Hindawi

Research Article

Hourly Forecasting of Solar Photovoltaic Power in Pakistan Using Recurrent Neural Networks

Sohrab Khan [iD],[1] **Faheemullah Shaikh** [iD],[1] **Mokhi Maan Siddiqui,**[1] **Tanweer Hussain** [iD],[2] **Laveet Kumar** [iD],[2] **and Afroza Nahar** [iD][3]

[1]*Department of Electrical Engineering, Mehran University of Engineering & Technology, Jamshoro 76062, Pakistan*
[2]*Department of Mechanical Engineering, Mehran University of Engineering & Technology, Jamshoro 76062, Pakistan*
[3]*Department of Computer Science, Faculty of Science and Technology, American International University – Bangladesh,*
 Dhaka 1229, Bangladesh

Correspondence should be addressed to Afroza Nahar; afroza@aiub.edu

The solar photovoltaic (PV) power forecast is crucial for steady grid operation, scheduling, and grid electricity management. In this work, numerous time series forecast methodologies, including the statistical and artificial intelligence-based methods, are studied and compared fastidiously to forecast PV electricity. Moreover, the impact of different environmental conditions for all of the algorithms is investigated. Hourly solar PV power forecasting is done to confirm the effectiveness of various models. Data used in this paper is of one entire year and is acquired from a 100 MW solar power plant, namely, Quaid-e-Azam Solar Park, Bahawalpur, Pakistan. This paper suggests recurrent neural networks (RNNs) as the best-performing forecasting model for PV power output. Furthermore, the bidirectional long-short-term memory RNN framework delivered high accuracy results in all weather conditions, especially under cloudy weather conditions where root mean square error (RMSE) was found lowest 0.0025, R square stands at 0.99, and coefficient of variation of root mean square error (RMSE) Cv was observed 0.0095%.

1. Introduction

Electricity is a fundamental instrument to continue socio-economic activities. Pakistan has been facing an electricity shortage for many years due to heavy reliance on expensive imported fuel, suboptimal transmission and distribution systems, and poor revenue collection [1]. Pakistan's current power generation mix is heavily skewed towards imported and high-carbon fuels. Pakistan produced 67.5% of its total power consumption through thermal (furnace oil, coal, regasified liquified gas, and nuclear), 32% from renewables (hydel, wind, solar, and bagasse), and the rest is being imported from Iran. Among other renewable energy sources, solar PV power only contributes less than 1% of country's total power consumption [2].

Solar PV has emerged as a reliable technology and competitive power source globally among its renewable and non-renewable counterparts. The solar PV power installed

capacity increased by 22% in 2019, with the second-largest renewable generation growth slightly behind wind [3]. Pakistan is also blessed with a vast potential to generate solar PV power and could be an essential and clean source of energy for country's future energy needs. World Bank (WB) reported that Pakistan could meet its electricity demand by only utilizing as small as 0.071% of Pakistan's land area for solar PV power generation [4]. Since the output of solar PV power is intermittent, it depends entirely on the availability of sunshine hours throughout the day, solar irradiance, angle of incident, the circuit of a cell, and metrological conditions [5, 6]. Hence, integrating solar PV-generated power with a grid creates challenges in grid planning and operation. Forecasting solar PV power is an important planning activity, and a robust solar power forecast is essential to mitigate solar-induced variability and facilitate solar PV grid integration [7]. There has been little study done indigenously in Pakistan on solar power forecasting. Most of the

studies focus on solar system performance and solar radiations [8–11]. Pakistan has taken steps towards the adoption of renewable energy. In a recent meeting for climate change, leaders of the world met in Washington, where Pakistan assured that it would shift to 60% and 30% on renewable energy and use of electric vehicles, respectively, by 2030 [12]. To achieve this huge target, it is essential to accurately forecast the power output of available renewable energy sources, especially PV systems.

In reference [13], a statistical model autoregressive integrated moving average (ARIMA) for weather forecast was proposed, the model with the lowest mean square error value was found accurate with MSE value 0.00029, this work also suggests that by reducing training dataset, accuracy decays (mean square error value increases). In reference [14], a study for the monthly mean temperature forecast for 36 months was completed. The original temperature historical data was obtained from automatic weather station Nanjing. Seasonal autoregressive integrated moving average (SARIMA) model is suggested best fit model for temperature forecast, the mean square error (MSE) for the last three years of validation data is 0.84, 0.89, and 0.94, respectively. MSE values are relatively low, with a slight increase of 0.05 every year. Since the increase in error is not of more importance, the model can be safely utilized for temperature forecast.

In reference [15], by utilizing data from National Solar Radiance DataBase (NSRDB), a multisite (i.e., Atlanta, New-york, Huawei)-based study was conducted for accurate forecasting of short-term solar irradiance. This study compared different models, i.e., ARIMA, support vector regression (SVR), back propagation neural network (BPNN), and RNN. In conclusion to this study, LSTM results were more accurate, especially on mixed days and cloudy days. Moreover, root mean square error (RMSE), R squared, and mean absolute percentage error (MAPE) values for LSTM in other complicated weather situations were also found competitive than models as mentioned above. Hence, this study suggests LSTM as an accurate model for short-term solar irradiance forecast. In reference [16], a comparative analysis of general regression (GR) and back propagation (BP) models was conducted. First, temperature and irradiance were found most important factors for input data by using Pearson's correlation coefficient. Then, learning vector quantization neural network (LVQ NN) was implemented to observe three different weather conditions, sunny, cloudy, and rainy. Finally, a comparative study of the backpropagation neural network and general regression neural network was conducted. The conclusion to this study suggests the backpropagation model as a more accurate model than the general regression model for PV power generation forecast.

In reference [17], the case study of South Korea was investigated to predict the amount of PV power generation at new sites. Dataset of 164 different sites that contained weather information, estimated solar irradiance, plant capacity, and electricity trading was studied on LSTM model for prediction. It was observed that LSTM could learn complex and nonlinear patterns between power output and factors affecting it at different sites. It is concluded that the proposed LSTM model can be beneficial in accurately predicting PV power output, in any region, with known historical weather data. In reference [18], the backpropagation method is proposed to forecast 24-hour PV power accurately. Before model implementation, a correlation analysis was conducted to investigate the relationship between power output and ambient temperature. Based on this correlation analysis, hourly solar radiation intensity, highest daily temperature, lowest daily temperature, average daily temperature, and hourly PV output were given as inputs to forecast PV power. According to results obtained, a model with settings of 28 neurons in the input layer, 11 neurons in the output layer, and 20 hidden nodes was found to perform the BP model best for PV power output forecast.

A study in reference [19] proposed a novel model to forecast the one-day power output of a single 20 MW PV power plant. Support vector machines (SVM) with weather classification was studied in this work. The process divided weather conditions into four types, clear sky, cloudy, foggy, and rainy. The presented model showcased promising results with low forecasting errors. RMSE was observed at 2.10, and mean relative error (MRE) was found at 8.64% for the chosen site. In reference [20], a hybrid approach is proposed for three different neural networks (FFNN, GRNN, and MLP) to forecast 24 hours PV output for 16 different rooftop solar panels of capacity 250 W. Firstly, stepwise regression was used to select meteorological parameters that are strongly correlated with PV power generation. Then, these parameters were used as input to three different single-stage models (feedforward neural network, general regression neural network, and multilayer regression) and their corresponding hybrid models. Hybrid models prediction results were found very close to measured values. Furthermore, accuracy metrics also pointed out that hybrid models are slightly better than single-staged counterparts.

In reference [21], a case study compared artificial neural networks (ANNs) and RNNs to predict solar irradiance. The study proposed deep learning RNN as a better performing model for forecasting solar radiations. In the results, compared to ANNs, for RNN, a significant improvement of 47% was observed in normalized mean bias error (NMBE), and a 26% improvement in RMSE was also observed. It was observed that with an increase in sampling frequency from 1 hour to 10 minutes, coefficient of variation of RMSE (Cv(RMSE)) of ANN dropped by approximately 30%, and CV (RMSE) of RNN dropped by about 2.19%. This study also suggested that adding a moving average algorithm to predicting model accuracy can improve RNN. In reference [22], the study proposes two PV output prediction models using LSTM and GRU (gate recurrent unit) without knowledge of future meteorological information. This study utilized meteorological information of morning hours to estimate the PV power output around noon. The results found that the proposed GRU-based model could capture the seasonal trend between PV power output in peak zone and its preceding zones more effectively than that of LSTM-based model. Furthermore, even in increased difficulty levels, the GRU-based model performs more accurately than other models.
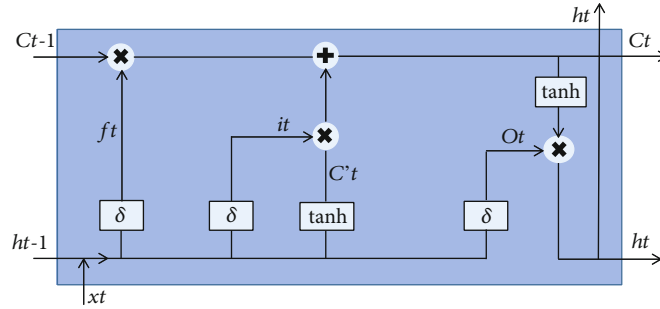
FIGURE 1: Cell structure of LSTM.

In reference [23], Bi-LSTM for accurate forecasting of solar irradiance hourly and daily was proposed in the study based on two different sites. Multiple models, i.e., vanilla LSTM, attention-based LSTM, Bi-LSTM, and convolutional neural networks, were invested gated for this study, and the models were developed based on single location univariate and multiple location data. Performance and evaluation of model were investigated based on rolling window evaluation. Results indicated that Bi-LSTM and attention-based LSTMs could be used for daily solar irradiance forecasts. In reference [24], a study of single-layer and multilayer LSTM models was conducted for the accurate forecast of PV power generation. Cv (RMSE) was used as a precision method, the results observed for single-layer and multilayer model were 13.8% and 13.2%, respectively. A very little difference in error was observed between the single- and multilayer LSTM model's forecast. Multilayer LSTM showed reduced error. Hence, the accurate forecast is achievable through multilayer LSTM.

Different authors have tested numerous studies and different methodologies. The goal of every author was to propose an accurate model for forecasting. Much work has been already done, and a lot more is completed for precise forecasting. It is considered that deep learning methods have proved themselves in discussed studies for accurate short term time series forecasting. In few studies [13, 14, 25, 26], statistical models were also highlighted as better performing methods to forecast accurately.

In the current study, real-world data has been utilized for accurate PV power output forecast using deep learning LSTM and Bi-LSTM for the first time in the scenario of Pakistan. In addition, a dropout mechanism has been incorporated to prevent overfitting and ensure model's accuracy. Moreover, this study will test different hidden layers for LSTM and consider the best fit of these models. Conclusively, LSTM and Bi-LSTM will be compared, and a final accurate model will be suggested. This study comprises different time series forecasting algorithms for PV power forecasts. Statistical and artificial intelligence-based methodologies are both included in the procedure. The seasonal autoregressive integrated moving average (SARIMA) is a statistical model utilized in this study. The goal of SARIMA is to secure seasonality in data, while our dataset comprises seasonality. long-short-term memory (LSTM) and bidirectional LSTM (Bi-LSTM) are two forms of recurrent neural networks (RNNs) that are studied. The following are the key novelties of this paper:

(i) A deep learning Bi-LSTM is proposed as an accurate power forecasting model for grid-connected PV systems in the study

(ii) Evaluation and comparison of various forecasting models, including statistical and neural network techniques, for time series forecasting of large-scale PV systems

(iii) For accuracy concerns, the study examined over multilayers of LSTM

(iv) The paper includes the time frames for which the forecasting models under consideration are effective

2. Methodology

2.1. Data Description. Data utilized in current work is provided by Quaid-e-Azam Solar Park, Bahawalpur, which is 100 MW power plant. It is a collective 1000 MW project under the China Pakistan Economic Corridor (CPEC) energy section. In the first phase, 100 MW was completed and has been in operation since August 2016 [27]. Data were recorded at intervals of 15 minutes at the power plant. Using Equation (1) [28], the data was averaged on an hourly basis to make signal smoother and improve algorithms' operation.

Furthermore, during operation, it was found that time series forecasts work more accurately on hourly averaged data than that of 15 minutes interval data [29]. The dataset is of one-year time period from 01-January-2019 to 31-December-2019. During the given time, it was observed that power output was constantly zero between 7 PM to 7 AM. Therefore, we only considered the data between 7 AM to 7 PM. The dataset was further separated according to weather conditions such as sunny days, cloudy days, rainy days, partially cloudy days, dusty days, and foggy days.

$$\text{Average} = \frac{t_1 + t_2 + t_3 + t_4}{n}, \tag{1}$$

where n is the number of observations, and $t1$, $t2$, $t3$, and $t4$ are time intervals of an hour.

2.2. Forecasting Models. Forecasting models used for current work are long-short-term memory (LSTM) and bidirectional long-short-term memory (Bi-LSTM), both special types of recurrent neural networks. A statistical approach, seasonal autoregressive integrated moving average (SARIMA), was also studied and compared with RNNs. A brief yet necessary introduction to these models is given, respectively.

2.2.1. Recurrent Neural Networks. Artificial neural networks (ANNs) are a set of algorithms that mimic the human brain. RNNs are ANN types that contain a loop that helps information pass from one step to another. RNNs have a memory based on the previous information, and they look at the previous state to predict the next state [30]. The special kind of RNN, LSTM, was introduced in 1997 to overcome long-term dependency issues [31]. LSTM contains three types of layers, namely input, hidden, and output layer. Unlike other neural networks, LSTMs include memory blocks that are connected through layers. Block's state and output are handled by gates present in each block [32]. Three gates (forget gate, input gate, and output gate) and a cell state make a single LSTM Block. Figure 1 shows the cell structure of LSTM, where $ct - 1$ is the previous cell state, $ht - 1$ is the hidden layer at time $t - 1$, ht is the hidden layer at time t, ct is preliminary input, and xt is input at time t [33].

(1) *Forget gate:* forget gate decides what information to keep and discard from the cell state. The sigmoid layer of forget gate generates either 0 or 1 as a value (ft in Figure 1), 0 means discard, and 1 means keep [34]

(2) *Input gate:* input gate updates the values to cell state; simply, it decides what to store in cell state. A sigmoid layer generates what values to update in cell state (it in Figure 1), and tanh (hyperbolic tangent) layer generates a new vector candidate value to be added to cell state ($C^{'}t$ in Figure 1); information from the sigmoid layer and tanh layer is then combined and updated to cell state [30]

(3) *Output gate:* the output gate generates output based on the input and memory of the block. The sigmoid layer has information about the required output (Ot in Figure 1), then values are pushed between -1 to 1 using tanh layer. Now, the output from the sigmoid layer and tanh layer is multiplied to get desired output [32]

Besides LSTM, another recurrent neural network algorithm Bi-LSTM is also being employed in this work, a special type of RNNs. The core idea of Bi-LSTM is to present two different neural networks, forward and backward connections to the same output [33]. Therefore, Bi-LSTM takes information from previous contexts and gets information from the future [35]. Bi-LSTM is the LSTMs operating forward and also backward. The forward generates output using information from history, while the backward uses information from the future, which helps in accurate forecasting [33]. Figure 2 depicts Bi-LSTM, the $xt - 1$, xt,
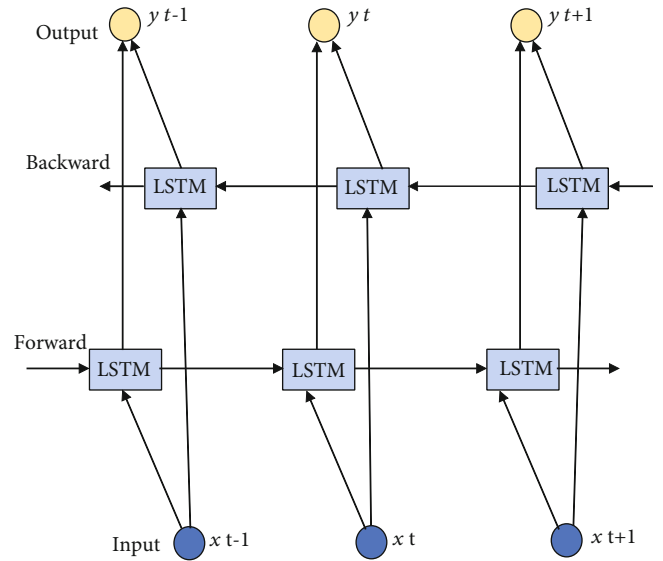

Figure 2: Bidirectional LSTM.

Table 1: Configuration of layers of RNNs.

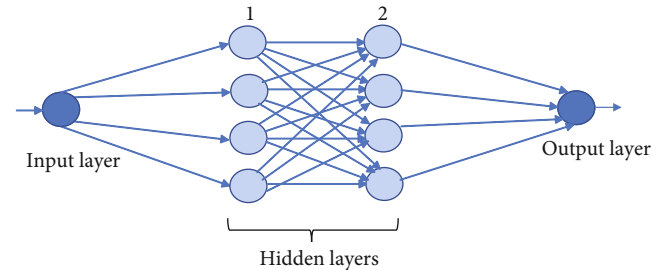| Model	Optimizer	Neurons	Hidden layers	Neurons division	Epochs
Bi-LSTM	Adam	32	1	32	100
LSTM	Adam	64	2	32 + 32	100

Figure 3: Block diagram of layers configuration of RNN.

and $xt + 1$ is a set of inputs from past information, the current input and future information, respectively. The mentioned *LSTM* in Figure 2 is the cell structure of LSTM described in Figure 1 [33].

2.2.2. Statistical Model. For time series forecasting, the widely used statistical approach is autoregressive integrated moving average, which is acronymically called ARIMA. An extension to ARIMA for seasonal data is seasonal ARIMA (SARIMA) model. With autoregression, integration, and moving average, additional seasonality parameters are added to form SARIMA, as mentioned below in Equation (2) [36].

$$\text{SARIMA} = (p, d, q) * (P, D, Q)^{n}. \tag{2}$$

TABLE 2: Cv (RMSE) results of hourly forecast of PV power output using LSTM with four different hidden layers.

No. of layers	Sunny	Cloudy	Rainy	Partial cloudy	Dusty	Foggy
1	0.08%	0.14%	1.51%	0.73%	0.65%	0.85%
2	0.15%	0.21%	0.60%	0.51%	0.50%	0.84%
3	0.47%	1.32%	2.12%	0.95%	0.86%	1.45%
4	0.52%	1.47%	0.88%	0.92%	1.08%	1.31%

TABLE 3: Hourly forecast of PV power output using LSTM and Bi-LSTM, RMSE, R square, and Cv (RMSE) of the respective model.

Weather	Model	RMSE	R square	Cv (RMSE) %
Sunny	LSTM (2 layers)	0.06	0.99	0.15
	Bi LSTM	0.06	0.99	0.15
Cloudy	LSTM (2 layers)	0.058	0.99	0.21
	Bi LSTM	0.0025	0.99	0.0095
Rainy	LSTM (2 layers)	0.157	0.91	0.60
	Bi LSTM	0.12	0.95	0.54
Partial cloudy	LSTM (2 layers)	0.18	0.81	0.51
	Bi LSTM	0.06	0.99	0.17
Dusty	LSTM (2 layers)	0.18	0.80	0.50
	Bi LSTM	0.08	0.99	0.22
Fog	LSTM (2 layers)	0.17	0.85	0.84
	Bi LSTM	0.072	0.98	0.33

Here, p, d, q are nonseasonal parameters and P, D, Q are seasonal parameters with n being order of seasonality, which could be 4 for the quarter and 12 for annual [37]. Seasonal and nonseasonal parts are quite similar, except backshifts are involved in seasonal time [36]. SARIMA in mathematical representation is given below in Equation (3) [38].

$$\varnothing_p(B)\Phi_P(B^n)W_t = \theta_q(B)\Phi_Q(B^n)\omega_t. \quad (3)$$

As described in Equation (1), p, q notations are nonseasonal parameters. They are the same in Equation (2) also. In Equation (2) the P, Q represent order of seasonal autoregressive (AR) and moving average (MA), respectively, and n is the length of period. AR and MA are mathematically represented in Equations (4) and (5), respectively where Φ is respective weight of lagged values and ε is error at respective lagged values [39]. In Equation (3), the B represents backshifts and ω value for noise at time t [38]. W_t is stationary variable, which can further be mathematically explained as Equation (6) [40]. Equation (7) is obtained by merging the value of W_t (given in Equation (6)) in Equation (3).

$$Y_t = \beta_1 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_P Y_{t-P}, \quad (4)$$

$$Y_t = \beta_2 + \omega_1 \varepsilon_{t-1} + \omega_2 \varepsilon_{t-2} + \cdots + \omega_Q \varepsilon_{t-Q} + \varepsilon_t, \quad (5)$$

$$W_t = (1 - B^n)^D (1 - B)^d X'_t, \quad (6)$$

$$\varnothing_p(B)\Phi_P(B^n)(1 - B^n)^D(1 - B)^d X'_t = \theta_q(B)\Phi_Q(B^n)\omega_t, \quad (7)$$

where n is seasonality length and X'_t is appropriate predifferencing of series X_t (the input value) to ensure constant variance of transformed series over time.

2.3. Implementation of Forecasting Models

2.3.1. Recurrent Neural Networks. This study applied recurrent neural networks to datasets of different weather conditions, sunny days, cloudy days, rainy days, partially cloudy days, dusty days, and foggy days to forecast hourly PV power. The standard LSTM and Bi-LSTM model are available under the *Keras* package that runs on Tensorflow in R programming (R Studio). For PV power forecasting, realtime data in the current study is utilized. For an accurate forecast of PV power output, the real-time power output variable of the dataset is taken into account in the modeling. Then, split the data of PV output into training and validation datasets, with 80% of the data for training and 20% for validation. The models LSTM and Bi-LSTM were trained on the training data and tested on validation data. Min-max normalization method was applied to datasets, and normalized the data between 0 and 1 values, making it easier for models to understand [41]. For LSTM, in this study, four different hidden layers were tested, with the help of different precision methods, the best fit out of these four different models was considered. In this work, number of neurons has been fixed for all datasets based on mean square error [42]. The maximum number of chosen epochs was 100 for current work. Forecasting performance of models is validated by RMSE, R square, and coefficient of variation of RMSE acronymically known Cv(RMSE). The configuration of the layers of RNNs is shown in Table 1 and Figure 3.
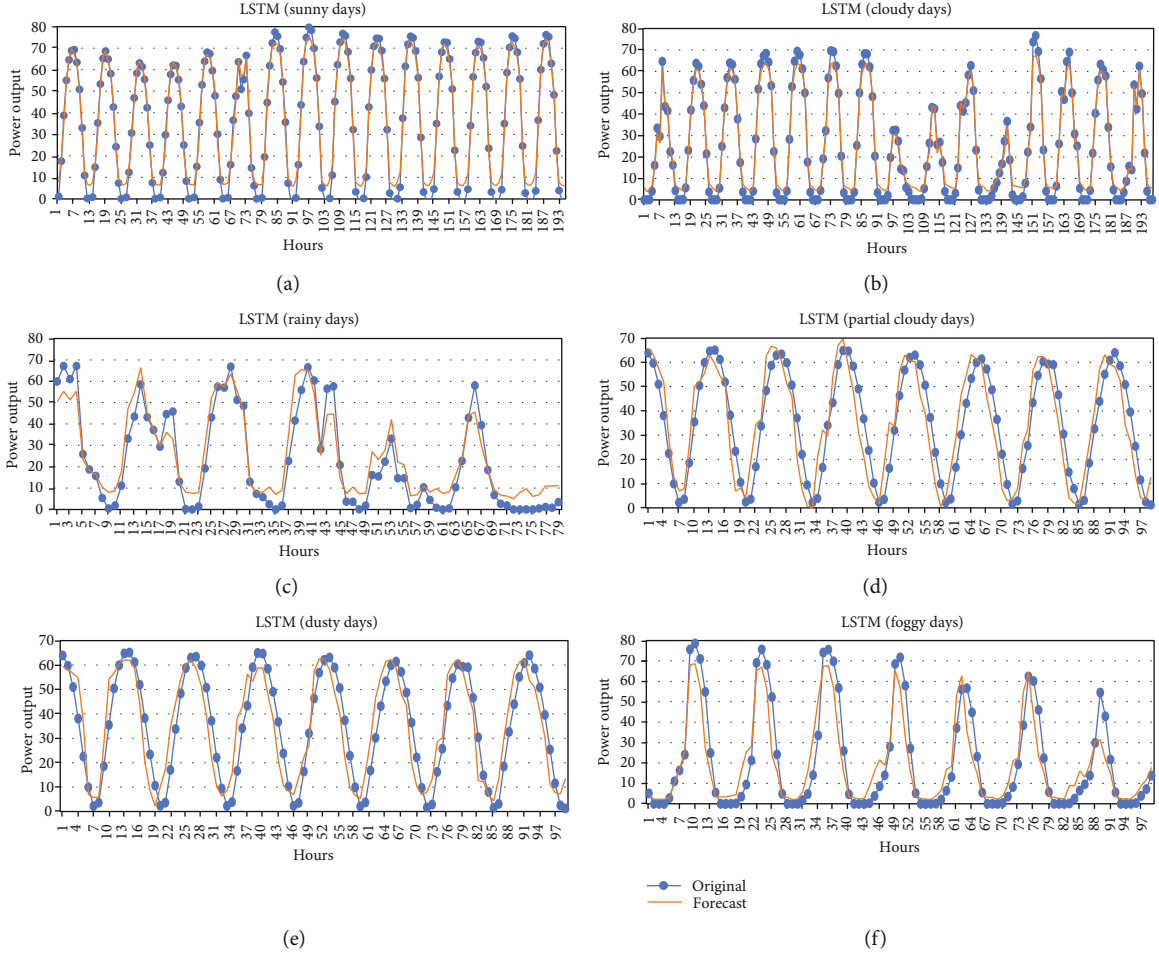
FIGURE 4: LSTM results for (a) sunny days, (b) cloudy days, (c) rainy days, (d) partial cloudy days, (e) dusty days, and (f) foggy days.

2.3.2. Statistical Model. Akaike's information criterion (AIC) is used to make model parameters effective in ARIMA, ARMA, and SARIMA [36]. The seasonality in the dataset is inherent. Therefore, the SARIMA model was adopted in this study. For the PV power output forecast, the PV power output variable from a real-world dataset has been used. Furthermore, the dataset was partitioned into two sets, one for training and one for testing, to achieve accuracy. Training data makes up around 80% of the dataset, whereas testing data makes up about 20% of real-world datasets. AIC estimates model fit, but AIC does not have any sign of unquestionable quality [43]. SARIMA model for this work has been chosen with a low AIC value (9.387). AIC is determined in [29].

$$\text{AIC}\,(p+q) = N\,\log\,(r^2) + 2(p+q), \tag{8}$$

where N represents observations and p and q represents autoregressive and moving average order, respectively.

2.4. Precision Methods. This study adopted the following three precision measures to check the performance of the LSTM and Bi-LSTM in hourly solar PV power forecasts.

$$\text{RMSE} = \sqrt{\sum_{i=1}^{M} \frac{(P_i - O_i)^2}{M}}, \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{M}(P_i - O_i)^2}{\sum_{i=1}^{M}(P_i - O_i)^2}, \tag{10}$$

where M is the number of samples, P is the forecasted value, and O is the original value.

$$\text{Cv(RMSE)} = \frac{\text{RMSE}_{\text{forecast}}}{\text{Average}_{\text{forecast}}} \times 100. \tag{11}$$

The idea is that the lower the Cv(RMSE) and RMSE values, the better is accuracy, and higher the R square value, the better is accuracy.

3. Results and Discussion

Table 2 represents the experimental results of different hidden layers tested for LSTM. Table 3 represents the RMSE, R square, and Cv (RMSE) for LSTM (2 layers) and Bi-LSTM with different weather conditions. LSTM 2-layer model was
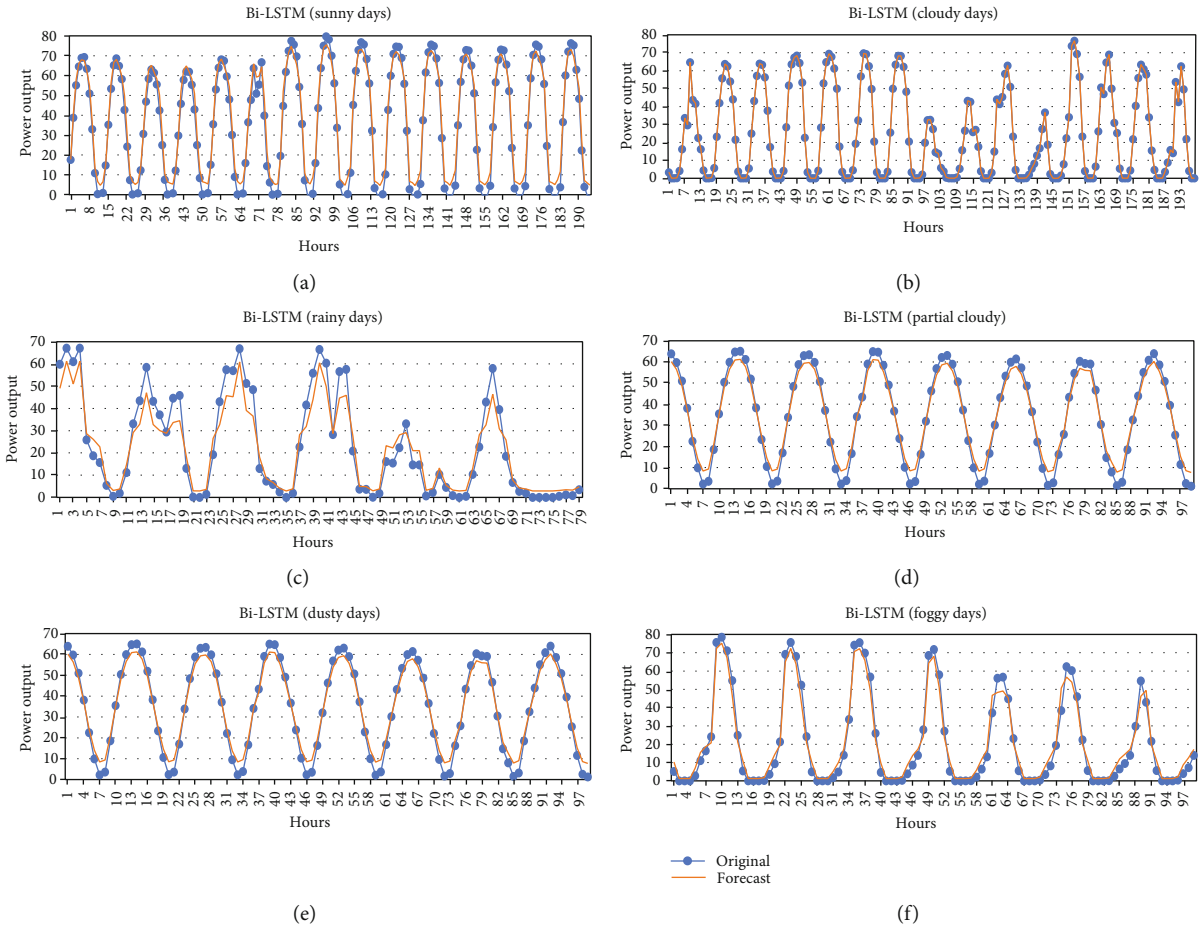
FIGURE 5: Bi-LSTM results for (a) sunny days, (b) cloudy days, (c) rainy days, (d) partial cloudy days, (e) dusty days, and (f) foggy days.

chosen best fit after conducting a multilayer experiment of LSTM (refer to Table 1). It is important to recall that the dataset of all the weather conditions is certainly not the same. All of the weather conditions have a different number of hours.

3.1. Recurrent Neural Networks

3.1.1. Long-Short-Term Memory RNN. Figure 4 shows a graphical illustration of the LSTM model with 2 hidden layers. The overall performance of LSTM is remarkable in graphical representations.

Figure 4(a)shows that sunny day results of the LSTM model are shown. RMSE, R square, and Cv (RMSE) values for sunny days have been found as 0.06, 0.99, and 0.15%, respectively, which justifies model's accuracy. Figure 4(b) shows that cloudy day results of the LSTM model are shown. RMSE, R square, and Cv (RMSE) values for cloudy weather found are 0.058, 0.99, and 0.21%, respectively. Figure 4(c) shows that rainy day results of the LSTM model are shown; some notable deviations in original and forecasted values can be seen in rainy data results. RMSE, R square, and Cv (RMSE) values for rainy data is 0.157, 0.91, and 0.60%, respectively. Figure 4(d) shows that partial cloudy day results of the LSTM model are shown. It defines a few devi-

ations in forecasted and original values. The RMSE, R square, and Cv (RMSE) 0.18, 0.81, and 0.51%, respectively, were found for partial cloudy data. Figure 4(e) shows that dusty day results of the LSTM model are shown. Dusty days observed nearly identical outcomes as partial cloudy days. RMSE, R square, and Cv (RMSE) 0.18, 0.80, and 0.50%, respectively. Figure 4(f) shows that foggy day results of the LSTM model are shown; few notable deviations can be seen in its graph, RMSE, R square, and Cv (RMSE) value for foggy weather stand at 0.17, 0.85, and 0.84%, respectively. Few authors [32, 44] have also used LSTM for forecasting of photovoltaic power, and the RMSE values achieved are within the similar range as obtained in this research study.

3.1.2. Bidirectional LSTM. Figure 5 shows graphical outcomes for the Bi-LSTM model. Figure 5(a) shows that sunny day results of the Bi-LSTM model are shown. It is found that there are minor deviations in original and forecasted values. RMSE, R square, and Cv (RMSE) values of the Bi-LSTM model for sunny day situations are 0.06, 0.99, and 0.15%, respectively. Figure 5(b) shows that cloudy day results of the Bi-LSTM model are shown, in which results of forecasts are pretty just like the original values. RMSE, R square, and Cv (RMSE) values found are 0.0025, 0.99, and 0.0095%, respectively. Bi-LSTM executed greater accuracy than LSTM

TABLE 4: RMSE and R square values for respective weather parameters of SARIMA model.

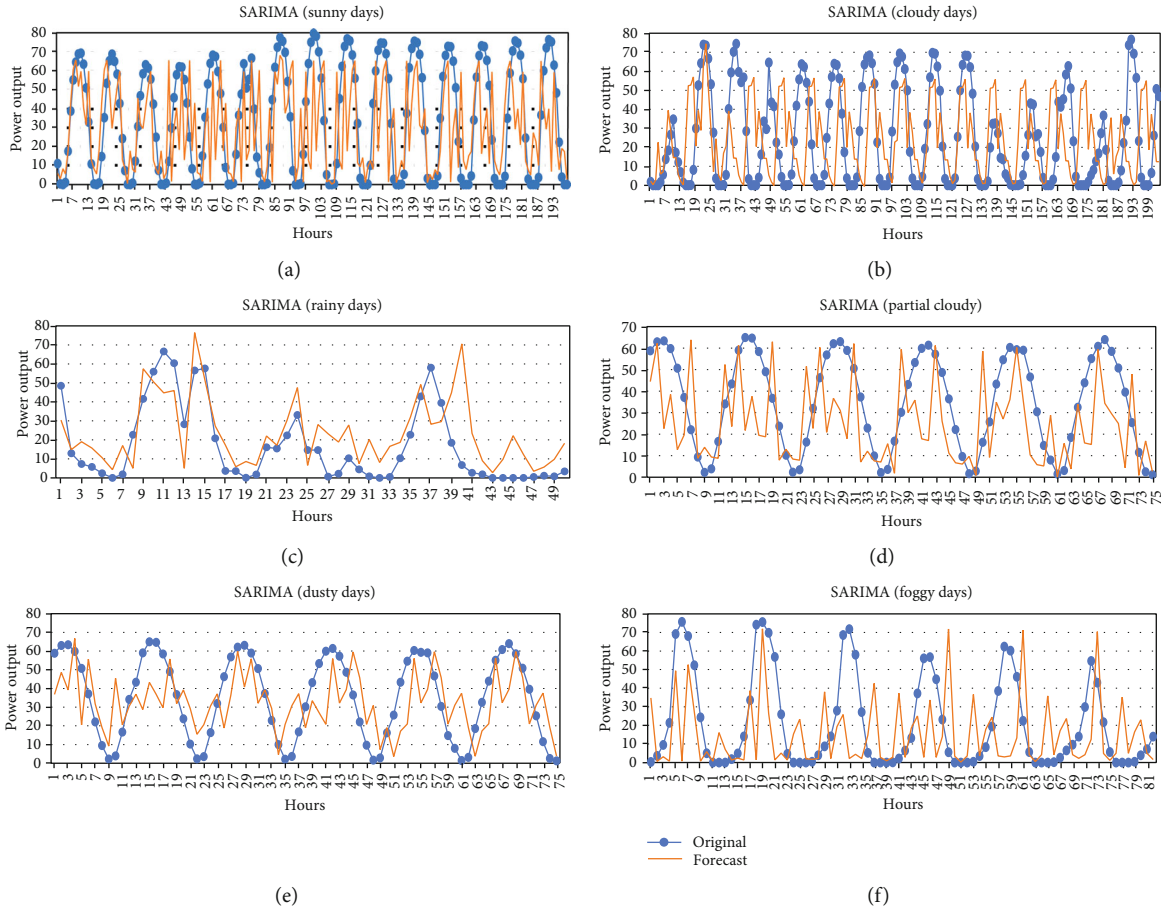| Weather	Model	RMSE	R square	Cv (RMSE) %
Sunny	SARIMA	1.43	0.26	4.57
Cloudy	SARIMA	1.76	0.40	7.03
Rainy	SARIMA	2.545	0.48	10.79
Partial cloudy	SARIMA	1.654	0.229	6.55
Dusty	SARIMA	2.53	0.274	7.75
Fog	SARIMA	3.81	0.19	25.4

(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 6: SARIMA results for (a) sunny days, (b) cloudy days, (c) rainy days, (d) partial cloudy days, (e) dusty days, and (f) foggy days.

on cloudy weather situations. Figure 5(c) shows that rainy day outcomes of Bi-LSTM are shown, a few substantial deviations are found at a few factors of original and forecasted data. RMSE, R square, Cv (RMSE) values found are 0.12, 0.95, and 054%, respectively. Figure 5(d) shows that partial cloudy weather outcomes are shown, the consequences are exceptional, and deviations are negligible; furthermore, RMSE, R square, and Cv (RMSE) values stand at 0.06, 0.99, and 0.17%, respectively. This defines the accuracy of the Bi-LSTM model. Figure 5(e) shows that dusty day results are shown. Bi-LSTM results are excellent, the graph appears pleasant with low deviations. RMSE, R square, and Cv (RMSE) results are 0.08, 0.99, and 0.22%, respectively, which also indicates the accuracy of the model. Figure 5(f) shows
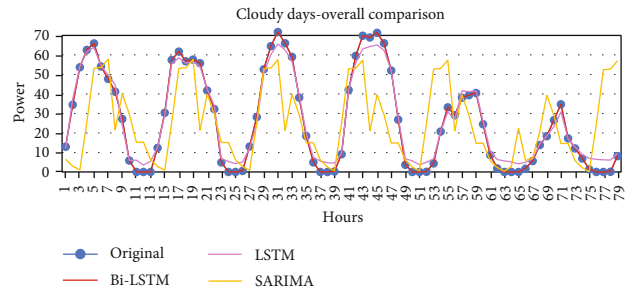

FIGURE 7: Forecasting results of LSTM, Bi-LSTM, and SARIMA.

TABLE 5: Results of LSTM, Bi-LSTM, and SARIMA.

Weather	Model	RMSE	R square	CV (RMSE) %
	LSTM (2 layers)	0.06	0.99	0.15
Sunny	Bi-LSTM	0.06	0.99	0.15
	SARIMA	1.43	0.26	4.57
	LSTM (2 layers)	0.058	0.99	0.21
Cloudy	Bi-LSTM	0.0025	0.99	0.0095
	SARIMA	1.76	0.40	7.03
	LSTM (2 layers)	0.157	0.91	0.60
Rainy	Bi-LSTM	0.12	0.95	0.54
	SARIMA	25.45	0.48	10.79
	LSTM (2 layers)	0.18	0.81	0.51
Partial cloudy	Bi-LSTM	0.06	0.99	0.17
	SARIMA	16.54	0.229	6.55
	LSTM (2 layers)	0.18	0.80	0.50
Dusty	Bi-LSTM	0.08	0.99	0.22
	SARIMA	2.53	0.274	7.75
	LSTM (2 layers)	0.17	0.85	0.84
Fog	Bi-LSTM	0.072	0.98	0.33
	SARIMA	3.81	0.19	25.4

that foggy day results of Bi-LSTM are shown. Very minor deviations too on few of the points are found. RMSE, R square, and Cv (RMSE) value found are 0.072, 0.98, and 0.33%, respectively. The results of those precision techniques endorse accuracy for the Bi-LSTM model.

3.2. Statistical Model. In this study, a statistical forecast method was also taken into consideration for the time series forecast of PV power. The overall performance of RNNs was superior to SARIMA. Table 4 represents the precision outcomes for the SARIMA model, and graphical results also are shown in Figure 6.

3.3. Comparative Analysis. The findings of the deep learning RNN model were found to be the most accurate on a cloudy weather dataset. Figure 7 and Table 5 show a thorough comparison of all the models studied over a dataset of cloudy days.

4. Conclusion

The hourly PV electricity output forecast is vital for operation, maintenance, and overcoming the demanding situations faced by the grid-linked PV plants. A simple statistical model for time series forecasting of hourly solar PV electricity, SARIMA, and the performances of LSTM and Bi-LSTM recurrent neural networks have been examined in this paper. The models have been trained and examined on absolutely distinctive weather parameters.

The study recommends that the RNNs overall performance is superior to that of the SARIMA model. The RNNs have deep systems to remedy intense troubles which include vanishing gradients; for this reason, they have carried out

forecasting with such excessive accuracy. Furthermore, in the assessment of 2 distinctive RNNs, findings recommend bidirectional long-short-term memory (Bi-LSTM) carried out greater accuracy than long-short-term memory (LSTM). It is found that Bi-LSTM scored higher R squares in each weather condition, whilst it remained lowest 0.95 on rainy days. The RMSE and Cv (RMSE) values for Bi-LSTM have been also recorded lowest in all weathers; RMSE was observed maximum at rainy weather with 0.12 value, and maximum Cv (RMSE) for Bi-LSTM was found at rainy data with 0.54%. The graphical representations of Bi-LSTM (shown in Figure 4) also propose that actual values are very near the forecasted values under distinctive weather conditions. Hence, for short time forecasting of the power output of grid-linked PV power plant, this paper suggests bidirectional LSTM recurrent neural network as exceptional model with high accuracy.

Nomenclature

PV: Photovoltaic
ANN: Artificial neural network
RNN: Recurrent neural network
WB: World Bank
LSTM: Long-short-term memory
Bi-LSTM: Bidirectional long-short-term memory
RMSE: Root mean square error
Cv (RMSE): Coefficient of variation of room mean square error
ARIMA: Autoregressive integrated moving average
SARIMA: Seasonal autoregressive integrated moving average
AIC: Akaike's information criterion.

Data Availability

Data are available upon request to submitting or the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] NEPRA, *Annual report annual report 2019-20*, Islamabad, Pakistan, 2019https://nepra.org.pk/publications/Annual.Reports.ph.

[2] NEPRA, "State of industry report 2020," *Regulation*, vol. 53, no. 9, pp. 1689–1699, 2020.

[3] "Solar - fuels & technologies - IEA," June 2021, https://www.iea.org/fuels-and-technologies/solar.

[4] World Bank, "Expanding renewable energy in Pakistan's electricity mix," 2020, September 2021, https://www.worldbank.org/en/news/feature/2020/11/09/a-renewable-energy-future-for-pakistans-power-system.

[5] "How environmental factors affect solar panel performance - naked solar," September 2021, https://nakedsolar.co.uk/solar-pv/environmental-factors/.

[6] M. A. Al-Nimr, S. Kiwan, and H. Sharadga, "Simulation of a novel hybrid solar photovoltaic/wind system to maintain the cell surface temperature and to generate electricity," *International Journal of Energy Research*, vol. 42, no. 3, pp. 985–998, 2018.

[7] K. N. Nwaigwe, P. Mutabilwa, and E. Dintwa, "An overview of solar power (PV systems) integration into electricity grids," *Materials Science for Energy Technologies*, vol. 2, no. 3, pp. 629–633, 2019.

[8] Z. Abbas, K. Harijan, P. H. Shaikh, G. D. Walasai, and F. Ali, "Effect of ambient temperature and relative humidity on solar PV system performance: a case study of Quaid-e-Azam Solar Park, Pakistan," *Sindh University Research Journal-SURJ (Science Series)*, vol. 49, no. 4, pp. 721–726, 2017.

[9] Z. Abbas, K. Harijan, P. Hameed, and F. Bhayo, "Effect of dust on the performance of photovoltaic system (a case study of Quaid-E-Azam Solar Park Bahawalpur)," *Sindh University Research Journal-SURJ (Science Series)*, vol. 1, no. 6, pp. 73–79, 2017.

[10] A. A. Khosa, T. U. Rashid, N. U. H. Shah, M. Usman, and M. S. Khalil, "Performance analysis based on probabilistic modelling of Quaid-e-Azam Solar Park (QASP) Pakistan," *Energy Strategy Reviews*, vol. 29, article 100479, 2020.

[11] M. H. Aziz, "Performance evaluation of an on grid 100 MW Quaid-E-Azam Solar Power Bahawalpur," 2018, https://www.sciencedirect.com/science/article/pii/S2211467X2030033X.

[12] A. Iqbal, *Pakistan will shift to 60pc clean energy by 2030, world assured - newspaper - DAWN.COM*, Dawn.Com, 2021.

[13] A. G. Salman and B. Kanigoro, "Visibility forecasting using autoregressive integrated moving average (ARIMA) models," *Procedia Computer Science*, vol. 179, no. 2019, pp. 252–259, 2021.

[14] P. Chen, A. Niu, D. Liu, W. Jiang, and B. Ma, "Time series forecasting of temperatures using SARIMA: an example from Nanjing time series forecasting of temperatures using SARIMA: an example from Nanjing," *IOP Conference Series: Materials Science and Engineering*, vol. 394, no. 5, article 052024, 2018.

[15] Y. Yu, J. Cao, and J. Zhu, "An LSTM short-term solar irradiance forecasting under complicated weather conditions," *IEEE Access*, vol. 7, pp. 145651–145666, 2019.

[16] J. Zhong, L. Liu, Q. Sun, and X. Wang, "Prediction of photovoltaic power generation based on general regression and back propagation neural network," *Energy Procedia*, vol. 152, pp. 1224–1229, 2018.

[17] Y. Jung, J. Jung, B. Kim, and S. U. Han, "Long short-term memory recurrent neural network for modeling temporal patterns in long-term power forecasting for solar PV facilities: case study of South Korea," *Journal of Cleaner Production*, vol. 250, article 119476, 2020.

[18] L. Liu, D. Liu, Q. Sun, H. Li, and R. Wennersten, "Forecasting power output of photovoltaic system using a BP network method," *Energy Procedia*, vol. 142, pp. 780–786, 2017.

[19] J. Shi, W. J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *IEEE Transactions on Industry Applications*, vol. 48, no. 3, pp. 1064–1069, 2012.

[20] P. Ramsami and V. Oree, "A hybrid method for forecasting the energy output of photovoltaic systems," *Energy Conversion and Management*, vol. 95, pp. 406–413, 2015.

[21] Z. Pang, F. Niu, and Z. O'Neill, "Solar radiation prediction using recurrent neural network and artificial neural network: a case study with comparisons," *Renewable Energy*, vol. 156, pp. 279–289, 2020.

[22] D. Lee and K. Kim, "PV power prediction in a peak zone using recurrent neural networks in the absence of future meteorological information," *Renewable Energy*, vol. 173, pp. 1098–1110, 2021.

[23] B. Brahma and R. Wadhvani, "Solar irradiance forecasting based on deep learning methodologies and multi-site data," *Symmetry*, vol. 12, no. 11, article 1830, 2020.

[24] M. K. Park, J. M. Lee, W. H. Kang, J. M. Choi, and K. H. Lee, "Predictive model for PV power generation using RNN (LSTM)," *Journal of Mechanical Science and Technology*, vol. 35, no. 2, pp. 795–803, 2021.

[25] F. Shaikh and Q. Ji, "Forecasting natural gas demand in China: logistic modelling analysis," *International Journal of Electrical Power & Energy Systems*, vol. 77, pp. 25–32, 2016.

[26] F. Shaikh, Q. Ji, P. H. Shaikh, N. H. Mirjat, and M. A. Uqaili, "Forecasting China's natural gas demand based on optimised nonlinear grey models," *Energy*, vol. 140, pp. 941–951, 2017.

[27] *Quaid-e-Azam 1000MW Solar Park (Bahawalpur) Quaid-e-Azam*China-Pakistan Economic Corridor (CPEC) Official WebsiteJune 2021, http://cpec.gov.pk/project-details/10.

[28] Cuemath.com, "Average formula - what is the average formula? Examples," September 2021, https://www.cuemath.com/average-formula/.

[29] H. Sharadga, S. Hajimirza, and R. S. Balog, "Time series forecasting of solar power generation for large-scale photovoltaic plants," *Renewable Energy*, vol. 150, pp. 797–807, 2020.

[30] C. Olah, "Understanding LSTM networks – colah's blog," 2015, September 2021, https://colah.github.io/posts/2015-08-Understanding-LSTMs/.

[31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[32] M. Abdel-Nasser and K. Mahmoud, "Accurate photovoltaic power forecasting models using deep LSTM-RNN," *Neural*

Computing and Applications, vol. 31, no. 7, pp. 2727–2740, 2019.

[33] B. Iung, "Cœur et grossesse," *EMC – Trait médecine AKOS*, vol. 8, no. 2, pp. 1–4, 2013.

[34] M. AlKandari and I. Ahmad, "Solar power generation forecasting using ensemble approach based on deep learning and statistical methods," *Applied Computing and Informatic*, 2020.

[35] A. Graves, A. R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, pp. 6645–6649, Vancouver, BC, Canada, 2013.

[36] P. Chen, A. Niu, D. Liu, W. Jiang, and B. Ma, "Time series forecasting of temperatures using SARIMA: an example from Nanjing," *IOP Conference Series: Materials Science and Engineering*, vol. 394, no. 5, 2018.

[37] R. J. Hyndman and G. Athanasopoulos, *8.9 seasonal ARIMA models | forecasting: principles and practice*, O Texts, 2nd edition, 2018.

[38] M. Farsi, D. Hosahalli, B. R. Manjunatha et al., "Parallel genetic algorithms for optimizing the SARIMA model for better forecasting of the NCDC weather data," *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 1299–1316, 2021.

[39] G. Shmueli and K. C. Lichtendahl Jr., "Practical time series forecasting with R: a hands-on guide," 2015, http://www.amazon.com/Practical-Time-Series-Forecasting-Hands-On/dp/0991576632.

[40] J. Flaherty and R. Lombardo, "Modelling private new housing starts in Australia," in *A paper presented in the Pacific-Rim Real Estate Society conference*, pp. 24–27, University of Technology Sydney, Australia, 2000.

[41] C. Martínez, "Neural networks for business analytics with R | Udemy," 2021, https://www.udemy.com/course/neural-networks-for-business-analytics-with-r/learn/practice/1302928?start=start-page#questions.

[42] A. Mellit, S. Sağlam, and S. A. Kalogirou, "Artificial neural network-based model for estimating the produced power of a photovoltaic module," *Renewable Energy*, vol. 60, pp. 71–78, 2013.

[43] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[44] D. Lee and K. Kim, "Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information," *Energies*, vol. 12, no. 2, 2019.