid strategies to improve the accuracy and implementation of the prediction model.

1.1. Literature Review on Solar PV Prediction Models. Recent studies have developed different hybrid methods for forecasting solar PV to enhance the performance of individual methods [15, 16, 17, 18, 19, 20, 21, 22]. A detailed comparison of the previous studies on solar PV prediction models is listed in Table 1. Many of these hybrid models employ combinations of various statistical methods, particularly machine learning technologies [15, 16, 17, 18]. Several studies [15, 16, 17, 18] integrate the long short-term memory (LSTM) network with other methods. LSTM, a recurrent neural network (RNN), improves the performance of the traditional RNN model by overcoming the gradient vanishing problem [15, 16, 17, 18, 19, 20]. For instance, the LSTM was combined with wavelet packet decomposition (WPD) to decompose the original PV power series and weather variables (global horizontal radiation (GHR), diffuse horizontal radiation (DHR), ambient temperature (AT), wind speed (WS), and relative humidity (RH)) as inputs to develop the LSTM network in a study [15]. In this study [15], the WPD was applied to decompose the original PV power time series into four sublayers, which are the inputs in the LSTM networks to produce forecasting results from each LSTM network which are integrated based on linear weighting method to obtain improved forecasting results.

Conversely, other studies [16, 17, 18] combined the LSTM with a convolutional neural network (CNN), a deep learning algorithm used as a hierarchical feature extractor with layer architecture. In a different study [16], by analyzing historical patterns, CNN was utilized to classify daily weather as sunny or cloudy. At the same time, LSTM was split into models trained separately to extract long-term dependent features from the raw data to predict solar PV power generation according to the weather condition type. Another study [17] developed a prediction model that combines LSTM with stochastic differential equations to consider the randomness of solar PV power in different seasons and simultaneously extract the potential laws of historical data. This study [17] used a wavelength analysis and automatic encoder to decompose data and extract its essential features. This hybrid prediction model [17] provided a root mean square error (RMSE) of 4.4647, lower than the presented individual prediction models. In another study [18], a short-term forecasting approach based on bidirectional LSTM-CNN (BiLSTM-CNN) was proposed to develop a prediction of regional PV power plants using a representative power plant. A *k*-means algorithm was used to divide power plants with similar generation characteristics, and then a representative power plant in each subregion was selected based on correlation coefficients [18]. Using the representative plant's historical operation and meteorological data, the prediction model was developed based on the BiLSTM-CNN method [18].

Alternatively, studies [19, 20, 21, 22] integrated multiple methods to formulate the solar PV prediction model. The study [19] involved data processing for extracting daytime data, data augmenting through time-series generative adversarial networks, clustering based on *k*-medoids considering various weather conditions, and predicting solar PV power using CNN-gated recurrent units. The PV power data augmentation method based on TimeGAN was considered to get the expanded PV power data [19]. The soft-DTW and *k*-medoids clustering was developed to classify PV power under different weather conditions [19]. Finally, a hybrid NN model was developed to integrate CNN and GRU into a unified framework for accurate PV power forecasting [19]. Another study [20] developed a prediction model combining the Pearson correlation coefficient (PCC), ensemble empirical modal decomposition (EEMD), sample entropy (SE), sparrow search algorithm (SSA), and LSTM. In this study [20], PCC screened meteorological factors: EEMD decomposed temperature, global horizontal irradiance (GHI), and PV power; calculated the SE values of the components and reconstructed them according to the SE values to reduce the computational cost of the prediction model; and SSA algorithm optimizes the structural parameters of the LSTM to minimize the error of the prediction model.

Recent studies [21, 22] developed an ensemble prediction model that combines clustering, classification, and regression to improve the accuracy of solar PV prediction. A recent study [21] incorporated *k*-means clustering, random forest (RF) models, regression-based methods with LASSO, and ridge regularizations. This approach utilized *k*-means clustering to divide the historical daily average PV power and

Table 1: Detailed comparison of hybrid solar PV prediction models with previous works.

References	Method	Region	Forecasting horizon	Data	Rated capacity	Best metrics
Li et al. [15]	WPD-LSTM	Australia	1 hr ahead (15 min interval)	PV output power, GHR, DHR, AT, WS, and RH	26.5 kW	MBE = 0.0396 kW MAPE = 1.8681% RMSE = 0.1526 kW
Lim et al. [16]	CNN-LSTM	Korea	1 day ahead	PV power data	2,500 W	MAPE = 4.58% RMSE = 43.87 W MAE = 34 W $R^2 = 0.99$
Zhanga and Kong [17]	LSTM-SDE	China	1 day ahead	PV power data, ambient temperature, air humidity ratio, solar radiation intensity, average soil temperature, wind speed	10 MW	RMSE = 4.4647 MW
Li et al. [18]	BiLSTM-CNN	China	1 day ahead	PV power data, temperature, dew point temperature, air pressure, and wind speed	18 MW	MSE = 0.037 MW RMSE = 0.191 MW MAE = 0.144 MW
Li et al. [19]	TimeGAN-Soft DTW-based K-medoids-CNN-GRU	China	1 day ahead (15-min interval)	PV power data	50 MW	RMSE = 1.003 MW MAE = 0.643 MW $R^2 = 0.997$
Li et al. [20]	PCC-EEMD-SSA-LSTM	China	1 day ahead (15-min interval)	PV power data, temperature, and GHI	100 MW	RMSE = 1.7596 MW $R^2 = 0.9971$
Lateko et al. [21]	*k*-means clustering, RF models, regression-based methods with LASSO, and ridge regularizations	Taiwan	1 day ahead	PV power data, irradiance, temperature, precipitation, relative humidity, and wind speed	2,000 kW	MRE = 1.599% MAE = 31.976 kW $R^2 = 0.964$
Lin and Li [22]	Hybrid Kmeans-GRA-SVR	Australia	1 day ahead	PV power data, GHI, diffuse horizontal irradiance, relative humidity, and temperature	5.83 kW	MAE = 0.1618 kW RMSE = 0.2132 kW

hourly weather information into clusters [21]. Each cluster was then trained using RF models and regression-based methods, including linear regression, support vector regression (SVR), LASSO, and ridge regularization, to select the optimal set of weights to combine with the result of RF models [21]. Similarly, another recent study [22] also developed a prediction model based on *k*-means clustering, gray relational analysis (GRA), and SVR. This approach utilized *k*-means clustering to cluster the historical PV power data of the four seasons [22]. The historical data for four seasons was clustered using *k*-means, and GRA was used to assess correlations within the formed clusters; the SVR was then trained to create the final prediction model [22].

1.2. Research Gaps and Contributions. The hybrid solar PV prediction models discussed above have provided accurate and promising forecasting results. However, continuous research and improvements are necessary to address specific issues. Recent studies focused on enhancing the performance of forecasting methods by combining various approaches. Studies [15, 16, 17, 20, 21, 22, 23] used historical solar PV power and weather data to formulate the prediction model. However, some of these studies [16, 20] do not consider the seasonal effect of these variables, which could impact the accuracy of the prediction model. Addressing these aspects in future research could further improve solar PV prediction models. Hence, this paper considers the seasonal effect of predicting solar PV power, contributing to a more accurate forecasting result.

In addition to the seasonal trend of solar PV power, the variations and patterns of solar PV power data were also considered in several studies [19, 21, 22, 23]. In studies [21, 22], a *k*-means clustering algorithm was used to divide historical solar PV power and weather data into clusters based on similarities measured by Euclidean distance. Another study [19] employed *k*-medoids for clustering solar PV power based on the characteristics of weather conditions. Unlike *k*-means, *k*-medoids rely on medoids calculated by minimizing the absolute distance between the points and the selected centroid, as opposed to minimizing the square distance [24]. While *k*-medoids reduced the drawbacks of *k*-means, the effectiveness of both algorithms is directly influenced by the chosen centroids. In contrast, study [23] used hierarchical clustering to group solar irradiance data in a multilevel hierarchy for forecasting solar PV power. Moreover, the hybrid hierarchical *k*-means clustering method was developed to avoid random centroid selection [25, 26]. Thus, the use of hybrid hierarchical *k*-means clustering contributes to the improvement of solar PV prediction.

Several studies [15, 16, 17, 20, 21, 22, 23] have already considered weather data in predicting solar PV power, recognizing the significant influence of these variables on solar PV power generation. Similarly, previous studies [27, 28, 29] also considered the effects of dynamic environmental conditions in providing a good model for the lifetime prediction of PV modules. These studies present that evaluating the impact of weather data in solar PV prediction will further improve the accuracy of the prediction model. In studies [17, 20, 21, 22], the PCC was used to select the weather data correlating with solar PV power. Meanwhile, GRA was employed to measure the degree of correlation of factors to the reference variable for similar days in the prediction model formulation in a study [22]. In study [23], GRA was used to determine the influential factors affecting the solar irradiance to be used in forecasting solar PV power. GRA reflects the degree of correlation of each element to a reference, while PCC analyzes the correlation between each factor and the reference value. Combining GRA and PCC allows the identification of the most relevant factors influencing the reference value [30]. Thus, a hybrid gray relational analysis-Pearson correlation coefficient (GRA-PCC) approach has the potential to enhance the accuracy of the solar PV power prediction model.

This study proposes a hybrid approach to the solar PV power prediction model, incorporating various phases to enhance the accuracy and performance of the prediction model used in previous studies. This paper considers historical solar PV power data and weather data in formulating the proposed prediction model. Five phases are considered: grouping, clustering, linking, classifying, and predicting. In the grouping phase, the solar PV power and weather data are grouped according to season to consider the seasonal variation. A hybrid hierarchical *k*-means clustering is proposed to enhance data division of the grouped solar PV power and weather data in the clustering phase. In addition, a hybrid GRA-PCC is used to identify significant weather factors impacting solar PV power in the linking phase. The classification phase employs a support vector machine (SVM) to establish the relationship between the clusters and the relevant weather factors. These first four phases are used to process the training data to formulate the prediction model in the last phase. A NN is used to prepare the training data to predict solar PV power in the prediction phase.

Compared to earlier studies on solar PV power prediction models, the proposed multiphase solar PV prediction model considers different parameters that affect solar PV power, such as seasonal variation, daily pattern variation, and the effect of various weather factors. As another hybrid technique and method for accurately predicting solar PV power, the contributions of this paper are the following: (1) A multiphase solar PV prediction model, which includes five phases: grouping, clustering, linking, classifying, and predicting, is developed using historical hourly solar PV power and weather data of different weather factors. Each phase enhances the accuracy of the solar PV prediction model, which considers seasonal variation, daily pattern variation, and relevant weather factors that significantly affect solar PV power. (2) Seasonal fluctuation of solar PV power is considered. Discounting the seasonal trends in formulating the solar PV prediction model affects its performance. (3) A hybrid hierarchical *k*-means clustering is developed to account for the daily pattern variations in solar PV power. The daily pattern variations should be considered when formulating the prediction model of solar PV since it provides more informative data that will improve the model's accuracy. (4) A hybrid GRA-PCC is established to identify

relevant weather factors significantly affecting solar PV power. It is essential to consider only those factors that significantly affect solar PV power among various weather factors when formulating the solar PV prediction model. (5) A relatively accurate forecasting result is achieved compared to the conventional method. The best metrics demonstrate that the uncertainties obtained using the proposed prediction model are six times lower than those of the conventional method, NN, and 12 times lower than those of autoregressive moving averages with exogenous (ARMAX).

The remainder of this paper is organized as follows: the proposed multiphase solar PV prediction model, which describes the process in each phase, is explained in Section 2. Simulation results, which illustrate the results of each phase and the solar PV power profiles, are presented in Section 3 to demonstrate the effectiveness of the proposed prediction model. Finally, Section 4 concludes the paper with a summary of the findings.

2. Multiphase Solar PV Prediction Model

The proposed multiphase solar PV prediction model is formulated using solar PV power data which is collected from one of the solar PV power plants in the Philippines and various weather factors such as temperature, wind speed, precipitation, humidity, visibility, pressure, cloud cover, heat index, dewpoint, windchill, UV index, and weather description collected from the location of the chosen solar PV power plant through the weather source online [31]. Twenty-one weather descriptions, such as clear, cloudy, fog, and heavy rain, are represented by values one to 21 [31]. Two-year hourly solar PV power data and weather data from a solar PV power plant with 18 MW rated power capacity, collected throughout the day (i.e., 0:00 to 23:00) from January 1, 2019, to December 31, 2020, was used in this paper. The data collected in 2019 were considered training data, while data from 2020 were testing data. The data were divided to represent each month's data in the training and testing data.

These weather factors were considered as they affect the solar PV system's performance, affecting the actual solar PV power output. The temperature plays a crucial role in the PV conversion process, in which the efficiency and the power output of the PV are reduced by 0.03%–0.05% for every 1°C increased temperature [32, 33]. Wind speed affects the operating temperature of PV panels; hence, PV panels with wind speed can generate a higher PV power output [34]. In addition, a study [35] reported a wind chill of 15–20°C for wind speeds of 10 m/s at solar irradiance of about 1,000 W/m^2; hence, the wind chill was affected by wind speed and solar irradiance, which further affected the PV power production. A study [36] concluded that precipitation has a 26% increase in power generation. Humidity is the amount of water vapor measured in the air, affecting performance, estimating a loss of approximately 15%−30% in PV power [37, 38]. Air pressure is directly proportional to light intensity, and light intensity is directly proportional to solar panel power output [38]. A study [39] stated that cloud cover has an immediate effect, resulting in lower PV power production. Heat index increases the PV panels' operating temperature, causing a 0.45% decrease in the PV power output [40]. An increase in dew point, the temperature at which air becomes saturated and condenses, decreases the PV system's efficiency, thus decreasing the production of solar PV [37]. Study [41] considered the UV index in forecasting solar PV power, which provided higher accuracy than the model without considering the UV index. Hence, all the considered weather factors impact the generation of PV power. Factors such as temperature, humidity, visibility, cloud cover, and heat index are inversely proportional to the PV power production output. In contrast, wind speed, precipitation, pressure, wind chill, and UV index are directly proportional to the PV power output. At the same time, the effect of weather descriptions on PV power output can be either direct or inverse. Thus, considering the mentioned weather factors is necessary in formulating an enhanced solar PV prediction model.

Figure 1 shows the operational framework for formulating the multiphase solar PV prediction model, which comprises the following five phases: Phase 1: Grouping, Phase 2: Clustering, Phase 3: Linking, Phase 4: Classifying, and Phase 5: Predicting. The input data are the historical solar PV power data and the historical weather factor data (temperature, wind speed, precipitation, humidity, visibility, pressure, cloud cover, heat index, dewpoint, windchill, UV index, and weather description). The procedure for formulating the multiphase solar PV prediction model is summarized as follows:

(i) These input data undergo normalization and are then grouped according to season in Phase 1.
(ii) In Phase 2, the grouped solar PV power data were clustered based on similarities using the proposed hybrid hierarchical *k*-means clustering algorithm. The solar PV power data will be clustered first using hierarchical clustering and then using the means of each cluster as the initial centers. The *k*-means clustering will generate the final solar PV power data clusters by season.
(iii) Phase 3 is used to identify the most relevant and significant factors affecting solar PV power from the considered weather factors (temperature, wind speed, precipitation, humidity, visibility, pressure, cloud cover, heat index, dewpoint, windchill, UV index, and weather description) using the proposed hybrid GRA-PCC method. The significant weather factors are first determined using GRA, and then from these identified factors, the most relevant weather factors affecting solar PV power are determined using the PCC method.
(iv) In Phase 4, a classification model links cluster outputs of Phase 2 and relevant weather factors in Phase 3 using SVM classification. Lastly, the prediction phase uses the NN to forecast solar PV power. A detailed discussion of the steps and algorithms used in each phase is mentioned in the following subsections.

2.1. Phase 1: Grouping. Phase 1 of the model is the grouping wherein the input data are the solar PV power data (i.e., $X_{i,j}$)

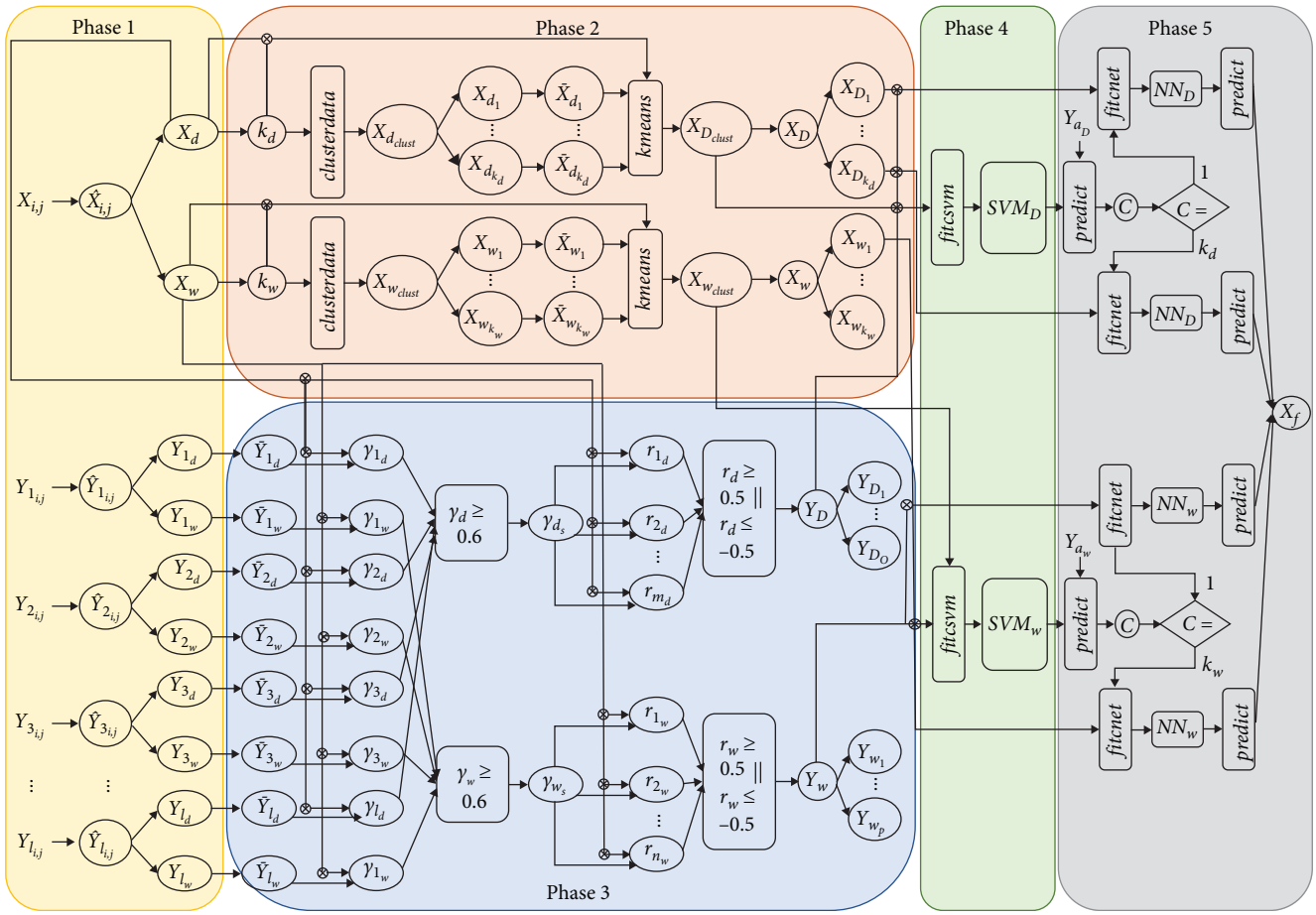


FIGURE 1: Operational framework of the multiphase solar PV prediction model.

and weather factor data (i.e., $Y_{1_{i,j}}$, $Y_{2_{i,j}}$, …, $Y_{l_{i,j}}$) where i is the day from 1 to 365, j is the hour from 0 to 23, and l is the number of weather factors which is 12. These inputs undergo normalization for comparison using the following:

$$\widehat{X}_{i,j} = \frac{X_{i,j} - X_{\min}}{X_{\max} - X_{\min}}, \tag{1}$$

$$\widehat{Y}_{l_{i,j}} = \frac{Y_{l_{i,j}} - Y_{l_{\min}}}{Y_{l_{\max}} - Y_{l_{\min}}}, \tag{2}$$

where $\widehat{X}_{i,j}$ is the normalized solar PV power data, $X_{i,j}$ is the solar PV power data, $X_{\min}$ and $X_{\max}$ are the minimum and maximum solar PV power data, respectively, $\widehat{Y}_{l_{i,j}}$ is the normalized data of each of the weather factor data, $Y_{l_{i,j}}$ is each of the weather factor data, $Y_{l_{\min}}$ and $Y_{l_{\max}}$ are the minimum and maximum data for each weather factor, respectively. The next step involves grouping based on the season. Since the data were collected in the Philippines, which experiences two seasons, this study distinguishes between the dry and wet seasons. The dry season comprises January, February, March, April, May, and December, while the wet season spans June, July, August, September, October, and November. Both solar PV power data and weather data will be grouped according to the following:

$$\widehat{X}_{i,j} = \begin{cases} X_d & \begin{matrix} i = 1:151:335:365 \\ j = 7:20 \end{matrix} \\ X_w & \begin{matrix} i = 152:334 \\ j = 7:20 \end{matrix} \end{cases}, \tag{3}$$

$$\widehat{Y}_{l_{i,j}} = \begin{cases} Y_{l_d} & \begin{matrix} i = 1:151:335:365 \\ j = 7:20 \end{matrix} \\ Y_{l_w} & \begin{matrix} i = 152:334 \\ j = 7:20 \end{matrix} \end{cases}, \tag{4}$$

where X_d and X_w are the solar PV power data in the dry and wet seasons, respectively, Y_{l_d} is each of the weather factor data in the dry season, and Y_{l_w} is each of the weather factor data in the wet season. This grouped data will be the basis of the subsequent phases.

2.2. Phase 2: Clustering. Phase 2 involves clustering, which divides solar PV power data based on similarities. The two most popular clustering algorithms are k-means clustering and hierarchical clustering. K-means is an iterative algorithm that partitions data into k mutually exclusive clusters using a

Input: Solar PV power data in the dry season (X_d) and solar PV power data in the wet season (X_w)

Output: Solar PV power data in each cluster in the dry season ($X_D = (X_{D_1}, X_{D_2}, \cdots, X_{D_{k_d}})$) and solar PV power data in each cluster in the wet season ($X_W = (X_{W_1}, X_{W_2}, \cdots, X_{W_{k_w}})$)

1. Read X_d and X_w.
2. Calculate k_d and k_w.
   $k_d = \text{optimal}(evalclusters\ (X_d))$
   $k_w = \text{optimal}(evalclusters\ (X_w))$
3. Perform hierarchical clustering using k_d and k_w.
   $X_{d_{clusters}} = clusterdata(X_d, k_d)$
   $X_{w_{clusters}} = clusterdata(X_w, k_w)$
4. Split X_d and X_w according to the cluster outputs ($X_{d_{clusters}}$ and $X_{w_{clusters}}$), respectively.

for $i = 1:365$	for $i = 1:365$
if $X_{d_{clusters}}(i) = 1$	if $X_{w_{clusters}}(i) = 1$
place $X_d(i)$ to X_{d_1}	place $X_w(i)$ to X_{w_1}
⋮	⋮
else if $X_{d_{clusters}}(i) = k_d$	else if $X_{w_{clusters}}(i) = k_w$
place $X_d(i)$ to $X_{d_{k_d}}$	place $X_w(i)$ to $X_{w_{k_w}}$
end	end

5. Find the cluster means in each cluster.

$\overline{X}_{d_1} = \text{mean}\ (X_{d_1})$	$\overline{X}_{w_1} = \text{mean}\ (X_{w_1})$
⋮	⋮
$\overline{X}_{d_{k_d}} = \text{mean}\ (X_{d_{k_d}})$	$\overline{X}_{w_{k_w}} = \text{mean}\ (X_{w_{k_w}})$

6. Perform k-means clustering using the cluster means ($[\overline{X}_{d_1}, \overline{X}_{d_2}, \cdots, \overline{X}_{d_{k_d}}]$ and $[\overline{X}_{w_1}, \overline{X}_{w_2}, \cdots, \overline{X}_{w_{k_w}}]$) as the initial clusters.
   $X_{D_{clusters}} = kmeans\ (X_d, k_d, [\overline{X}_{d_1}, \overline{X}_{d_2}, \cdots, \overline{X}_{d_{k_d}}])$
   $X_{W_{clusters}} = kmeans\ (X_w, k_w, [\overline{X}_{w_1}, \overline{X}_{w_2}, \cdots, \overline{X}_{w_{k_w}}])$
7. Split X_d and X_w according to the cluster outputs ($X_{D_{clusters}}$ and $X_{W_{clusters}}$).

for $i = 1:365$	for $i = 1:365$
if $X_{D_{clusters}}(i) = 1$	if $X_{W_{clusters}}(i) = 1$
place $X_d(i)$ to X_{D_1}	place $X_w(i)$ to X_{W_1}
⋮	⋮
else if $X_{D_{clusters}}(i) = k_d$	else if $X_{W_{clusters}}(i) = k_w$
place $X_d(i)$ to $X_{D_{k_d}}$	place $X_w(i)$ to $X_{W_{k_w}}$
end	end

8. Generate $X_D = (X_{D_1}, X_{D_2}, \cdots, X_{D_{k_d}})$ and $X_W = (X_{W_1}, X_{W_2}, \cdots, X_{W_{k_w}})$.

Algorithm 1: Hybrid hierarchical k-means clustering.

distance metric [42, 43, 44]. On the other hand, hierarchical clustering groups similar data into distinct sets. This study adopts a hybrid approach, which incorporates both hierarchical and k-means clustering, to address potential variability in k-means outputs due to random initialization. Consequently, a hybrid hierarchical k-means clustering algorithm was employed to enhance cluster accuracy and improve the overall prediction model.

Algorithm 1 outlines the hybrid hierarchical k-means clustering employed in this study with eight steps. Step 1 involves reading the solar PV power data in dry and wet seasons (X_d and X_w). Step 2 calculates the optimal number of clusters for both seasons (k_d and k_w) using the silhouette coefficient, determined by MATLAB's *evalclusters* function, with the silhouette coefficient calculated as follows [44]:

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}, \tag{5}$$

where a_i is the average distance from the ith point to the other points in the same cluster as i, and b_i is the minimum average distance from the ith point to the points in a different cluster. The result of this function determines the optimal number of clusters for each season.

Moving on to Step 3, agglomerative hierarchical clustering, a bottom-up approach, was initially employed to cluster the solar PV power data. The *clusterdata* function in MATLAB is utilized for this purpose using the optimal number of clusters for dry and wet seasons calculated in Step 2. This function performs the following actions in this study:

(i) Calculate the distance between objects x_s and x_t using Euclidean distance [44]:

$$d = \sqrt{\sum(x_s - x_t)^2}. \tag{6}$$

(ii) Link pairs of objects that are in close proximity.

(iii) Cut hierarchical trees into clusters.

The outcome of Step 3 is the cluster output of each of the days in dry and wet seasons that are listed in $X_{d_{clusters}}$ and $X_{w_{clusters}}$.

In Step 4, the solar PV power data in dry and wet seasons (X_d and X_w) will be split according to the cluster outputs ($X_{d_{clusters}}$ and $X_{w_{clusters}}$), respectively. This step will provide the solar PV power data for each cluster in the dry season (X_{d_1}, X_{d_1},…$X_{d_{k_d}}$) and wet season (X_{w_1}, X_{w_1},…$X_{w_{k_w}}$) using hierarchical clustering. Step 5 involves calculating the mean of each cluster in the dry season ($\overline{X}_{d_1}$, $\overline{X}_{d_2}$,…$\overline{X}_{d_{k_d}}$) and wet season ($\overline{X}_{w_1}$, $\overline{X}_{w_2}$,…$\overline{X}_{w_{k_w}}$) to be used by the k-means clustering as the initial centers in the next step.

In Step 6, the k-means clustering method was executed using the estimated means as initial centers. The *kmeans* function in MATLAB is used to perform this step with the following actions:

(i) Choose the calculated means of each cluster as initial cluster centers.

(ii) Compute the distance of all observations to each centroid using the squared Euclidean distance metrics [44]:

$$d = \sqrt{\sum(x_s - c)^2}. \tag{7}$$

(iii) Assign observations to a different centroid.

(iv) Compute the average of the observations in each cluster to obtain new centroid locations.

(v) Repeat steps ii to iv until cluster assignments do not change.

The output of Step 6 is to produce the cluster output of each of the days in dry and wet seasons, listed in $X_{D_{clusters}}$ and $X_{W_{clusters}}$, respectively. These outputs will be utilized in Step 7 to further partition the solar PV power data in the dry (X_d) and wet (X_w) seasons. In the last step, Step 7 generates clustered solar PV power data in the dry season ($X_D = (X_{D_1}, X_{D_2}, \cdots, X_{D_{k_d}})$) and wet season ($X_W = (X_{W_1}, X_{W_2}, \cdots, X_{W_{k_w}})$) using the hybrid hierarchical k-means clustering method.

2.3. Phase 3: Linking. Phase 3 aims to identify the most relevant and significant weather factors affecting the solar PV power data using a hybrid GRA-PCC method. The GRA assesses the similarity between reference data and several comparative data, with the gray relational grade (GRG) indicating the correlation scale between them [45, 46, 47, 48, 49]. Additionally, correlation analysis, specifically PCC, further evaluates the relationship between factors and solar PV power. The PCC ranges from −1 to +1, where −1 signifies a perfect negative correlation, and +1 indicates a perfect positive correlation [50, 51]. Through this hybrid GRA-PCC approach, the study aims to pinpoint the most crucial weather factors that significantly influence solar PV power data, enhancing the accuracy of the solar PV power prediction model.

Algorithm 2 outlines the hybrid GRA-PCC algorithm applied in this study. In Step 1, the algorithm reads inputs, including solar PV power data in dry and wet seasons (X_d and X_w) and weather factor data in the dry season (Y_{1_d}, Y_{2_d}, …, Y_{l_d}) and wet season (Y_{1_w}, Y_{2_w},…, Y_{l_w}). Then, in Step 2, the mean of each input (i.e., $\overline{X}_d$, $\overline{X}_w$, ($\overline{Y}_{1_d}$, …,$\overline{Y}_{l_d}$), ($\overline{Y}_{1_w}$, …, $\overline{Y}_{l_w}$)) is calculated. Step 3 involves calculating the GRG using gray relational coefficients, determined as follows [47, 49]:

$$\xi_l(k) = \frac{\Delta_{l\min}(k) + \rho\Delta_{l\max}(k)}{\Delta_l(k) + \rho\Delta_{l\max}(k)}, \tag{8}$$

where $\Delta_l(k) = \|X(k) - Y_l(k)\|$, $\Delta_{l\min}(k) = \min_l \min_k \|X(k) - Y_l(k)\|$, $\Delta_{l\max}(k) = \max_l \max_k \|X(k) - Y_l(k)\|$, l is the number of comparative data, k is the number of data series, and ρ is a coefficient ranging between 0 and 1, set to 0.5 in this study for simplicity. After, the GRG can be calculated as follows [47, 49]:

$$\gamma_l = \frac{1}{n}\sum_{k=1}^{n}\xi_l(k). \tag{9}$$

In Step 3, the GRG for each weather factor in the dry season ($\gamma_d = (\gamma_{1_d}, \ldots, \gamma_{l_d})$) and wet season ($\gamma_w = (\gamma_{1_w}, \ldots, \gamma_{l_w})$) was calculated using Equations (8) and (9), with solar PV power data (X) as the reference data and weather factor data (Y_l) as the comparative data.

In Step 4, weather factors with a grade of 0.6 and above are identified as the significant factors in both dry and wet seasons, listed in Y_{d_s} and Y_{w_s}, respectively. Weather factors with grades below 0.6 are excluded from determining relevant weather factor data in the next step. In Step 5, the PCC of each obtained significant weather data in dry and wet seasons (Y_{d_s} and Y_{w_s}) was calculated as follows [50, 51]:

$$r = \frac{\sum_{i=1}^{k}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{k}\left(\mathrm{X} - \overline{X}\right)^2\sum_{i=1}^{k}\left(Y_i - \overline{Y}\right)^2}}, \tag{10}$$

where $\overline{X}$ and $\overline{Y_s}$ are the mean of solar PV power data and significant weather factor data, respectively, and k is the number of data series. The obtained coefficients in the dry season are denoted as $r_d = (r_{1_d}, \ldots, r_{m_d})$, where m is the number of significant weather data in the dry season. Similarly, in the wet season, the coefficients are represented as $r_w = (r_{1_w}, \ldots, r_{n_w})$,

Input: Solar PV power data in dry and wet seasons (X_d and X_w) and weather data in the dry season (Y_{1_d}, Y_{2_d},…, Y_{l_d}) and wet seasons (Y_{1_w}, Y_{2_w},…, Y_{l_w})

Output: Relevant weather data in the dry season ($Y_D = (Y_{1_D}, Y_{2_D}, \ldots, Y_{o_D})$) and relevant weather data in the wet season ($Y_W = (Y_{1_W}, Y_{2_W}, \ldots, Y_{p_W})$)

1. Read X_d, X_w, (Y_{1_d}, Y_{2_d}, …, Y_{l_d}), and (Y_{1_w}, Y_{2_w}, …, Y_{l_w}).
2. Calculate the mean ($\overline{X}_d$, $\overline{X}_w$, ($\overline{Y}_{1_d}$, …,$\overline{Y}_{l_d}$), ($\overline{Y}_{1_w}$, …, $\overline{Y}_{l_w}$)).
3. Calculate the GRG of each weather factor using Equation (9).
   $\gamma_{l_d} = \frac{1}{n}\sum_{k=1}^{n}\xi_l(k)$
   $\gamma_{l_w} = \frac{1}{n}\sum_{k=1}^{n}\xi_l(k)$
4. Choose the significant weather data ($Y_{d_s}$and Y_{w_s}) based on the grade (γ_d and γ_w).

   for $i = 1{:}l$
   if $\gamma_d(i) \geq 0.6$
   set $Y_d(i)$ to Y_{d_s}
   else
   proceed to the next weather data
   end

   for $i = 1{:}l$
   if $\gamma_w(i) \geq 0.6$
   set $Y_w(i)$ to Y_{w_s}
   else
   proceed to the following weather data
   end
5. Calculate the PCC (r_d and r_w) of the significant weather data (Y_{d_s} and Y_{w_s}) using Equation (10).
   $r_{m_d} = \frac{\sum_{i=1}^{k}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^{k}(X-\overline{X})^2\sum_{i=1}^{k}(Y_i-\overline{Y})^2}}$
   $r_{n_d} = \frac{\sum_{i=1}^{k}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^{k}(X-\overline{X})^2\sum_{i=1}^{k}(Y_i-\overline{Y})^2}}$
6. Choose the relevant factors ($Y_D$and Y_W) among the significant factors ($Y_{d_s}$and Y_{w_s}) based on the coefficient.

   for $i = 1{:}m$
   if $r_d(i) \geq 0.5$ & $r_d(i) \leq -0.5$
   set $Y_{d_s}(i)$ to Y_D
   else
   proceed to the next factor
   end

   for $i = 1{:}n$
   if $r_w(i) \geq 0.5$ & $r_w(i) \leq -0.5$
   set $Y_{w_s}(i)$ to Y_W
   else
   proceed to the next factor
   end
7. Generate $Y_D = (Y_{D_1}, Y_{D_2}, \ldots, Y_{D_o})$ and $Y_W = (Y_{W_1}, Y_{W_2}, \ldots, Y_{W_p})$.

Algorithm 2: Hybrid GRA-PCC method.

where n is the number of significant weather data in the wet season.

Proceeding to Step 6, weather data with coefficients greater than or equal to 0.5 and less than or equal to −0.5 are considered relevant weather data in this study. In the dry season, relevant weather factors are represented as $Y_D = (Y_{D_1}, Y_{D_2}, \ldots, Y_{D_o})$, where o is the number of relevant weather factors. Similarly, in the wet season, relevant weather factors are denoted as $Y_W = (Y_{W_1}, Y_{W_2}, \ldots, Y_{W_p})$, where p is the number of relevant factors. These relevant weather factors serve as the generated outputs of this algorithm in Step 8.

2.4. Phase 4: Classifying. In Phase 4, a classification model links cluster outputs and relevant weather data in dry and wet seasons. Algorithm 3 is employed for this purpose using SVM classification. SVM classifies data by identifying the optimal hyperplane that separates it into two classes [44]. The inputs for this classification model include the cluster outputs for each day in dry and wet seasons ($X_{D_{clusters}}$ and $X_{W_{clusters}}$) and the relevant weather data in dry and wet seasons (Y_D and Y_W). Once these inputs are obtained from Phases 2 and 3, SVM classification is executed using the *fitcsvm* function in MATLAB, which trains an SVM model for one-class and two-class classification on a low-dimensional or moderate-dimensional predictor data set [44]. This function performs the following:

(i) Generate hyperplanes that separate all data points of one class from those of the other class using [44]:

$$f(y) = y'\beta + b, \tag{11}$$

where y is an observation corresponding to matrix Y, the vector β contains the coefficients that define an orthogonal vector to the hyperplane, and b is the bias term.

(ii) Select the best hyperplane, which is the one with the largest margin between the two classes.

(iii) Separate the data points using the best hyperplane.

This function returns an SVM classifier model trained using the relevant weather data in each season. The classification of relevant weather data in dry and wet seasons is based on the cluster outputs of each season. The result of this algorithm comprises the SVM classifier model for dry and wet seasons, which will be utilized in Phase 5 to determine the forecast day's cluster based on its weather data.

Input: Cluster outputs of each day in dry and wet seasons ($X_{D_{clusters}}$ and $X_{W_{clusters}}$), relevant weather data in the dry season ($Y_D = (Y_{1_D}, Y_{2_D}, \ldots, Y_{o_D})$), and relevant weather data in the wet season ($Y_W = (Y_{1_W}, Y_{2_W}, \ldots, Y_{p_W})$)

Output: SVM models for dry and wet seasons (SVM_D and SVM_W)

1. Get X_D, X_W, Y_D, and Y_W.
2. Perform SVM classification.
   $SVM_D = fitcsvm\ (Y_D, X_{D_{clusters}})$
   $SVM_W = fitcsvm\ (Y_W, X_{W_{clusters}})$
3. Generate SVM_D and SVM_W.

Algorithm 3: SVM classification model.

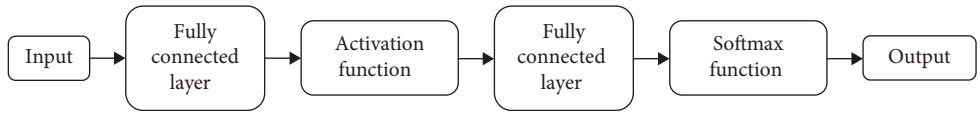


Figure 2: NN structure [44].

Input: Solar PV power data in each cluster in the dry season ($X_D = (X_{D_1}, X_{D_2}, \ldots, X_{D_{k_d}})$), solar PV power data in each cluster in the wet season ($X_W = (X_{W_1}, X_{W_2}, \ldots, X_{W_{k_w}})$) and relevant weather data in the dry season ($Y_D = (Y_{1_D}, Y_{2_D}, \ldots, Y_{o_D})$) and relevant weather data in the wet season ($Y_W = (Y_{1_W}, Y_{2_W}, \ldots, Y_{p_W})$)

Output: NN prediction model (NN_D and NN_W)

1. Get X_D, X_W, Y_D, and Y_W.
2. Perform NN.

   For the dry season,
   for $c = 1$
   $NN_D = fitcnet\ (Y_D, X_{D_1})$
   ⋮
   for $c = k_d$
   $NN_D = fitcnet\ (Y_D, X_{D_{k_d}})$

   For the wet season,
   for $c = 1$
   $NN_W = fitcnet\ (Y_W, X_{W_1})$
   ⋮
   for $c = k_w$
   $NN_W = fitcnet\ (Y_W, X_{W_{k_w}})$

3. Generate (NN_D) and (NN_W).

Algorithm 4: NN prediction model.

2.5. Phase 5: Predicting. In Phase 5, the prediction phase focuses on forecasting solar PV power using NN. The NN, a machine learning approach, employs interconnected nodes or neurons in a layered structure. Algorithm 4 outlines the development of the NN prediction model. The inputs for this model include solar PV power data in each cluster in the dry and wet seasons (X_D and X_W) and relevant weather data in dry and seasons (Y_D and Y_W). Once these inputs are collected using Algorithms 1 and 2, MATLAB's *fitcnet* function is utilized to develop the NN prediction model with the layer structure shown in Figure 2. This function performs using the following actions:

(i) The predictor data is inputted.

(ii) The activation function is used for every first fully connected layer using the rectified linear unit function as follows [44]:

$$f(y) = \begin{cases} y, & y \geq 0 \\ 0, & y < 0 \end{cases}, \tag{12}$$

where y is an observation corresponding to matrix Y.

(iii) The activation function is applied to the final fully connected layer using the softmax function [44]:

$$f(y_i) = \frac{\exp(y_i)}{\sum_{j=1}^{z} y_j}, \tag{13}$$

where y_i is an observation corresponding to matrix Y, and z is the total number of response variables.

(iv) The predicted value is generated.

The input data in the structure shown in Figure 2 are the relevant weather data in dry and wet seasons. The first fully connected layer has 10 outputs with the weights and biases set by the function. Then, the activation function using the rectified linear unit function, which performs a threshold operation on each element of the relevant weather data in which any value less than zero is set to zero, shown in Equation (12), was used in these first fully connected layers. The number of solar PV power data in each dry and wet

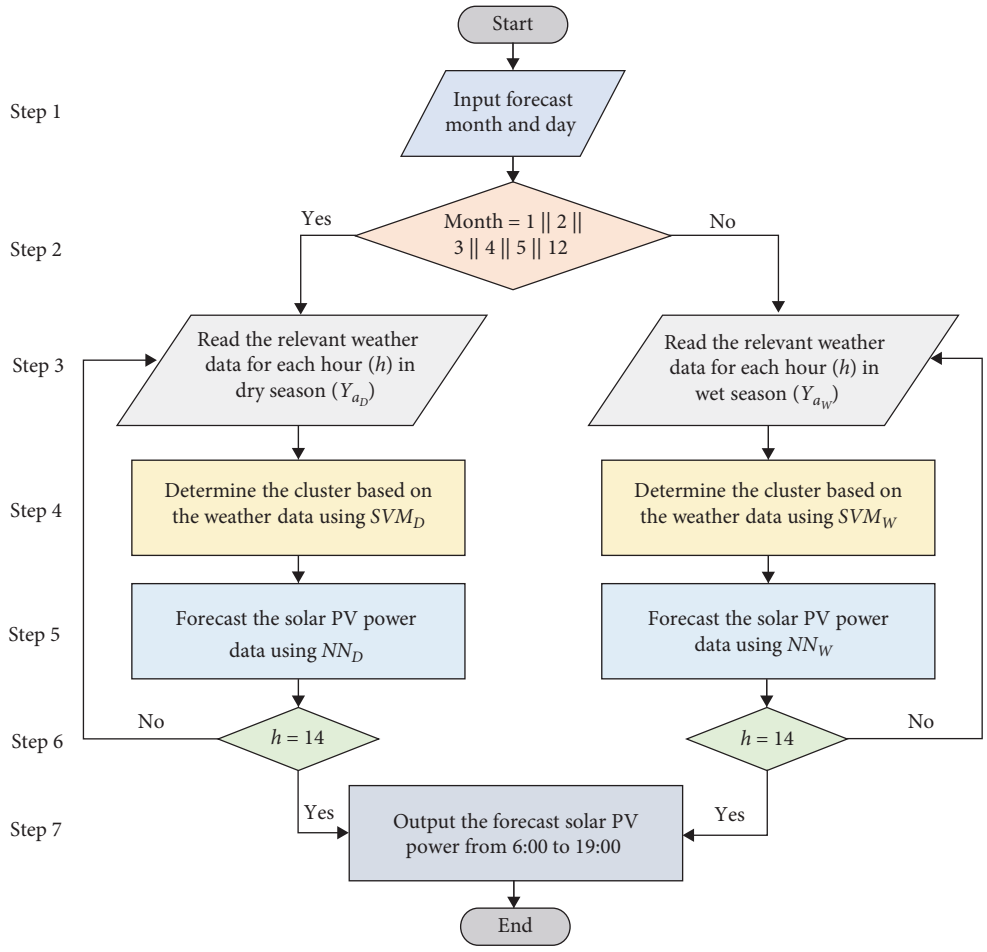


FIGURE 3: Flowchart of the predicting phase of the solar PV prediction model.

season cluster is the output in the final fully connected layer with the weights and biases set by the function. Then, the softmax function in Equation (13) was used in these final fully connected layers. Lastly, the output is the layer that corresponds to the predicted solar PV power. This *fitcnet* function developed a model that trains the relevant weather data in each season using the solar PV power data in each season's cluster. The output of this function includes NN prediction models for dry and wet seasons (NN_D and NN_W) to be used to forecast solar PV power data.

Figure 3 illustrates the flowchart for forecasting solar PV power in the prediction phase. There are seven steps in this prediction procedure.

(i) In Step 1, the inputs include the month and day to be forecast.

(ii) Step 2 decides the season of the entered forecast month. If the month corresponds to 1, 2, 3, 4, 5, or 12 (i.e., January, February, March, April, May, and December), the training data and the models formulated using the training data from Phase 1 to Phase 5 in the dry season are utilized. Alternatively, if the month corresponds to 6, 7, 8, 9, 10, or 11 (i.e., June, July, August, September, October, and November), the training data and the models formulated using the training data from Phase 1 to Phase 5 in the wet season are employed.

(iii) In Step 3, the forecast month and day weather data were collected using the testing data. The relevant weather data information for the forecast day is determined based on the identified relevant weather factors of the season (Y_{a_D} and Y_{a_W}) in Phase 3. The system reads the relevant weather data from hour 1 (i.e., 6:00) from the testing data.

(iv) Step 4 determines the cluster based on the weather factors gathered in Step 3. The SVM classification model of the season, established in Algorithm 3 using the training data, is used to determine the cluster based on that particular hour's relevant weather data information. The relevant weather data information for each hour serves as the input to the SVM classification model, which provides the cluster of the forecast day.

(v) Step 5 is to forecast the solar PV power data for the particular hour of the forecast day. Utilizing the cluster output of the SVM classification model for

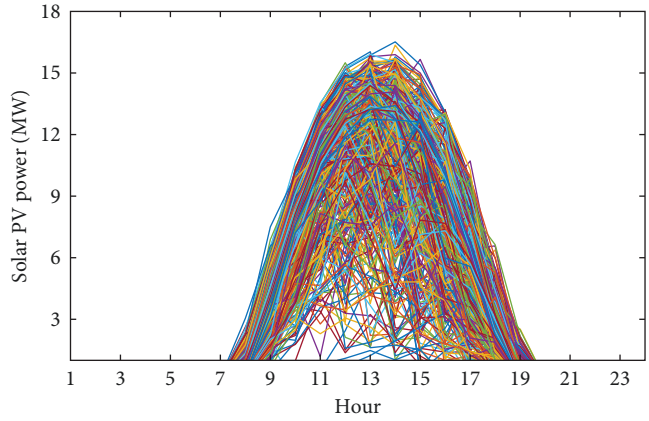


Figure 4: Training data for formulating a multiphase solar PV prediction model.

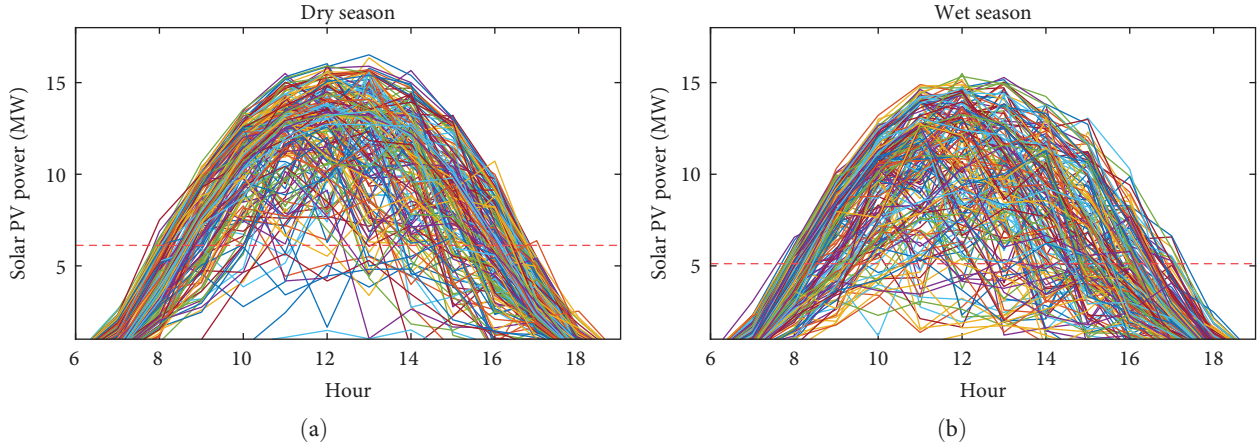


Figure 5: Grouped solar PV power data according to season: (a) dry season; (b) wet season.

that hour in Step 4, the solar PV power is forecast using the NN prediction model developed using the training data in Algorithm 4.

(vi) In Step 6, the simulation model decides if the simulation continues to the next step or back to Step 3. This process continues until the hour reaches 14 (i.e., 19:00).

(vii) Lastly, the output of this model is the solar PV power data from 6:00 to 19:00 of the forecast month and day will be generated in Step 7.

3. Numerical Example

This section illustrates the effectiveness of the proposed multiphase solar PV prediction model through a numerical example. The solar PV prediction model was formulated based on the gathered data, and the solar PV power was predicted using the developed prediction model. The prediction model's performance was also evaluated for its efficiency.

3.1. Description of Data. Solar PV power and various weather factors, including temperature, wind speed, precipitation, humidity, visibility, pressure, cloud cover, heat index, dewpoint, windchill, UV index, and weather description, were utilized in this study. The training dataset comprised hourly solar PV power data and weather data from a solar PV power plant with 18 MW rated power capacity, collected throughout the day (i.e., 0:00 to 23:00) from January 1, 2019, to December 31, 2019. Figure 4 illustrates the solar PV power data for the 18 MW solar PV plant, indicating power generation only from 6:00 to 19:00, thus serving as the training data for the formulation of the solar PV power prediction model.

3.2. Grouping Result. In Phase 1 of the solar PV power prediction model, the data were categorized into two seasons in the Philippines (i.e., dry and wet). The training data was split accordingly into dry and wet seasons. Figure 5 illustrates the grouped solar PV power data for both seasons, with Figure 5(a) representing the dry season and Figure 5(b) the wet season. The higher solar PV power during the dry season is evident from their respective means.

3.3. Hybrid Hierarchical k-Means Clustering Result. In Phase 2, the solar PV power data in each season, shown in Figure 5, were partitioned into clusters. The silhouette coefficients of dry and wet seasons in Figure 6 indicate that the optimal number of clusters for both seasons is two, as it yields the

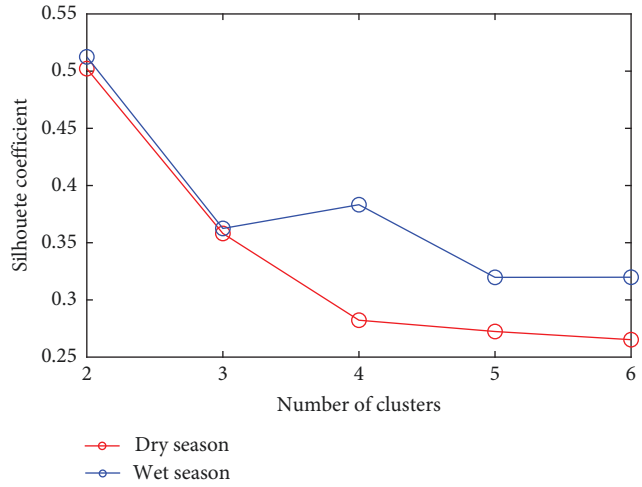


FIGURE 6: Silhouette coefficients for dry and wet seasons.

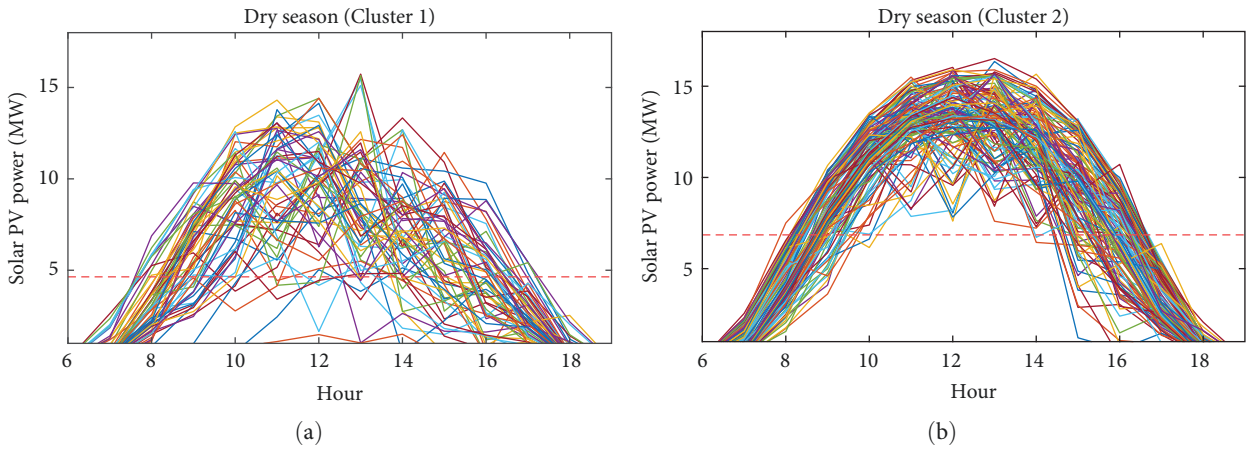


FIGURE 7: Solar PV power data in each cluster in the dry season: (a) Cluster 1; (b) Cluster 2.

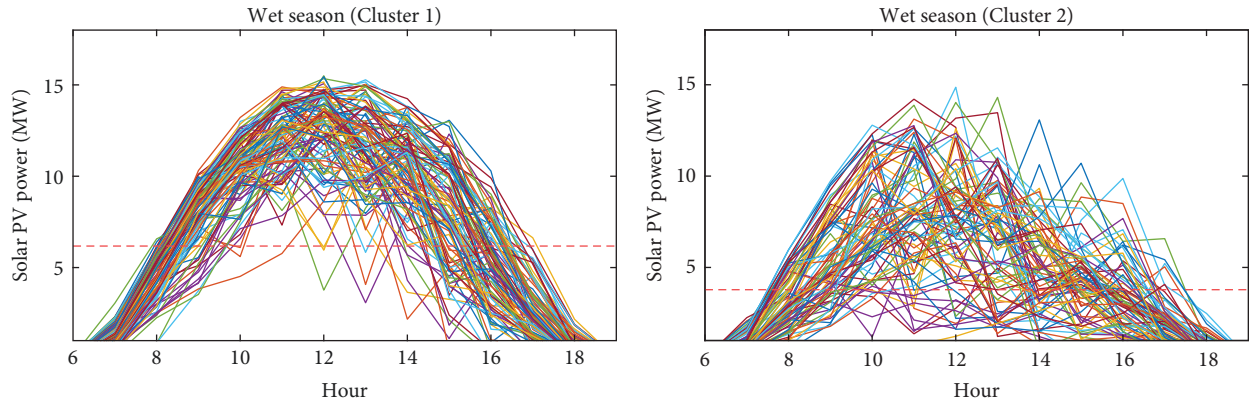
highest silhouette coefficient. Hence, the solar PV power data for both seasons will be divided into two clusters.

Figures 7 and 8 display the solar PV power for each cluster in dry and wet seasons, respectively. In Figure 7(a), the solar PV power Cluster 1 exhibits low values during the dry season, while Figure 7(b) depicts solar PV power Cluster 2 with high solar PV power values in the dry season. For wet season data, Figure 8(a) illustrates high solar PV power values for Cluster 1, while Figure 8(b) shows low values for Cluster 2 in the wet season.

3.4. Hybrid GRA-PCC Method Result. In Phase 3, relevant weather factors were determined from the 11 studied weather factors (i.e., temperature, wind speed, precipitation, humidity, pressure, cloud cover, heat index, dewpoint, windchill, UV index, and weather description) using the hybrid GRA-PCC method. Table 2 lists these GRGs in dry and wet seasons, considering those with grades 0.6 and above as significant. The significant factors in the dry season (Y_{d_s}) include wind speed, humidity, pressure, cloud cover, heat index, dewpoint, and UV index. While temperature, wind speed, cloud cover, heat index, wind chill, and UV index are considered significant in the wet season (Y_{w_s}).

Next, the PCC of these significant factors listed in Table 3 was computed to identify the relevant factors affecting the solar PV power data. Those with a coefficient ≥ 0.5 or ≤ -0.5 are considered the relevant weather factors. Three factors—humidity, cloud cover, and UV index—are deemed relevant in the dry season. Meanwhile, five factors—temperature, cloud cover, windchill, and UV index—are identified as relevant in the wet season.

3.5. Multiphase Solar PV Prediction Model Result. The multiphase solar PV prediction model is validated with test data from January 1, 2020, to December 31, 2020, including solar PV power data and weather data. The model's effectiveness was demonstrated using the first week of each month in 2020, compared to conventional methods such as NN and ARMAX. The NN method was considered as it is also used in the prediction phase of the proposed model. In this regard, the



(a) (b)

Figure 8: Solar PV power data in each cluster in the wet season: (a) Cluster 1; (b) Cluster 2.

Table 2: GRG of each weather factor in the dry and wet seasons.

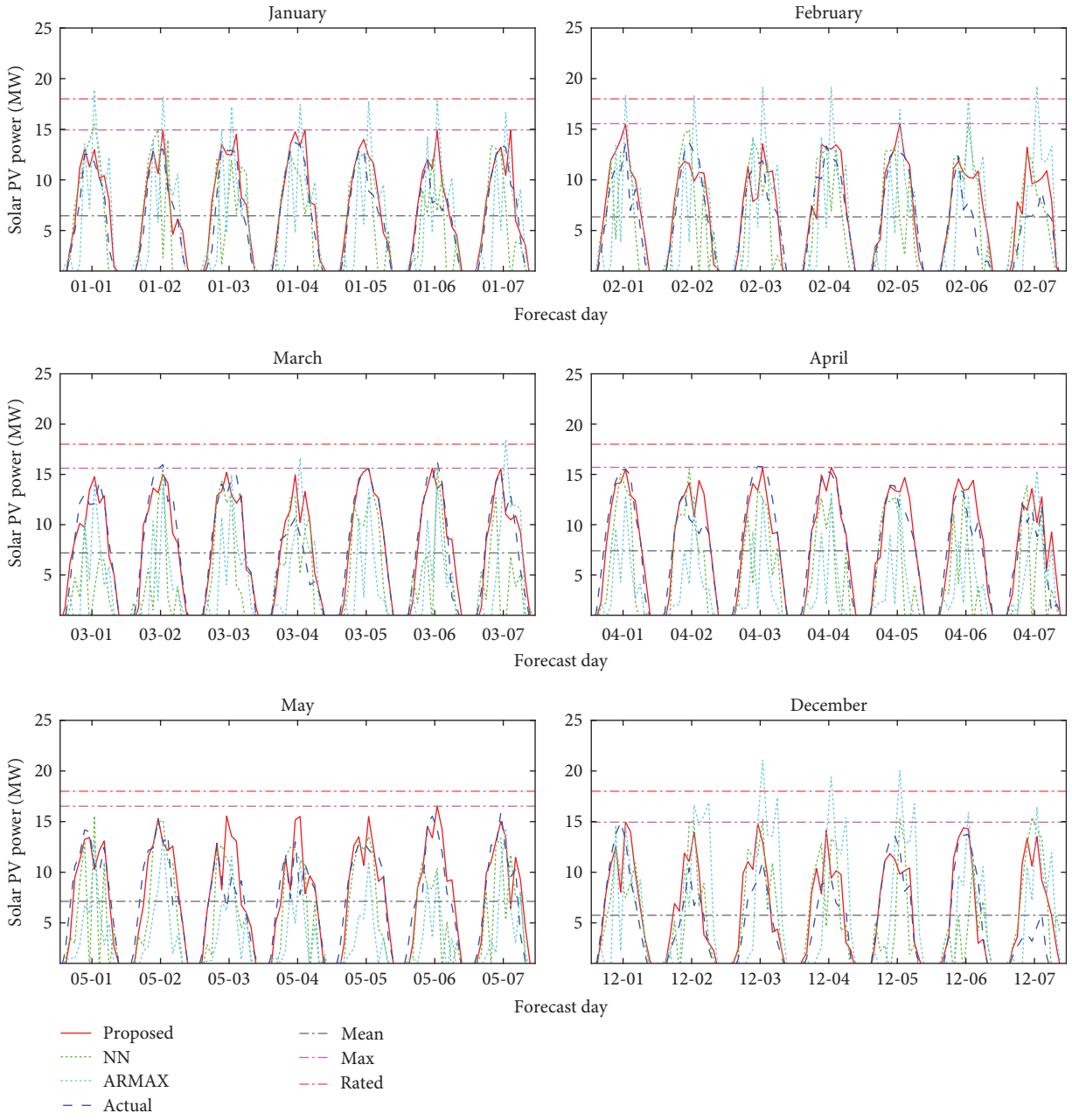
Weather factor (Y_l)	GRG	
	Dry (γ_d)	Wet (γ_w)
Temperature	0.5848	0.7086
Wind speed	0.6781	0.6248
Precipitation	0.5242	0.5756
Humidity	0.7168	0.5963
Pressure	0.6531	0.5332
Cloud cover	0.6799	0.7165
Heat index	0.6232	0.6577
Dewpoint	0.6215	0.5496
Windchill	0.5848	0.7086
UV index	0.7286	0.6790
Weather description	0.5380	0.5387

Table 3: The PCC of each significant factor in the dry and wet seasons.

Dry season		Wet season	
Significant weather data (Y_{d_s})	PCC (r_d)	Significant weather data (Y_{w_s})	PCC (r_w)
Wind speed	−0.2041	Temperature	0.6208
Humidity	−0.7359	Wind speed	−0.2206
Pressure	0.0688	Cloud cover	−0.6456
Cloud cover	−0.5795	Heat index	0.5814
Heat index	0.3458	Windchill	0.6028
Dewpoint	−0.4040	UV index	0.6863
UV index	0.7311	—	—

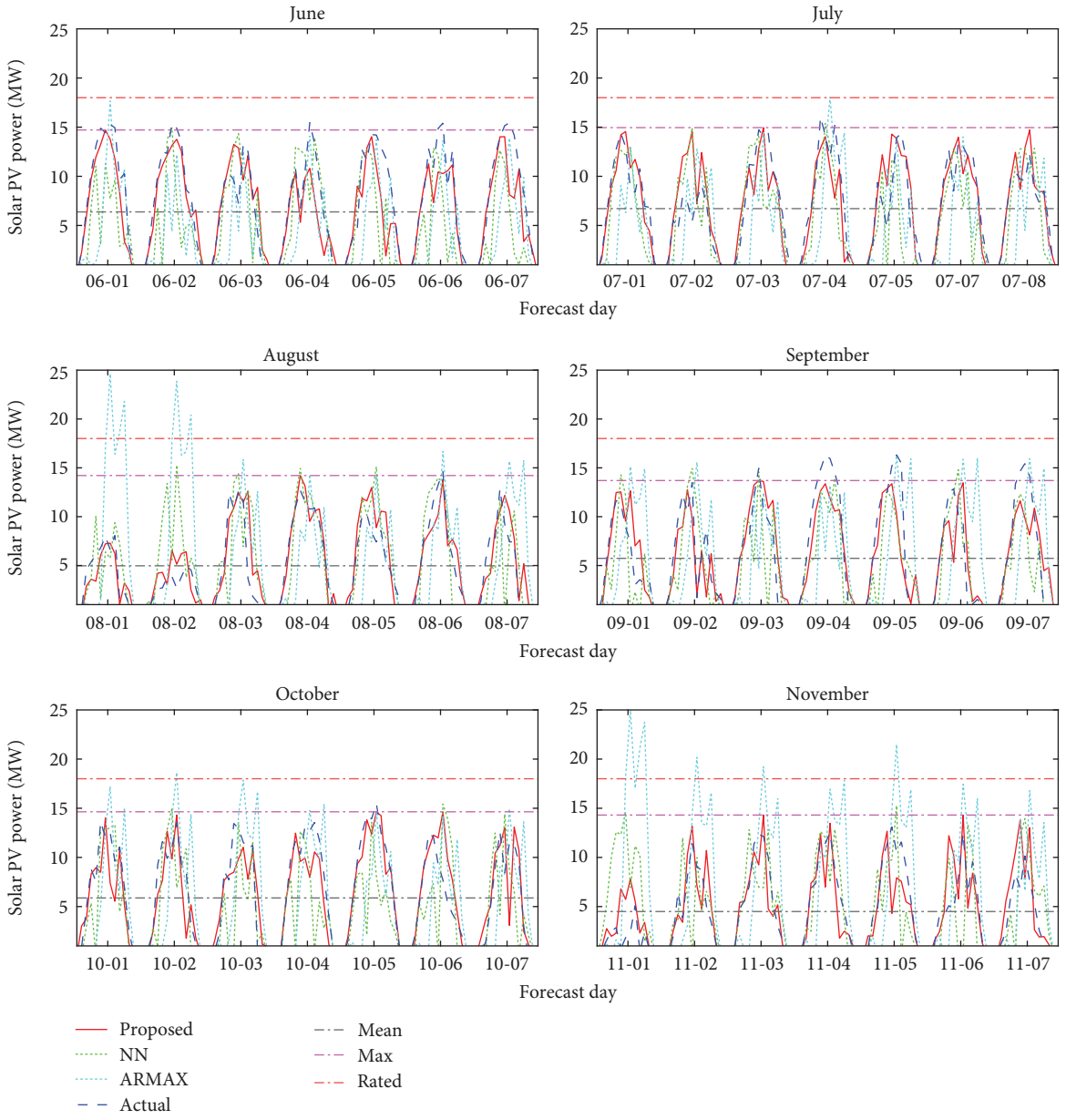
result of the proposed model considering the phases can be compared to the result of the conventional method, NN, without consideration of the phases. The ARMAX method was also considered to compare the result further with conventional methods used in forecasting other than the NN method. The proposed model's parameters include historical solar PV power and weather data with hyperparameters focusing on each cluster's significant weather data of the forecast month and day. In contrast, conventional methods (NN and ARMAX) utilized all the weather data and weather factors of the forecast month and day as the hyperparameters.

Figure 9 shows the forecast solar PV power profiles for every first week of all the months of 2020, comparing the proposed prediction model with conventional methods and actual data. Figures 9(a) and 9(b) show the solar PV power for the dry and wet seasons, respectively, which illustrates



(a)

Figure 9: Continued.



(b)

FIGURE 9: Solar PV power data on each forecast day of every month: (a) dry season; (b) wet season.

higher solar PV power in the dry season. Maximum and mean values displayed in the figures confirm this trend. Table 4 summarizes both seasons' average actual and forecast solar PV power. In the dry season, the average actual solar PV power is 6.2057 MW; the proposed method indicates 6.7273 MW, while conventional NN and ARMAX methods forecast 4.8110 and 4.5248 MW, respectively. In the wet season, the average actual solar PV power is 5.8594 MW; the proposed method predicts 5.6955 MW, while NN and ARMAX indicated 4.8016 and 5.2065 MW, respectively. The proposed prediction model aligns closely with the actual mean values, and it is noted that ARMAX forecasts in the wet season exceed the rated capacity of the solar PV system. Thus, understanding the seasonal trend is crucial for accurately predicting solar PV power.

In addition, the one-sample t-test, conducted at a significance level of 0.05, was employed to assess the significance of differences between the means of the actual and forecast data. If the p-value is less than or equal to 0.05, it is concluded that there is a substantial difference between the means of the forecast and actual values. Table 5 presents the p-values for both seasons using a one-sample t-test. The proposed method's p-values exceed 0.05 in both seasons, indicating an insignificant difference between forecast and actual means. In contrast, NN exhibits a significant difference. In the dry season, ARMAX has a p-value less than 0.05;

Table 4: Mean solar PV power values for each month in dry and wet seasons.

Month	Dry season				Month	Wet season			
	Mean solar PV power values (MW)					Mean solar PV power values (MW)			
	Actual	Proposed	NN	ARMAX		Actual	Proposed	NN	ARMAX
January	5.7046	6.4551	4.9882	5.2851	June	7.3321	6.3804	4.6774	4.0503
February	5.5187	6.3456	5.1469	5.4565	July	6.6084	6.7144	5.2991	4.4593
March	7.1527	7.1942	4.1472	4.1058	August	4.6507	4.9711	4.9573	5.7203
April	7.2651	7.4073	4.6634	3.3497	September	5.9912	5.7354	4.6516	5.2265
May	6.9939	7.1271	5.0682	3.2320	October	6.2057	5.8767	4.4982	5.2210
December	4.5993	5.7521	4.8522	5.7198	November	4.3683	4.4950	4.7257	6.5614
Average	6.2057	6.7136	4.8110	4.5248	Average	5.8594	5.6955	4.8016	5.2065

Table 5: The p-value of the dry and wet seasons using a one-sample t-test.

Season	p-Value		
	Proposed	NN	ARMAX
Dry season	0.107	≤0.001	0.014
Wet season	0.652	≤0.001	0.134

in the wet season, it is more significant than 0.05, indicating a contradiction due to the neglect of seasonal effects. The proposed multiphase prediction model outperforms conventional methods, providing more accurate solar PV power, which is crucial for precise power estimation.

Moreover, this study conducted hourly comparisons of forecast and actual solar PV power values, analyzing variations and patterns based on specific weather conditions. A paired t-test was employed, examining the significance of the output of the proposed prediction model. A significance level of 0.05 was used, with a p-value ≤ 0.05 indicating a substantial difference between the forecast and the actual values. Table 6 lists the p-values from the paired t-tests for the proposed and conventional methods. In the dry season, the proposed method shows eight out of 14 forecast hours with p-values > 0.05, while NN and ARMAX have four and two, respectively. In the wet season, 13 out of 14 forecast hours using the proposed method and seven and five using NN and ARMAX, respectively, have p-values > 0.05. The proposed multiphase prediction model demonstrates superior accuracy in hourly solar PV power forecasting, which is crucial for effectively monitoring daily fluctuations.

Furthermore, various forecasting measures, including mean absolute error (MAE), mean squared error (MSE), RMSE, normalized root mean square error (nRMSE), and mean relative error (MRE), were employed to assess the effectiveness of the proposed multiphase solar PV prediction model. These measures calculate accuracy errors regarding solar PV power (MW) for MAE, MSE, and RMSE, and as percentage (%) for nRMSE and MRE. The actual and forecast solar PV power values were considered in computing these measures given as follows [10]:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|X_f - X_a\right|, \tag{14}$$

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(X_f - X_a\right)^2, \tag{15}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_f - X_a\right)^2}, \tag{16}$$

$$\mathrm{nRMSE} = \left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_f - X_a\right)^2}\right) \times \frac{100}{\max(X_a)}, \tag{17}$$

$$\mathrm{MRE} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|X_f - X_a\right|}{X_T} \times 100, \tag{18}$$

where X_f is the forecast solar PV power, X_a is the actual solar PV power, X_T is the total power capacity of the PV system (i.e., 18 MW), and N is the number of forecast hours equal to 14.

Table 7 presents the performance metrics results using the proposed and the conventional methods, highlighting the forecast day with the best error metrics for each month. The complete metrics for all the forecast days are listed in Table 8. Table 7 shows consistently lower values for most metrics with the proposed prediction model than the conventional methods. Notably, the proposed prediction model achieved the lowest performance metrics on March 5, 2020, which is a day in the dry season, with the following results: MAE of 0.408 MW, MSE of 460.51 MW, RMSE of 0.679 MW, nRMSE of 4.345%, and MRE of 2.266%. In comparison, NN recorded MAE of 2.458 MW, MSE of 10,311.59 MW, RMSE of 3.211 MW, nRMSE of 20.561% and MRE of 13.653%, while ARMAX showed MAE of 4.772 MW, MSE of 40,943.09 MW, RMSE of 202.344 MW, nRMSE of 1,425.378% and MRE of 26.510%. To further illustrate, the MRE results were compared, which measures the uncertainty compared to the total power capacity. The MRE result of the proposed prediction model is 2.266%, while that of NN is 13.653%, and that of ARMAX is 26.510%. These results state that the uncertainties using the proposed prediction model are six times lower than those of the NN method and 12 times lower than those of the ARMAX method. These results show that a significantly improved solar PV power prediction was achieved using the proposed prediction model compared to conventional methods.

Table 6: The p-value of each hour in the dry and wet seasons using a paired t-test.

Hour	Dry season			Hour	Wet season		
	p-Value				p-Value		
	Proposed	NN	ARMAX		Proposed	NN	ARMAX
6:00	0.907	0.627	0.156	6:00	0.048	0.440	0.910
7:00	0.116	0.004	0.004	7:00	0.242	≤0.001	≤0.001
8:00	0.042	0.001	≤0.001	8:00	0.418	≤0.001	≤0.001
9:00	0.176	≤0.001	≤0.001	9:00	0.203	0.003	≤0.001
10:00	0.069	0.102	≤0.001	10:00	0.775	0.108	≤0.001
11:00	0.210	0.049	≤0.001	11:00	0.799	0.631	≤0.001
12:00	0.150	0.031	≤0.001	12:00	0.835	0.821	0.049
13:00	0.001	0.115	≤0.001	13:00	0.308	0.009	≤0.001
14:00	0.001	0.009	0.038	14:00	0.089	0.039	0.477
15:00	0.054	≤0.001	0.007	15:00	0.332	0.057	0.673
16:00	0.002	≤0.001	0.018	16:00	0.593	0.015	≤0.001
17:00	0.003	≤0.001	0.012	17:00	0.730	0.154	0.038
18:00	0.046	0.145	0.129	18:00	0.717	≤0.001	0.348
19:00	0.258	0.078	0.019	19:00	0.154	0.995	0.466

Table 7: Performance metrics result from the proposed and conventional methods.

Day	MAE (MW)			MSE (MW)			RMSE (MW)		
	Proposed	NN	ARMAX	Proposed	NN	ARMAX	Proposed	NN	ARMAX
01/04	0.743	1.494	2.768	1,570.67	4,869.31	18,102.46	1.253	2.207	134.545
02/05	0.808	2.206	2.638	1,322.42	13,167.10	15,236.03	1.150	3.629	123.434
03/05	0.408	2.458	4.772	460.51	10,311.59	40,943.09	0.679	3.211	202.344
04/04	0.849	3.330	4.954	1,238.99	22,834.13	41,888.38	1.113	4.779	204.667
05/01	1.014	3.851	4.455	1,819.55	32,578.83	31,818.76	1.349	5.708	178.378
06/02	1.060	3.701	4.276	1,849.28	27,802.01	33,144.12	1.360	5.273	182.055
07/08	0.988	2.382	2.859	2,018.74	9,037.07	17,568.30	1.421	3.006	132.545
08/04	0.834	1.494	3.259	1,198.66	5,098.03	19,015.98	1.095	2.258	137.898
09/03	1.450	2.567	3.311	3,848.43	13,849.99	25,014.82	1.962	3.722	158.161
10/05	0.828	3.281	3.620	1,469.35	23,024.94	25,900.17	1.212	4.798	160.935
11/04	1.184	1.783	3.576	3,504.57	6,134.04	26,624.45	1.872	2.477	163.170
12/07	1.439	1.937	3.959	5,651.07	7,540.64	30,580.123	2.377	2.746	174.872

Day	nRMSE (%)			MRE (%)		
	Proposed	NN	ARMAX	Proposed	NN	ARMAX
01/04	9.152	16.114	1,073.347	4.129	8.300	15.378
02/05	8.874	28.000	905.834	4.489	12.256	14.656
03/05	4.345	20.561	1,425.378	2.266	13.653	26.510
04/04	7.254	31.139	1,316.786	4.716	18.500	27.521
05/01	9.503	40.209	1,256.600	5.632	21.394	24.752
06/02	8.849	34.312	1,192.195	5.891	20.563	23.753
07/08	11.604	24.552	1,020.823	5.488	13.235	15.882
08/04	8.626	17.789	1,705.275	4.636	8.299	18.104
09/03	13.106	24.863	1,356.377	8.056	14.258	18.393
10/05	7.832	31.003	1179.989	4.598	18.225	20.112
11/04	14.961	19.793	3193.303	6.578	9.908	19.868
12/07	17.575	20.301	1172.409	7.992	10.763	21.996

In addition, the RMSE results of the proposed prediction model were compared with the best RMSE results obtained in the previous studies listed in Table 1 to evaluate the prediction model's performance further. To compare these values, the percentages between the RMSE value and the rated capacity of the proposed prediction model and the previous studies in Table 1 were obtained, as shown in Tables 9 and 10, respectively [15, 16, 17, 18, 19, 20, 22]. The smallest

Table 8: Detailed performance metrics result from the proposed and conventional method.

Day	MAE (MW)			MSE (MW)			RMSE (MW)		
	Proposed	NN	ARMAX	Proposed	NN	ARMAX	Proposed	NN	ARMAX
01/01	1.080	1.346	2.658	3,043.59	3,308.78	16,352.21	1.745	1.819	127.876
01/02	1.013	2.932	2.464	2,035.45	18,617.72	12,714.13	1.427	4.315	112.757
01/03	1.141	3.153	2.407	2,973.49	20,303.96	11,418.28	1.724	4.506	106.856
01/04	0.743	1.494	2.768	1,570.67	4,869.31	18,102.46	1.253	2.207	134.545
01/05	1.038	1.786	2.934	2,198.27	7,436.00	20,370.88	1.483	2.727	142.727
01/06	1.343	2.639	2.891	4,975.89	13,600.73	17,977.10	2.231	3.688	134.079
01/07	1.244	1.923	2.755	3,820.84	11,376.89	17,556.96	1.955	3.373	132.503
02/01	1.488	2.309	2.797	3,541.50	9,768.10	15,344.56	1.882	3.125	123.873
02/02	1.255	2.429	2.797	2,424.86	13,836.61	15,344.56	1.557	3.720	123.873
02/03	1.444	2.441	3.144	3,764.33	9,547.91	17,958.88	1.940	3.090	134.011
02/04	1.070	2.005	3.144	2,876.51	6,660.63	17,958.88	1.696	2.581	134.011
02/05	0.808	2.206	2.638	1,322.42	13,167.10	15,236.03	1.150	3.629	123.434
02/06	2.076	2.188	3.684	9,982.69	11,451.42	30,053.14	3.160	3.384	173.358
02/07	1.959	2.045	3.167	8,014.55	8,044.81	21,164.69	2.831	2.836	145.481
03/01	1.224	4.606	3.928	2,605.08	34,424.82	28,408.38	1.614	5.867	168.548
03/02	1.210	4.833	4.596	2,508.43	40,832.15	38,984.88	1.584	6.390	197.446
03/03	0.916	2.970	3.737	1,465.41	20,493.07	27,564.60	1.211	4.527	166.026
03/04	1.784	2.310	2.904	7,767.15	8,529.29	15,860.87	2.787	2.920	125.940
03/05	0.408	2.458	4.772	460.51	10,311.59	40,943.09	0.679	3.211	202.344
03/06	1.400	3.358	3.811	4,172.33	27,415.89	27,906.59	2.043	5.236	167.053
03/07	1.019	4.036	3.884	2,169.85	35,811.20	29,783.44	1.473	5.984	172.579
04/01	0.997	3.033	4.607	1,813.03	17,983.06	36,624.20	1.346	4.241	191.375
04/02	1.210	2.377	4.031	4,030.59	11,743.18	25,264.03	2.008	3.427	158.947
04/03	1.077	2.992	5.480	2,326.03	1,7150.96	49,230.13	1.525	4.141	221.879
04/04	0.849	3.330	4.954	1,238.99	22,834.13	41,888.38	1.113	4.779	204.667
04/05	0.936	2.550	4.499	1,956.20	17,406.72	33,475.37	1.399	4.172	182.963
04/06	1.303	3.516	3.896	3,926.84	2,2671.59	27,523.37	1.982	4.761	165.902
04/07	1.502	2.522	2.895	6,066.03	16,397.15	15,787.70	2.463	4.049	125.649
05/01	1.014	3.851	4.455	1,819.55	32,578.83	31,818.76	1.349	5.708	178.378
05/02	1.188	2.519	4.096	4,142.98	11,724.16	26,536.12	2.035	3.424	162.899
05/03	1.809	2.195	2.907	9,811.38	9,315.07	18,090.12	3.132	3.052	134.500
05/04	1.495	2.772	4.145	6,059.60	15,469.83	28,551.54	2.462	3.933	168.972
05/05	1.295	2.106	4.732	2,882.82	9,723.00	34,800.48	1.698	3.118	186.549
05/06	1.482	3.530	4.367	3,190.38	23,022.39	29,927.12	1.786	4.798	172.995
05/07	1.487	2.578	3.874	4,359.76	16,575.79	27,209.28	2.088	4.071	164.952
06/01	1.367	4.158	3.825	4,867.43	28,166.72	30,947.55	2.206	5.307	175.919
06/02	1.060	3.701	4.276	1,849.28	27,802.01	33,144.12	1.360	5.273	182.055
06/03	1.748	2.591	3.043	6,253.34	11,710.01	18,765.47	2.501	3.422	136.987
06/04	1.759	2.249	2.966	7,088.62	10,499.17	16,350.92	2.662	3.240	127.871
06/05	2.432	3.522	3.223	14,008.18	25,478.09	19,784.81	3.743	5.048	140.658
06/06	1.485	3.738	3.819	4,846.62	26,713.33	28,940.50	2.202	5.168	170.119
06/07	1.804	4.397	3.945	9,012.98	33,843.59	25,714.63	3.002	5.818	160.358
07/01	1.170	2.256	4.048	2,338.15	9,206.03	28,940.28	1.529	3.034	170.118
07/02	2.186	2.603	2.363	11,228.64	14,017.93	11,878.64	3.351	3.744	108.989
07/03	1.576	2.119	3.279	5,087.49	9,506.48	22,302.54	2.256	3.083	149.340
07/04	1.765	2.580	3.729	7,601.52	12,219.88	27,037.45	2.757	3.496	164.431
07/05	1.957	3.542	3.569	7,782.74	26,151.27	23,990.79	2.790	5.114	154.890
07/07	1.180	2.741	3.875	2,905.36	16,442.55	29,399.93	1.705	4.055	171.464
07/08	0.988	2.382	2.859	2,018.74	9,037.07	17,568.30	1.421	3.006	132.545
08/01	0.997	1.771	6.867	1,562.85	6,437.47	90,188.33	1.250	2.537	300.314
08/02	1.297	3.293	5.928	2,446.58	25,199.40	85,680.81	1.564	5.020	292.713
08/03	1.446	3.404	3.729	7,478.34	26,092.08	27,172.40	2.735	5.108	164.841

Table 8: Continued.

Day	MAE (MW)			MSE (MW)			RMSE (MW)		
	Proposed	NN	ARMAX	Proposed	NN	ARMAX	Proposed	NN	ARMAX
08/04	0.834	1.494	3.259	1,198.66	5,098.03	19,015.98	1.095	2.258	137.898
08/05	1.635	1.711	2.511	4,907.65	6,839.92	15,365.67	2.215	2.615	123.958
08/06	1.092	1.890	2.769	2,375.09	5,688.51	15,839.77	1.541	2.385	125.856
08/07	1.583	2.241	4.944	4,776.73	11,410.03	42,871.58	2.186	3.378	207.055
09/01	1.513	2.514	4.626	4,869.18	11,493.00	38,714.68	2.207	3.390	196.760
09/02	1.732	2.605	4.336	6,470.96	14,411.84	36,406.24	2.544	3.796	190.804
09/03	1.450	2.567	3.311	3,848.43	13,849.99	25,014.82	1.962	3.722	158.161
09/04	1.920	2.606	4.837	5,378.48	12,029.53	39,227.76	2.319	3.468	198.060
09/05	1.868	3.364	5.052	11,499.93	21,892.75	47,862.42	3.391	4.679	218.775
09/06	1.668	1.823	5.284	7,781.37	7,748.50	49,424.02	2.790	2.784	222.315
09/07	2.259	2.320	4.329	9,189.51	13,341.11	36,313.18	3.031	3.653	190.560
10/01	1.786	2.077	3.503	6,135.17	9,867.50	22,747.51	2.477	3.141	150.823
10/02	1.695	2.216	3.747	6,281.61	11,000.94	29,045.81	2.506	3.317	170.428
10/03	1.672	2.710	3.873	5,879.34	21,361.79	35,302.21	2.425	4.622	187.889
10/04	1.615	3.428	3.162	4,726.45	28,107.19	21,988.56	2.174	5.302	148.285
10/05	0.828	3.281	3.620	1,469.35	23,024.94	25,900.17	1.212	4.798	160.935
10/06	1.621	3.964	3.663	6,542.18	26,433.36	26,488.29	2.558	5.141	162.752
10/07	1.751	3.124	2.898	7,786.71	18,360.85	18,451.90	2.790	4.285	135.838
11/01	1.529	5.027	7.371	6,302.22	44,506.91	131,781.67	2.510	6.671	363.017
11/02	1.492	2.532	4.165	4,848.50	1,6875.78	33,958.82	2.202	4.108	184.279
11/03	1.198	2.219	3.720	3,547.80	9,022.73	26,485.68	1.884	3.004	162.744
11/04	1.184	1.783	3.576	3,504.57	6,134.04	26,624.45	1.872	2.477	163.170
11/05	1.793	2.939	3.841	8,671.33	17,745.34	28,903.97	2.945	4.213	170.012
11/06	1.649	3.550	3.008	6,965.81	25,329.71	16,907.41	2.639	5.033	130.028
11/07	1.969	2.629	4.261	7,917.41	12,543.19	32,085.73	2.814	3.542	179.125
12/01	1.810	4.917	3.148	7,658.13	44,728.40	26,539.631	2.767	6.688	162.910
12/02	2.326	3.454	4.306	10,917.09	22,826.63	370,18.826	3.304	4.778	192.403
12/03	1.818	4.108	4.565	6,824.79	31,256.52	37,976.105	2.612	5.591	194.875
12/04	2.580	3.697	3.798	130,88.61	25,882.62	24,970.734	3.618	5.087	158.021
12/05	3.804	4.909	4.549	29,263.15	41,158.69	39,574.402	5.410	6.416	198.933
12/06	3.842	4.995	2.833	35,960.91	53,043.80	15,297.578	5.997	7.283	123.683
12/07	1.439	1.937	3.959	5,651.07	7,540.64	30,580.123	2.377	2.746	174.872

Day	nRMSE (%)			MRE (%)		
	Proposed	NN	ARMAX	Proposed	NN	ARMAX
01/01	13.918	14.511	1,020.140	5.998	7.478	14.769
01/02	10.920	33.027	899.528	5.627	16.291	13.692
01/03	13.057	34.120	852.455	6.341	17.516	13.375
01/04	9.152	16.114	1,073.347	4.129	8.300	15.378
01/05	11.427	21.017	1,138.613	5.764	9.924	16.301
01/06	18.695	30.908	1,069.624	7.462	14.662	16.063
01/07	14.553	25.113	1,026.206	6.910	10.682	15.303
02/01	13.810	22.936	909.054	8.267	12.825	15.538
02/02	11.399	27.230	909.054	6.975	13.497	15.538
02/03	16.303	25.965	983.450	8.024	13.563	17.469
02/04	12.691	19.311	983.450	5.942	11.138	17.469
02/05	8.874	28.000	905.834	4.489	12.256	14.656
02/06	25.626	27.447	1,272.206	11.532	12.155	20.466
02/07	32.224	32.285	1,067.624	10.883	11.360	17.592
03/01	11.370	41.331	1,187.307	6.799	25.588	21.822
03/02	9.914	40.001	1,390.874	6.723	26.849	25.536
03/03	8.135	30.423	1,169.541	5.091	16.500	20.761
03/04	26.135	27.387	887.163	9.909	12.833	16.131

Table 8: Continued.

Day	nRMSE (%)			MRE (%)		
	Proposed	NN	ARMAX	Proposed	NN	ARMAX
03/05	4.345	20.561	1,425.378	2.266	13.653	26.510
03/06	12.519	32.091	1,176.774	7.778	18.656	21.171
03/07	9.599	38.994	1,177.284	5.661	22.425	21.579
04/01	8.663	27.284	1,231.268	5.537	16.851	25.595
04/02	15.145	25.852	1,022.633	6.722	13.204	22.395
04/03	9.654	26.214	1,427.525	5.983	16.624	30.444
04/04	7.254	31.139	1,316.786	4.716	18.500	27.521
04/05	10.048	29.973	1,177.148	5.200	14.165	24.993
04/06	14.681	35.275	1,067.380	7.236	19.535	21.644
04/07	20.281	33.345	808.403	8.347	14.010	16.081
05/01	9.503	40.209	1,256.600	5.632	21.394	24.752
05/02	13.330	22.423	1,147.557	6.599	13.992	22.753
05/03	24.824	24.188	947.494	10.051	12.197	16.152
05/04	18.907	30.210	1,190.338	8.306	15.398	23.030
05/05	13.096	24.050	1,314.160	7.197	11.697	26.288
05/06	11.498	30.886	1,218.676	8.235	19.610	24.261
05/07	13.131	25.604	1,179.327	8.261	14.324	21.520
06/01	14.448	34.755	1,152.012	7.592	23.100	21.252
06/02	8.849	34.312	1,192.195	5.891	20.563	23.753
06/03	18.847	25.790	897.064	9.709	14.393	16.905
06/04	17.194	20.926	837.365	9.770	12.497	16.478
06/05	26.289	35.455	921.107	13.509	19.568	17.906
06/06	14.282	33.530	1,114.030	8.251	20.769	21.218
06/07	19.431	37.653	1,050.108	10.020	24.430	21.915
07/01	10.622	21.078	1,181.775	6.499	12.531	22.491
07/02	36.943	41.277	757.124	12.145	14.461	13.128
07/03	15.307	20.924	1,037.434	8.754	11.774	18.219
07/04	17.189	21.794	1,142.264	9.803	14.331	20.718
07/05	19.449	35.651	1,075.984	10.874	19.678	19.830
07/07	12.628	30.040	1,191.123	6.558	15.228	21.528
07/08	11.604	24.552	1,020.823	5.488	13.235	15.882
08/01	15.459	31.376	3,713.730	5.542	9.842	32.933
08/02	31.752	101.902	3,619.736	7.206	18.296	20.717
08/03	21.645	40.431	2,038.446	8.033	18.913	18.104
08/04	8.626	17.789	1,705.275	4.636	8.299	13.949
08/05	21.572	25.467	1,532.890	9.082	9.504	15.385
08/06	10.547	16.323	1,556.358	6.069	10.502	27.467
08/07	16.993	26.264	2,560.471	8.793	12.451	32.933
09/01	18.924	29.074	1,687.406	8.406	13.965	25.700
09/02	18.842	28.119	1,636.325	9.620	14.475	24.088
09/03	13.106	24.863	1,356.377	8.056	14.258	18.393
09/04	14.266	21.336	1,698.551	10.665	14.477	26.872
09/05	20.758	28.641	1,876.200	10.380	18.687	28.066
09/06	21.712	21.666	1,806.133	9.266	10.128	29.358
09/07	19.637	23.660	1,548.149	12.549	12.887	24.049
10/01	18.161	23.032	1,105.843	9.923	11.538	19.463
10/02	18.500	24.482	1,249.593	9.418	12.313	20.817
10/03	18.020	34.348	1,377.614	9.289	15.056	21.515
10/04	16.027	39.083	1,087.239	8.974	19.046	17.569
10/05	7.832	31.003	1,179.989	4.598	18.225	20.112
10/06	20.620	41.448	1,193.311	9.007	22.021	20.349
10/07	22.196	34.083	1,082.929	9.729	17.357	16.098
11/01	49.130	130.561	7,104.398	8.493	27.930	40.952

Table 8: Continued.

Day	nRMSE (%)			MRE (%)		
	Proposed	NN	ARMAX	Proposed	NN	ARMAX
11/02	19.305	36.016	3,606.419	8.290	14.067	23.138
11/03	15.278	24.365	3,184.970	6.657	12.329	20.669
11/04	14.961	19.793	3,193.303	6.578	9.908	19.868
11/05	22.442	32.104	2,633.749	9.961	16.326	21.337
11/06	21.184	40.396	2,014.346	9.161	19.725	16.710
11/07	27.663	34.819	2,216.269	10.941	14.606	23.672
12/01	18.519	44.756	1,092.212	10.054	27.317	17.486
12/02	23.678	34.238	1,289.944	12.923	19.191	23.925
12/03	17.632	37.733	1,306.516	10.099	22.823	25.363
12/04	25.581	35.973	1,059.437	14.333	20.541	21.098
12/05	45.724	54.227	1,333.726	21.136	27.274	25.272
12/06	41.684	50.625	829.222	21.343	27.752	15.739
12/07	17.575	20.301	1,172.409	7.992	10.763	21.996

Table 9: Percentage between the RMSE value and the rated capacity of the proposed prediction model.

Day	Percentage (%)	Day	Percentage (%)	Day	Percentage (%)	Day	Percentage (%)
01/01	9.6944	04/01	7.4778	07/01	8.4944	10/01	13.7611
01/02	7.9278	04/02	11.1556	07/02	18.6167	10/02	13.9222
01/03	9.5778	04/03	8.4722	07/03	12.5333	10/03	13.4722
01/04	6.9611	04/04	6.1833	07/04	15.3167	10/04	12.0778
01/05	8.2389	04/05	7.7722	07/05	15.5000	10/05	6.7333
01/06	12.3944	04/06	11.0111	07/05	9.4722	10/06	14.2111
01/07	10.8611	04/07	13.6833	07/06	7.8944	10/07	15.5000
02/01	10.4556	05/01	7.4944	08/01	6.9444	11/01	13.9444
02/02	8.6500	05/02	11.3056	08/02	8.6889	11/02	12.2333
02/03	10.7778	05/03	17.4000	08/03	15.1944	11/03	10.4667
02/04	9.4222	05/04	13.6778	08/04	6.0833	11/04	10.4000
02/05	6.3889	05/05	9.4333	08/05	12.3056	11/05	16.3611
02/06	17.5556	05/06	9.9222	08/06	8.5611	11/06	14.6611
02/07	15.7278	05/07	11.6000	08/07	12.1444	11/07	15.6333
03/01	8.9667	06/01	12.2556	09/01	12.2611	12/01	15.3722
03/02	8.8000	06/02	7.5556	09/02	14.1333	12/02	18.3556
03/03	6.7278	06/03	13.8944	09/03	10.9000	12/03	14.5111
03/04	15.4833	06/04	14.7889	09/04	12.8833	12/04	20.1000
03/05	3.7722	06/05	20.7944	09/05	18.8389	12/05	30.0556
03/06	11.3500	06/06	12.2333	09/06	15.5000	12/06	33.3167
03/07	8.1833	06/07	16.6778	09/07	16.8389	12/07	13.2056

percentage of the proposed prediction model is 3.7722%, while 0.5758% was obtained by the previous studies [15]. Although the study [15] has the smallest percentage, it is only 1 hr ahead with a 15 min interval forecasting horizon. Compared with the same forecasting horizon of 1 day ahead, the percentage of the proposed prediction model is comparable. However, the percentage of the previous studies [16, 18, 19, 20, 21] is smaller than the proposed model. Moreover, the highest percentage obtained in the proposed prediction model is 33.3167%, lower than 44.6470% in the previous study [17]. Hence, the result of the proposed prediction model is comparable with that of the previous studies.

Furthermore, the uncertainty of the results was also analyzed to validate the proposed prediction model. Figure 10 depicts the solar PV power data of March 5, 2020, the day with the best forecasting result, simulated 100 times to assess the uncertainty. The scatter plot from the simulation closely aligns with the actual solar PV power data, indicating a consistent pattern. Table 11 further validates the result by listing the median and mean of the simulated solar PV power per

Table 10: Percentage between the RMSE value and the rated capacity of the studies in Table 1.

References	Rated capacity	RMSE	Percentage (%)
Li et al. [15]	26.5 kW	0.1526 kW	0.5758
Lim et al. [16]	2,500 W	43.87 W	1.7548
Zhanga and Kong [17]	10 MW	4.4647 MW	44.6470
Li et al. [18]	18 MW	0.191 MW	1.0611
Li et al. [19]	50 MW	1.003 MW	2.0060
Li et al. [20]	100 MW	1.7596 MW	1.7596
Lin and Li [22]	5.83 kW	0.2132 kW	3.6569

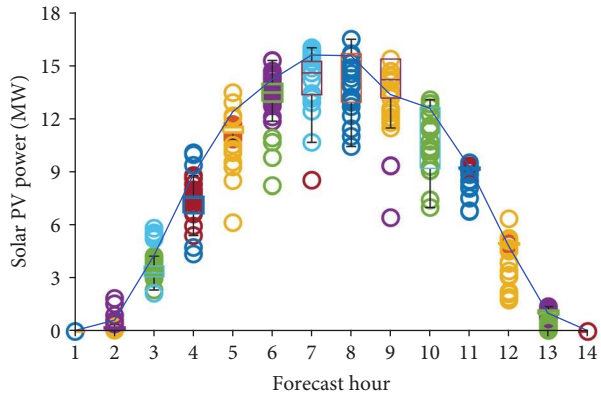


Figure 10: Simulated solar PV power data of March 5, 2020.

Table 11: Comparison of the median and mean of the simulated solar PV power data and the actual solar PV power data of March 5, 2020.

Hour	Proposed		Actual
	Median (MW)	Mean (MW)	
6:00	0	0	0
7:00	0.152	0.232	0.586
8:00	3.31	3.32	4.25
9:00	7.58	7.03	8.85
10:00	11.55	11.17	12.41
11:00	13.50	13.30	14.22
12:00	14.61	14.11	15.62
13:00	15.60	14.63	15.58
14:00	14.23	13.91	13.38
15:00	12.66	11.40	12.61
16:00	9.25	9.10	9.31
17:00	4.99	4.68	4.81
18:00	1.13	0.991	0.982
19:00	0	0.000132	0

hour and comparing it with actual data. Table 11 reveals that the median and mean of the 100 times simulated data closely match the actual solar PV power data. Consequently, the proposed prediction model generates reliable results, considering the uncertainty of its parameters. Therefore, this proposed prediction model developed accurate solar PV power profiles that effectively provide information on the shortage and excess solar PV power generation for reliable integration into the power system.

4. Conclusion and Discussion

This study developed a multiphase solar PV prediction model encompassing five phases: grouping, clustering, linking, classifying, and predicting. Demonstrating the model's effectiveness, solar PV power profiles for the first week of each month in both dry and wet seasons were presented. The data revealed higher solar PV power during the dry season, emphasizing the significance of accounting for seasonal

trends in predicting solar PV power. This observation highlights the crucial role of seasonal patterns in formulating the solar PV power prediction model for effective planning and design of power system networks.

In addition, the formulation of the prediction model accounted for the daily variations and patterns of solar PV power. The proposed model demonstrated superior accuracy in hourly forecasts, with 8 and 13 out of 14 forecast hours showing insignificant differences between the forecast and actual values in the dry and wet seasons, respectively. Therefore, the proposed multiphase solar PV prediction model provides more precise hourly forecasts, offering significant benefits for monitoring the day-long fluctuations of solar PV power.

This study extensively evaluated the proposed solar PV prediction model to verify its effectiveness. The comparison of the average forecast values with actual values revealed no significant differences, validating the model's efficacy. The prediction model's performance was further validated by comparing the forecast and actual values using the performance metrics. The best metrics obtained using the proposed prediction model were achieved during the dry season with MAE of 0.408 MW, MSE of 460.51 MW, RMSE of 0.679 MW, nRMSE of 4.345%, and MRE of 2.266%. These results demonstrate that the uncertainties obtained using the proposed prediction model are six times lower than those of the conventional method, NN, and 12 times lower than those of ARMAX.

In addition, the percentage between the RMSE value and the rated capacity was obtained and compared to further evaluate the performance of the proposed prediction model. The smallest percentage obtained using the proposed prediction model was 3.7722%, which is comparable to the result obtained by the previous studies. In comparison, the highest percentage obtained by the proposed model is 33.3167%, smaller than that obtained by one of the previous studies. This shows that the proposed prediction model provided accurate results comparable to previous studies that predicted solar PV power.

Moreover, uncertainty analysis affirmed the reliability and consistency of the forecasting result of the proposed solar PV power prediction model. It reveals that the median and mean of the 100 times simulated data of a particular day in the dry season closely match the actual solar PV power data from the conducted analysis. In conclusion, the developed multiphase solar PV prediction model provides valuable information about solar PV power generation availability, which is essential for ensuring reliable integration into power systems.

Nomenclature

GRA:	Gray relational analysis
GRG:	Gray relational grade
PCC:	Pearson correlation coefficient
$X_{i,j}$:	Solar PV power data in each day and hour
$Y_{1_{i,j}}, Y_{2_{i,j}}, \ldots, Y_{l_{i,j}}$:	Weather data of each weather factor in each day and hour
i:	Day (i.e., 1–365)
j:	Hour (i.e., 0–23)
l:	Number of weather factors
$\widehat{X}_{i,j}$:	Normalized solar PV power data
$\widehat{Y}_{l_{i,j}}$:	Normalized weather data of each weather factor
X_d and X_w:	Solar PV power data in dry and wet seasons
Y_{l_d} and Y_{l_w}:	Weather data of each weather factor in dry and wet seasons
k_d and k_w:	Optimal number of clusters in dry and wet seasons
$X_{d_{clusters}}$ and $X_{w_{clusters}}$:	Cluster output of each day in dry and wet seasons after hierarchical clustering
$X_{d_1},\ldots, X_{d_{k_d}}$ and $X_{w_1},\ldots, X_{w_{k_w}}$:	Solar PV power data in each cluster in the dry and wet seasons after hierarchical clustering
$\overline{X}_{d_1},\ldots, \overline{X}_{d_{k_d}}$ and $\overline{X}_{w_1},\ldots, \overline{X}_{w_{k_w}}$:	Mean solar PV power data of each cluster in the dry and wet seasons
$X_{D_{clusters}}$ and $X_{W_{clusters}}$:	Cluster output of each day in dry and wet seasons after hybrid clustering
$X_{D_1},\ldots, X_{D_{k_d}}$ and $X_{W_1},\ldots, X_{W_{k_w}}$:	Solar PV power data in each cluster in the dry and wet seasons after hybrid clustering
$\gamma_{1_d},\ldots,\gamma_{l_d}$ and $\gamma_{1_w},\ldots,\gamma_{l_w}$:	GRG of each weather factor in the dry and wet seasons
Y_{d_s} and Y_{w_s}:	Significant weather factors in dry and wet seasons
$r_{1_d},\ldots, r_{m_d}$ and $r_{1_w},\ldots, r_{n_w}$:	Correlation coefficient of each significant weather factor in the dry and wet seasons
m and n:	Number of significant weather factors in dry and wet seasons
Y_D and Y_W:	Data of each relevant weather factor in dry and wet seasons
o and p:	Number of relevant weather factors in dry and wet seasons
SVM_D and SVM_W:	SVM classification models in dry and wet seasons
NN_D and NN_W:	Neural network prediction model in dry and wet seasons
X_f:	Forecast solar PV power data.

Data Availability

The data used to support the findings of this study could not be released due to regulations by the Korean Government.

Conflicts of Interest

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

Acknowledgments

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea Government (MOTIE) (RS-2023-00234563, Development of power system modeling & analysis and interoperability evaluation technology applied with grid forming based on distributed energy, and 20224000000440, Sector coupling energy industry advancement manpower training program).

References

[1] Renewables 2022 Global Status Report, "REN21," Paris, France, 2022, [Online]. Available: [accessed 05.26.23]. https://www.ren21.net/wp-content/uploads/2019/05/GSR2022_Full_Report.pdf.

[2] M. Morey, N. Gupta, M. M. Garg, and A. Kumar, "A comprehensive review of grid-connected solar photovoltaic system: architecture, control, and ancillary services," *Renewable Energy Focus*, vol. 45, pp. 307–330, 2023.

[3] K. Jha and A. G. Shaik, "A comprehensive review of power quality mitigation in the scenario of solar PV integration into utility grid," *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, vol. 3, Article ID 100103, 2023.

[4] O. Gandhi, D. S. Kumar, C. D. Rodríguez-Gallegos, and D. Srinivasan, "Review of power system impacts at high PV penetration part I: factors limiting PV penetration," *Solar Energy*, vol. 210, pp. 181–201, 2020.

[5] E. Rakhshani, K. Rouzbehi, A. J. Sánchez, A. C. Tobar, and E. Pouresmaeil, "Integration of large scale PV-based generation into power systems: a survey," *Energies*, vol. 12, no. 8, Article ID 1425, 2019.

[6] S. Rahman, S. Saha, M. E. Haque et al., "A framework to assess voltage stability of power grids with high penetration of solar PV systems," *International Journal of Electrical Power & Energy Systems*, vol. 139, Article ID 107815, 2022.

[7] A. Sayed, M. El-Shimy, M. El-Metwally, and M. Elshahed, "Reliability, availability and maintainability analysis for grid-connected solar photovoltaic systems," *Energies*, vol. 12, no. 7, Article ID 1213, 2019.

[8] Y. Jang, Z. Sun, S. Ji et al., "Grid-connected inverter for a PV-powered electric vehicle charging station to enhance the stability of a microgrid," *Sustainability*, vol. 13, no. 24, Article ID 14022, 2021.

[9] M. J. Mayer and G. Gróf, "Extensive comparison of physical models for photovoltaic power forecasting," *Applied Energy*, vol. 283, Article ID 116239, 2021.

[10] K. J. Iheanetu, "Solar photovoltaic power forecasting: a review," *Sustainability*, vol. 14, no. 24, Article ID 17005, 2022.

[11] A. El hendouzi and A. Bourouhou, "Solar photovoltaic power forecasting," *Journal of Electrical and Computer Engineering*, vol. 2020, Article ID 8819925, 21 pages, 2020.

[12] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif, "A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization," *Renewable and Sustainable Energy Reviews*, vol. 124, Article ID 109792, 2020.

[13] M. N. Akhter, S. Mekhilef, H. Mokhlis, and N. M. Shah, "Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques," *IET Renewable Power Generation*, vol. 13, no. 7, pp. 1009–1023, 2019.

[14] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Solar Energy*, vol. 136, pp. 78–111, 2016.

[15] P. Li, K. Zhou, X. Lu, and S. Yang, "A hybrid deep learning model for short-term PV power forecasting," *Applied Energy*, vol. 259, Article ID 114216, 2020.

[16] S.-C. Lim, J.-H. Huh, S.-H. Hong, C.-Y. Park, and J.-C. Kim, "Solar power forecasting using CNN-LSTM hybrid model," *Energies*, vol. 15, no. 21, Article ID 8233, 2022.

[17] Y. Zhang and L. Kong, "Photovoltaic power prediction based on hybrid modeling of neural network and stochastic differential equation," *ISA Transactions*, vol. 128, pp. 181–206, 2022.

[18] G. Li, S. Guo, X. Li, and C. Cheng, "Short-term forecasting approach based on bidirectional long short-term memory and convolutional neural network for regional photovoltaic power plants, sustainable energy," *Sustainable Energy, Grids and Networks*, vol. 34, Article ID 101019, 2023.

[19] Q. Li, X. Zhang, T. Ma, D. Liu, H. Wang, and W. Hu, "A multi-step ahead photovoltaic power forecasting model based on TimeGAN, Soft DTW-based K-medoids clustering, and a CNN-GRU hybrid neural network," *Energy Reports*, vol. 8, pp. 10346–10362, 2022.

[20] Z. Li, R. Xu, X. Luo, X. Cao, S. Du, and H. Sun, "Short-term photovoltaic power prediction based on modal reconstruction and hybrid deep learning model," *Energy Reports*, vol. 8, pp. 9919–9932, 2022.

[21] A. A. H. Lateko, H.-T. Yang, and C.-M. Huang, "Short-term PV power forecasting using a regression-based ensemble method," *Energies*, vol. 15, no. 11, Article ID 4171, 2022.

[22] J. Lin and H. Li, "A short-term PV power forecasting method using a hybrid Kmeans-GRA-SVR model under Ideal weather condition," *Journal of Computer and Communications*, vol. 8, no. 11, pp. 102–119, 2020.

[23] M. B. Arias and S. Bae, "Solar photovoltaic power prediction using big data tools," *Sustainability*, vol. 13, no. 24, Article ID 13685, 2021.

[24] P. Arora, Deepali, and S. Varshney, "Analysis of K-means and K-medoids algorithm for big data," *Procedia Computer Science*, vol. 78, pp. 507–512, 2016.

[25] B. Chen, P. C. Tai, R. Harrison, and Yi Pan, "Novel hybrid hierarchical-K-means clustering method (H-K-means) for microarray analysis," in *2005 IEEE Computational Systems Bioinformatics Conference - Workshops (CSBW'05)*, pp. 105–108, IEEE, Stanford, CA, USA, 2005.

[26] A. D. Peterson, A. P. Ghosh, and R. Maitra, "Merging *K* -means with hierarchical clustering for identifying general-shaped groups," *Stat*, vol. 7, no. 1, Article ID e172, 2018.

[27] B. Subramaniyan, R. Pan, J. Kuitche, and G. TamizhMani, "Quantification of environmental effects on PV module degradation: a physics-based data-driven modeling method," *IEEE Journal of Photovoltaics*, vol. 8, no. 5, pp. 1289–1296, 2018.

[28] I. Kaaya, M. Koehl, A. P. Mehilli, S. de Cardona Mariano, and K. A. Weiss, "Modeling outdoor service lifetime prediction of PV modules: effects of combined climatic stressors on PV module power degradation," *IEEE Journal of Photovoltaics*, vol. 9, no. 4, pp. 1105–1112, 2019.

[29] Q. Liu, Q. Hu, J. Zhou, D. Yu, and H. Mo, "Remaining useful life prediction of PV systems under dynamic environmental conditions," *IEEE Journal of Photovoltaics*, vol. 13, no. 4, pp. 590–602, 2023.

[30] Q. Yang, Q. Kang, Q. Huang, Z. Cui, Y. Bai, and H. Wei, "Linear correlation analysis of ammunition storage environment based

on Pearson correlation analysis," *Journal of Physics: Conference Series*, vol. 1948, no. 1, Article ID 012064, 2021.
[31] World Weather Online, [Online]. Available: https://www.worldweatheronline.com/hwd/hfw.aspx.
[32] M. Sani and A. Sule, "Effect of temperature on the performance of photovoltaic module," *International Journal of Innovative Science and Research Technology*, vol. 5, no. 9, pp. 670–676, 2020.
[33] F. Shaik, S. S. Lingala, and P. Veeraboina, "Effect of various parameters on the performance of solar PV power plant: a review and the experimental study," *Sustainable Energy Research*, vol. 10, Article ID 6, 2023.
[34] W. Z. Leow, Y. M. Irwan, M. Irwanto, A. R. Amelia, and I. Safwati, "Influence of wind speed on the performance of photovoltaic panel," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 15, no. 1, pp. 60–68, 2019.
[35] C. Schwingshackl, M. Petitta, J. E. Wagner et al., "Wind effect on PV module temperature: analysis of different techniques for an accurate estimation," *Energy Procedia*, vol. 40, pp. 77–86, 2013.
[36] W. Jo, N. Ham, J. Kim, and J. Kim, "The cleaning effect of photovoltaic modules according to precipitation in the operation stage of a large-scale solar power plant," *Energies*, vol. 16, no. 17, Article ID 6180, 2023.
[37] P. Sarmah, D. Das, M. Saikia et al., "Comprehensive analysis of solar panel performance and correlations with meteorological parameters," *ACS Omega*, vol. 8, no. 50, pp. 47897–47904, 2023.
[38] M. R. Das, "Effect of different environmental factors on performance of solar panel," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 11, pp. 15–18, 2019.
[39] A. Bonkaney, S. Madougou, and R. Adamou, "Impacts of cloud cover and dust on the performance of photovoltaic module in Niamey," *Journal of Renewable Energy*, vol. 2017, Article ID 9107502, 8 pages, 2017.
[40] U. Ilo Frederick, "Determination of the effect of temperature changes on power output of solar panel," *International Journal of Research in Engineering and Applied Sciences*, vol. 7, no. 9, pp. 70–79, 2017.
[41] L. Sun and Y. Sun, "Photovoltaic power forecasting based on artificial neural network and ultraviolet index," *International Journal of Computing*, vol. 21, no. 2, pp. 153–158, 2022.
[42] M. Kaushik and B. Mathur, "Comparative study of k-means and hierarchical clustering techniques," *International Journal of Software & Hardware Research in Engineering*, vol. 2, no. 6, pp. 93–98, 2014.
[43] M. Kaur and U. Kaur, "Comparison between k-mean and hierarchical algorithm using query redirection," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 7, pp. 1454–1459, 2013.
[44] MATLAB and SIMULINK, [Online]. Available: https://www.mathworks.com/.
[45] J. Dai, X. Liu, and F. Hu, "Research and application for grey relational analysis in multigranularity based on normality grey number," *The Scientific World Journal*, vol. 2014, Article ID 312645, 10 pages, 2014.
[46] Z. A. Khan, A. N. Siddiquee, N. Z. Khan, U. Khan, and G. A. Quadir, "Multi response optimization of wire electrical discharge machining process parameters using Taguchi based Grey relational analysis," *Procedia Materials Science*, vol. 6, pp. 1683–1695, 2014.
[47] M. B. Arias and S. Bae, "Electric vehicle charging demand forecasting model based on big data technologies," *Applied Energy*, vol. 183, pp. 327–339, 2016.
[48] G. Fang, Y. Guo, X. Huang, M. Rutten, and Y. Yuan, "Combining Grey relational analysis and a Bayesian model averaging method to derive monthly optimal operating rules for a hydropower reservoir," *Water*, vol. 10, no. 8, Article ID 1099, 2018.
[49] C. Tritham, L. Lekawat, A. Arrayangkool, C. Viwatwongkasem, P. Satitvipawee, and P. Soontornpipit, "A comparison between correlation and Grey relational for big data and analytics," in *2018 International Electrical Engineering Congress (iEECON)*, pp. 1–4, IEEE, Krabi, Thailand, 2018.
[50] S. Kumar and I. Chong, "Correlation analysis to identify the effective data in machine learning: Prediction of depressive disorder and emotion states," *International Journal of Environmental Research and Public Health*, vol. 15, no. 12, Article ID 2907, 2018.
[51] M. M. Mukaka, "Statistics corner: a guide to appropriate use of correlation coefficient in medical research," *Malawi Medical Journal*, vol. 24, no. 3, pp. 69–71, 2012.