IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 January 2024, accepted 27 February 2024, date of publication 4 March 2024, date of current version 22 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3373312

RESEARCH ARTICLE

Prediction of Solar PV Power Using Deep Learning With Correlation-Based Signal Synthesis

M. DILSHAD SABIR[1], KAMRAN HAFEEZ[1], SAMERA BATOOL[2], GHANI AKBAR[3], LAIQ KHAN[1], GHULAM HAFEEZ[4], AND ZAHID ULLAH[5,6], (Graduate Student Member, IEEE)

[1]Department of Electrical and Computer Engineering, COMSATS University Islamabad, Islamabad 44000, Pakistan
[2]Department of Computer Science, COMSATS University Islamabad, Islamabad 44000, Pakistan
[3]Climate, Energy, and Water Research Institute, National Agricultural Research Centre, Islamabad 44000, Pakistan
[4]Department of Electrical Engineering, University of Engineering & Technology Mardan, Mardan 23200, Pakistan
[5]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milan, Italy
[6]Department of Electrical Engineering, University of Management and Technology Lahore, Sialkot Campus, Sialkot 51310, Pakistan

Corresponding author: Zahid Ullah (zahid.ullah@polimi.it)

ABSTRACT Enhancement of the dispatching capacity and grid management efficiency requires knowledge of photovoltaic power generation beforehand. Intrinsically, photovoltaic power generation is highly volatile and irregular, which impedes its prediction accuracy. This paper proposes deep learning-based approaches and a pre-processing algorithm to handle these constraints. The proposed scheme employs Pearson's Correlation Coefficient to find the similarity between atmospheric variables and PV power generation. Based on high PCC values, top atmospheric variables and PV power generated time series data are passed through the Empirical Mode Decomposition (EMD) to simplify the complex data streams into Intrinsic Mode Functions (IMFs). Further, to streamline the prediction process, the proposed correlation-based signal synthesis (CBSS) algorithm finds combinations of these IMFs, which have a high correlation value between atmospheric variables and PV power data. Deep learning models of algorithms Long Short Term Memory (LSTM) and Nonlinear Autoregressive Network with Exogenous Inputs (NARX) network with the configurations of three networks, a single network, and the direct approach employed for the prediction of IMFs combinations. The LSTM network was analyzed under the Adaptive moment estimation (ADAM), Stochastic Gradient Descent with Momentum (SGDM), and Root Mean Square Propagation (RMSP) optimization. Extensive experimentation was evaluated using atmospheric data from the Climate, Energy, and Water Research Institute (CEWRI), NARC, Islamabad, Pakistan. RMSE, MAE, MAPE, and R^2 performance measures show promising prediction results for the LSTM under the configuration of three networks and ADAM optimization.

INDEX TERMS Photovoltaic power prediction, Pearson correlation coefficient, empirical mode decomposition, correlation-based signal synthesis, long short-term memory, nonlinear autoregressive exogenous.

NOMENCLATURE

RIMA	Regressive Integrated Moving Average.
SVM	Support Vector machine.
WPD	Wavelet packet decomposition.
LSTM − DGM	Long Short-Term Memory -Deep Galerkin Method.
AD − LSTM	Antidecay long short-term memory.
CLSTM	Convolution long short term memory.
MRTPP	Multiple relevant and target variables prediction pattern.
MBE	Mean Bias Error.
MAE	Mean Absolute Error.
MAPE	Mean Absolute Percentage Error.
MSE	Mean Squared Error.
MPE	Main Percentage Error.
MRE	Mean Relative Error.

The associate editor coordinating the review of this manuscript and approving it for publication was Enamul Haque.

NRMSE	Normalized Root Mean Square Error.
NMAE	Normalized Mean Absolute Error.
MLP	Multi-layer perception.
IMF	Intrinsic Mode Function.
$E_l(t)$	Lower envelope of signal.
$E_u(t)$	Upper envelope of signal.
$\nabla L(\theta_i)$	Gradient of loss function.
m	Momentum.
θ_i	Training parameter i.

I. INTRODUCTION

Due to limitations in fossil fuels reserves and environmental constraints, renewable energy sources (RESs) are gaining popularity and expansion at a fast rate [1], [2]. Amongst these RESs, solar energy has emerged as a significant resource due to its continuity and pollution-free status [3]. Therefore, any developing country needs to increase its solar energy shares in a combined energy generation mix [4]. The frequent disruption in grid power and unavailability of large-scale grid infrastructure are limitations in providing electricity to a dense population [5]. Due to plenty of sunlight, solar PV can generate sufficient electric power and offer on-grid and off-grid solutions. However, unlike conventional power generation sources, solar PV depends on limited controllable variables. Since solar energy is inconsistent and varying, heavily dependent on weather data [6]. This data includes solar irradiance, relative humidity, temperature, and wind speed [7]. These weather-related factors cause instability and fluctuations in output power in grid-connected and stand-alone solar PV systems [8]. Therefore, precise and efficient prediction of solar PV is important in the smooth functioning and regulation of power systems [9]. Accuracy in solar PV prediction is a key challenge in interconnecting large-scale solar power to a conventional grid system [10].

A Unified Power Quality Conditioner (UPQC) is proposed to overcome quality issues of grid-connected PV systems [11], [12]. However, the connection of solar power to the main grid can be done smoothly by estimating the power obtained from the solar plants. Then, grid operators can utilize the predicted solar power for planning and decision processes efficiently [13]. The interdependence between climate input data and output power is considered. It is shown that solar irradiance has a strong correlation between output power and predicting PV power [14]. However, the more the number of input data variables, the accuracy of predicting the model increases but with an increase in complexity and computational cost [15]. Therefore, selecting important parameters relating to climate data using correlation is needed to achieve high accuracy with minimum computational cost.

In literature, solar PV prediction models can be classified as statistical, physical, and AI (Artificial Intelligence) based models [16]. Physical models forecast solar irradiance and output power using geographical and weather data [17]. These parameters include wind velocity, ambient temperature, humidity, and air pressure. This approach is directly related to the accuracy of climate data without considering previous solar data [18]. The statistical models co-relate time series and real-time data to predict future scenarios. This approach needs historical data but computational problems and accuracy in data remain a challenging issue [19]. AI methods include ANN, support vector machine (SVM), adaptive neuro-fuzzy interface system (ANFIS), LSTM, deep belief network (DBN), RNN, and CNN-based deep learning models are suggested for the prediction of solar power output [20]. LSTM deep learning model is implemented to forecast using time series data. This approach becomes difficult due to cyclic and seasonal variations in the data [21]. To predict solar irradiance and power a hybrid model is suggested using ANN [22]. An indirect deep learning method is implemented using the LSTM approach to estimate solar PV power from solar irradiance with the help of weather data [23]. The solar irradiance is predicted using the LSTM model with sunshine hours, humidity, and temperature as inputs. The proposed model results are compared with SVR (support vector regression). The LSTM model improves accuracy by evaluating root mean error criteria [24]. A multi-horizon prediction of solar irradiance with an LSTM model using inputs; irradiance, pressure, temperature, and wind speed is also implemented [25]. Different deep learning models, such as LSTM, SVR, RNN, and GRU (Gated Recurrent Unit) are implemented to predict solar irradiance with good accuracy [26]. In a comparative study between deep learning networks and machine learning-based models; Gradient Boosted Regression Tree (GBRT) and Feed Forward Neural Network (FFNN) are evaluated [27]. The physical models are directly related to climate parameters in predicting solar power, whereas AI-based methods can overcome these issues for short and long-term prediction [28]. The AI-based methods have limitations like higher computational cost and lower performance while handling large-size data [29]. Machine learning models can extract less complex features from multi-dimension data [30]. Hybrid models that combine several deep learning methods are also gaining importance to achieve better results in predicting solar power [31]. A hybrid model is developed using CNN and LSTM to forecast solar irradiance based on solar angle, wind speed, perceptible water, wind direction, and temperature parameters [32].

The research work in literature based on solar power forecasting related to machine learning models, ANN, RNN, and CNN is summarized in Table 1.

A. MOTIVATION AND RESEARCH CHALLENGES

The following motivation and key scientific challenges are associated with predicting PV power. These challenges require further investigation.

1) ***Energy Crisis*** Fossil fuel reserves are depleted in nature, unfriendly to the environment, and have severe implications for climate change. Environmental-friendly renewable energy has gotten a great deal of

TABLE 1. Comparative analysis of related literature review.

Reference	Proposed method	Forecast Horizon	Input parameters	Time for testing	Performance criteria
[33]	Hybrid RIMA-SVM	24h	Solar irradiance ambient temperature	4 months	2.7% (MPE) 9.40% (NRMSE)
[34]	ANN & Time-Series Data	1 week	Cloudiness, temperature, precipitation, and humidity	24 months	4.7% (MAPE)
[35]	WPD-LSTM	60 min	Horizontal radiation, diffuse ambient temperature, wind speed, and relative humidity	2 years	2.5% (MAPE)
[36]	LSTM-CNN	60 min	temperature Relative humidity, pressure, wind speed, and cloud type	4 years	27.3 Wm2 (MAE)
[37]	LSTM & LSTM-DGM	24 h	Solar irradiance, air temperature, relative humidity, wind speed, cloud, air pressure	12 months	4.6% (RMSE)
[38]	LSTM & ANN	4 months	Temperature, humidity, cloudiness, radiation,	38 months	LSTM 1.82% (RMSE) ANN; 8.02% (RMSE)
[39]	RNN and ANNN	60 min	solar radiation, dew-point temperature, humidity, wind speed, and wind direction	1 week	26% (RMSE) 0.2% (NMBE)
[40]	CLSTM	24 h	Solar irradiance	12 years	1.515% (RMSE) 4.672% (MAE)
[41]	ALSM & MRTPP	1 Hour	Wind speed, wind direction, radiation, rainfall, power, temperature, and humidity	2 years	97% (R2) 4.% (NMAE) 6.4% (NRMSE)

attention recently. Current energy crises, such as imbalance between supply and demand and overcoming power shortages, can be resolved using these renewable energy technologies [42]. Solar energy has achieved around 849 GW capacities, featuring almost 28%

2) ***Renewable Energy Resources*** Solar PV is considered an encouraging replacement for fossil fuels. It is turbulent and irregular as it is directly related to climate data. Due to its fluctuating property, there is an imbalance between the supply and demand of energy. Therefore,

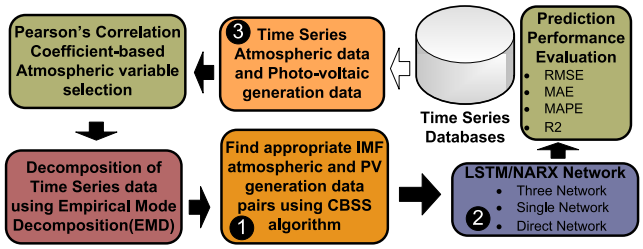


FIGURE 1. Optimization flow diagram of proposed deep learning-based PV power generation prediction.

solar energy prediction is highly recommended to maintain accurate supply and demand of solar power output and maintain a stable operation [43]. Large-scale connections of solar PV into the grid pose challenges and stability problems to the working of conventional power grids. Accurate prediction of solar PV will be effective in solving these issues.

3) ***Exploration of Deep Learning Techniques*** Recently, machine learning, recurrent neural networks (RNN), artificial neural networks (ANN), and convolution neural networks (CNN) are introduced to predict solar PV power. Deep learning models can perform well in solar irradiance prediction [36]. However, these methods are still early and need further exploration [44]. These deep learning approaches have several limitations, including slow convergence and efficiency degradation. The relationship between solar radiation and climate parameters has inspired different authors to predict solar power accurately but mostly relies on the linear relationship between different variables and selecting those inputs, achieving a minimum forecasting error [45].

Nowadays, Machine learning (ML) and deep learning (DL) methods, such as RNNs and LSTM, to forecast solar PV are extensively used [46]. The results of these studies depend on several factors. It includes climate conditions and forecasting horizon [47]. However, the focus should be on data selection and its effect on different models that can improve the accuracy of prediction results. Comparing several models can provide better forecast results compared to an individual network.

B. NOVEL CONTRIBUTIONS

Different weather conditions are considered to predict the highly fluctuated PV power. However, a correlation-based signal synthesis algorithm (CBSS) has been proposed to find the interdependence between Intrinsic Mode Function combinations of atmospheric variables and output PV power generation data. The following contributions are made to address the challenges mentioned earlier:

- A correlation-based signal synthesis algorithm (CBSS) has been proposed **(Figure 1 step 1)** to find the highly correlated IMF combinations of atmospheric variables with IMF combinations of PV power generation data.
- Deep learning approaches like the LSTM and NARX networks have been employed **(Figure 1 step 2)** in three different configurations of three networks, a single network, and a direct approach i.e., training on atmospheric and PV power data without any time series signal decomposition. Further, extensive experimentation is performed using three different optimization schemes of Stochastic Gradient Decent Momentum (SGDM), Adaptive moment estimation (ADAM), and Root Mean Square Propagation (RMSP) for the LSTM network.
- The consequent analysis demonstrates a high prediction rate performance **(Figure 1 step 3)** advantage as compared to the direct approach on collected atmospheric and PV power generated data.

Different weather conditions were pre-processed before a prediction by deep learning schemes of LSTM and NARX to forecast the highly fluctuated and volatile generation of PV power. To improve the forecast accuracy, the proposed algorithm, CBSS is employed to find the appropriate and related combinations of EMD components of vital weather variables and PV power data. The proposed technique is applied to locally collected data for different atmospheric sensors. For the CBSS-LSTM integrated proposed model, a three network composition achieved RMSE of 8.17, R^2 value of 0.99, MAE value of 3.3, and MAPE value of 2.72 for ADAM optimization. For the LSTM model and optimization, the direct approach without any CBSS pre-processing achieved an RMSE of 27.11, R^2 value of 0.90, MAE value of 15.61, and MAPE value of 9.72. For the NARX network, a three-network configuration achieved RMSE of 8.24, R^2 value of 0.96, MAE value of 1.46, and MAPE value of 2.63. For the NARX, the direct approach without any CBSS pre-processing achieved an RMSE of 25.43, R^2 value of 0.88, MAE value of 16.55, and MAPE value of 9.79.

The proposed work is divided into different sections in this paper. Each section provides the details of the proposed techniques. The introduction, related motivation, and research challenges are discussed in section I, which also represents the related literature review regarding the PV prediction. The different experimental streams and the proposed model are detailed in section II. Section III provides the experimental results and discussion details that include performance measures, different optimization schemes, and network configuration for deep learning approaches. The last section concludes the research work.

II. PREDICTION FRAMEWORK

A. OVERVIEW

The proposed methodology predicts photo-voltaic power generation using deep learning approaches. These approaches involve daily weather variables data like maximum temperature, pan evaporation, and humidity. However, multiple streams are employed to verify the effectiveness of the proposed scheme. Each stream encloses either the NARX or LSTM network as a basic deep-learning technique for prediction; however, additional preprocessing is an option

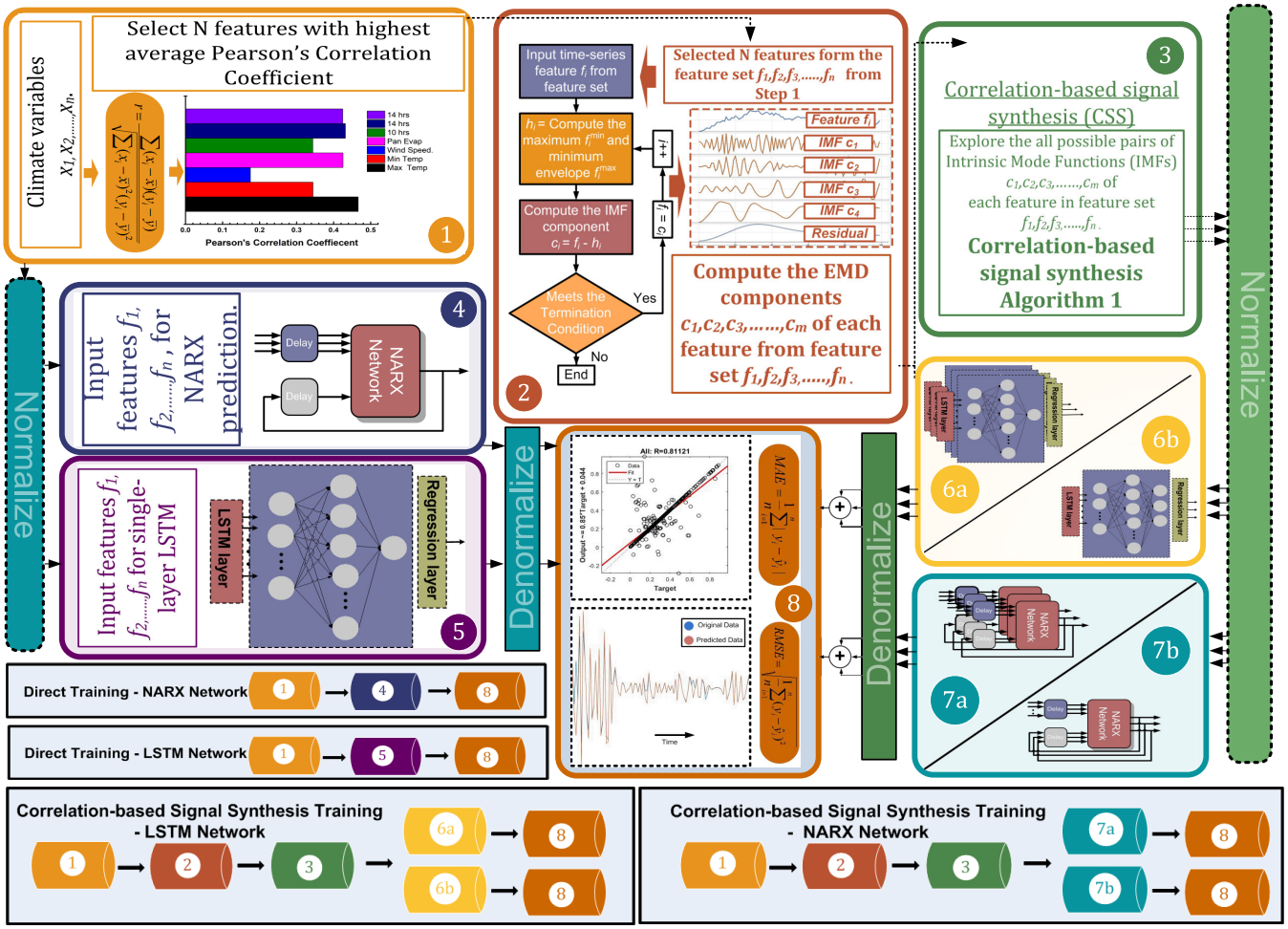


FIGURE 2. Complete block diagram of the step-by-step implementation of the proposed prediction technique.

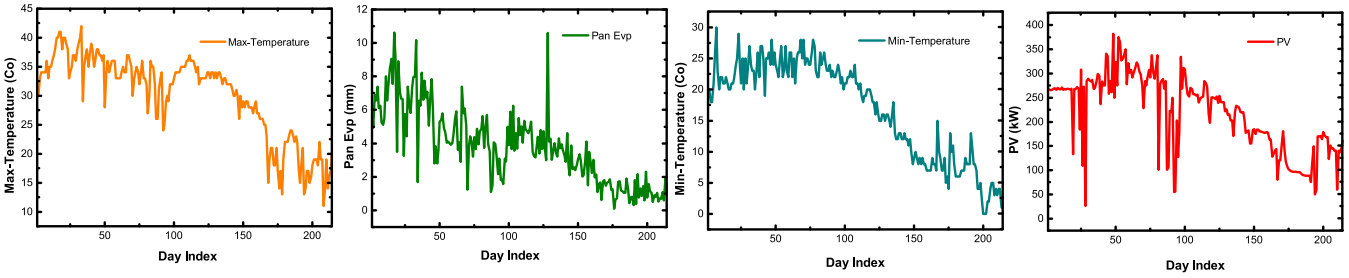


FIGURE 3. Graphs of correlation-based chosen weather variables maximum temperature pan evaporation, minimum temperature, and photo-voltaic power generation data.

that can further simplify the prediction process. Figure 2 provides the step-wise details of each stream separately. To present a comprehensive comparison, Figure 2 provides the complete block diagram of each experimental stream. The initial phase **(Figure 2 step 1)** measures the similarity between the weather variables and PV data using Pearson's correlation coefficient (r). It is much more convenient for the predictor function to have a similarity between the weather variables and PV data. This stage specifies the variables with the highest average correlation value with PV data and employs these variables to further process them for prediction. In the case of direct prediction **(Figure 2 step 4)**, the NARX network utilizes the prior selected weather variables along the PV data for prediction. A single-layer LSTM Network **(Figure 2 step 5)** is employed for prediction using previous weather variables. This network

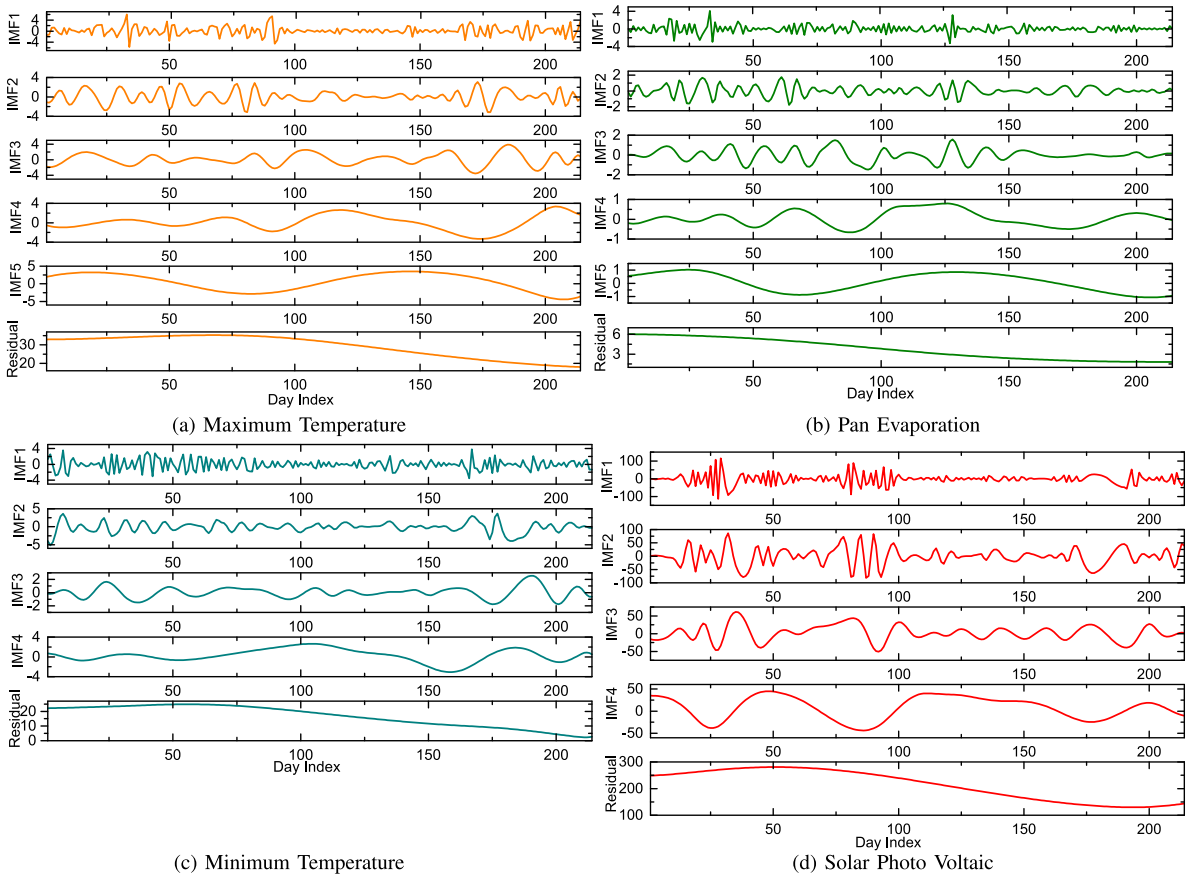


FIGURE 4. Intrinsic Mode Functions (IMFs) and residuals of atmospheric recordings of maximum and minimum temperature, pan evaporation, and photo-voltaic power data by Empirical Mode Decomposition (EMD).

contains a fully connected layer and a regression layer for forecasting.

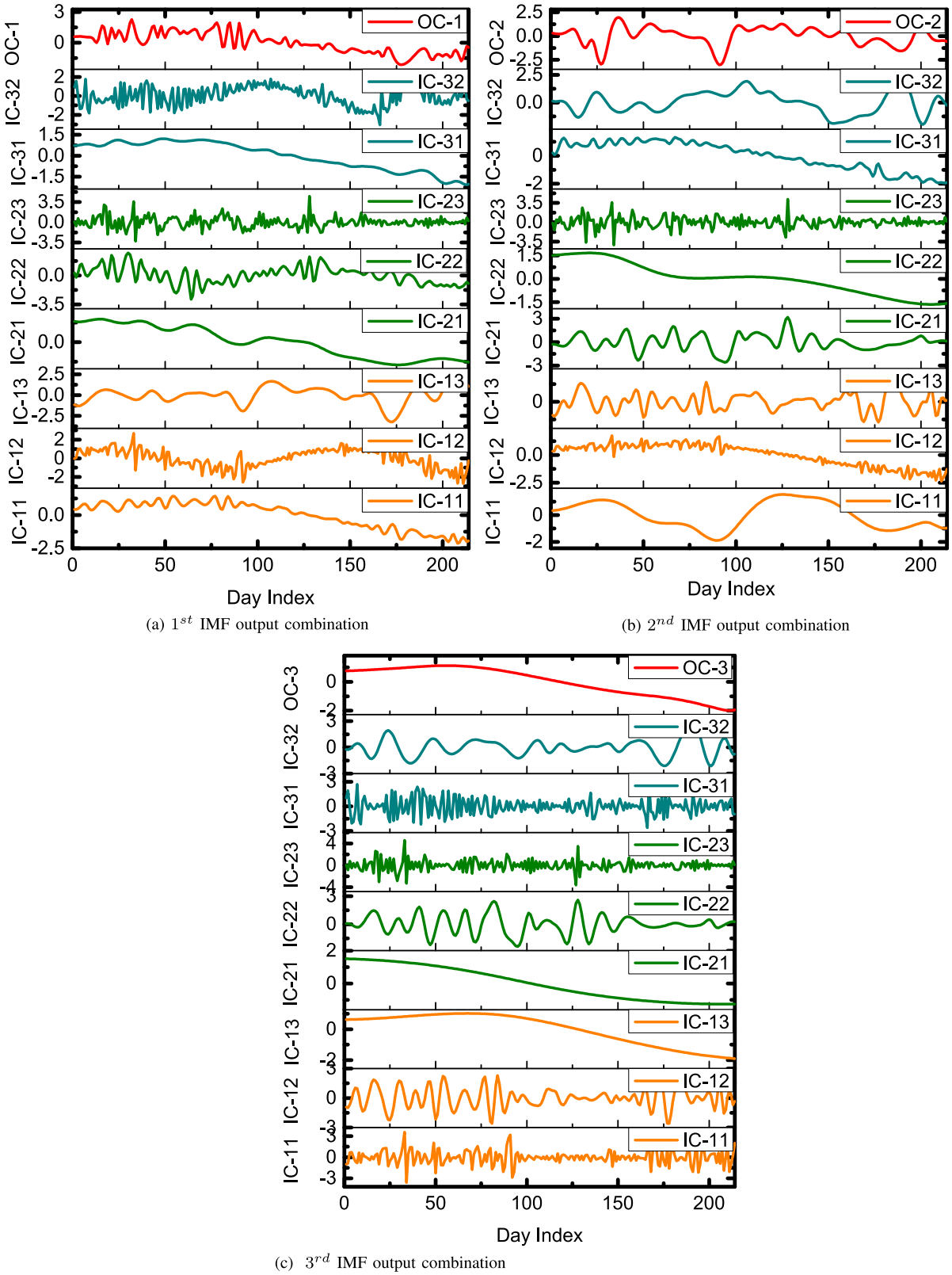
Empirical Mode Decomposition (EMD) employs these selected features in step 1 to give intrinsic mode functions (IMFs) and residual of each weather variable and PV generation **(Figure 2 step 2)**. The IMFs are further fed into the proposed correlation-based signal synthesis (CBSS) algorithm **(Figure 2 step 3)** to find the newly added IMFs of PV and weather variables having the highest correlation. The synthesized data is given to a single-layer LSTM Network **(Figure 2 step 6a)** holding three outputs and three single-layer LSTM Network **(Figure 2 step 6b)** holding one output. Similarly, newly constructed data is presented to a NARX network **(Figure 2 step 7a)** with three outputs and three NARX networks **(Figure 2 step 7b)** with three outputs. Finally, for each stream, evaluation **(Figure 2 step 8)** is used to determine the prediction accuracy. Conventional methods like Mean Square Error (MSE) and Mean Average Error (MAE) are employed for the performance evaluation.

B. CORRELATION-BASED SIGNAL SYNTHESIS (CBSS)

A strong correlation between input and output prediction data suggests any change in input data is related to predictable changes in output data that can improve predictable accuracy [48]. Application of Person's correlation formula results in selecting climate variables displayed in Figure 3. All the possible IMFs of these selected variables are shown in Figure 4. To find the strong impact of the possible IMF combinations of input data on the possible IMF combinations of photovoltaic power generation data, the dot product of both has been considered to analyze the dependency of IMF combinations of atmospheric input data and target PV power generation data. A scalar product or dot product identifies the closeness of two data vectors or variables. This measure supports understanding the impact of one variable on another variable. Element-wise multiplication and addition of two vectors are performed to compute their scaler product. Further, a R ratio is calculated by dividing this product by the absolute value of two vectors.

$$R = \frac{X.P}{|X||P|} \tag{1}$$

X is the IMF combination of weather variables and P is the IMF combination of PV power generation data. The measure results in a range from $+1$ to -1. It indicates a strong correlation if the value lies near $+1$, while -1 suggests a negative correlation. However, near zero or zero specifies no correlation between the two variables. This analysis aims to find the possible IMF combinations that maximize the



(a) 1^{st} IMF output combination

(b) 2^{nd} IMF output combination

(c) 3^{rd} IMF output combination

FIGURE 5. IMF combinations of PV power generation along with its corresponding maximum temperature, pan evaporation, and minimum temperature weather variable's IMF combinations.

correlation between input weather variables and PV power generation data.

The proposed algorithm 1 in Appendix aims to search for possible combinations of IMFs of selected input weather conditions that maximize the correlation value with possible IMF combinations of PV output data. Therefore, the first step is to acquire the pairs of each IMF of each weather variable, including the PV generation data. Algorithm X obtains these combinations from lines 1 to 9. Line 5 computes every sum of IMF's time series. To find the correlation between each IMF combination of each weather condition and each combination of PV generation data, algorithm 1 utilizes lines 10 to 17. Line 13 specifically computes the correlation values between these combinations. Then, line 15 selects the correlation values with the top three highest values. However, that selection is based on the condition that corresponding pairs should have unique IMF components. It is to ensure that each IMF should contribute equally to prediction. For each output IMF combination, lines 18 to 20 sum the selected correlation values at line 15 across all input IMF combinations of weather variables. It results in a sum value of Sy for each output. For lines 21 to 27, the algorithm selects the quotient (Ny divided by 2) number of combinations and its corresponding indics if the total number of combinations of PV data is odd. The *else* line 25 selects the $\frac{Ny}{2}$ number of combinations and their corresponding indices. Based on selected indices, Y combinations are chosen from $C_{k,v,f}$ for the next processing step of prediction.

Table 2 exhibits the best correlation-based combinations of IMFs of each weather condition and PV generation data. Algorithm 1 yields these combinations. The first column identifies the pairs of output combinations (OCs), where OC1, OC2, and OC3 represent IMF 1 and residual, IMF 2 and IMF 3, and IMF 4 and IMF 5, respectively. Similarly, input combinations (IC) are selected, where IC-IJ represents input IMF combination J^{th} combination number of weather condition I. The investigation considers three IMF combinations for 1^{st} and 2^{nd} weather variables and two IMF combinations for 3^{rd} weather variables. The table also mentions the correlation value R along with IMF combinations. These correlation values are between input combinations ICs and output combinations OCs. Figure 5 shows the IMF combinations of PV power generation (in red) and corresponding IMF combinations of selected atmospheric data like maximum temperature (in orange), minimum temperature (in blue), and pan evaporation (in green) data.

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. DATA ACQUISITION

To predict the solar PV power generation, a 100 KW grid-connected system is considered. The experiment evaluation utilizes 24-hour' data recorded at 0900 hours daily. The employed meteorological data is documented at the Climate, Energy and Water Research Institute (CEWRI) field station,

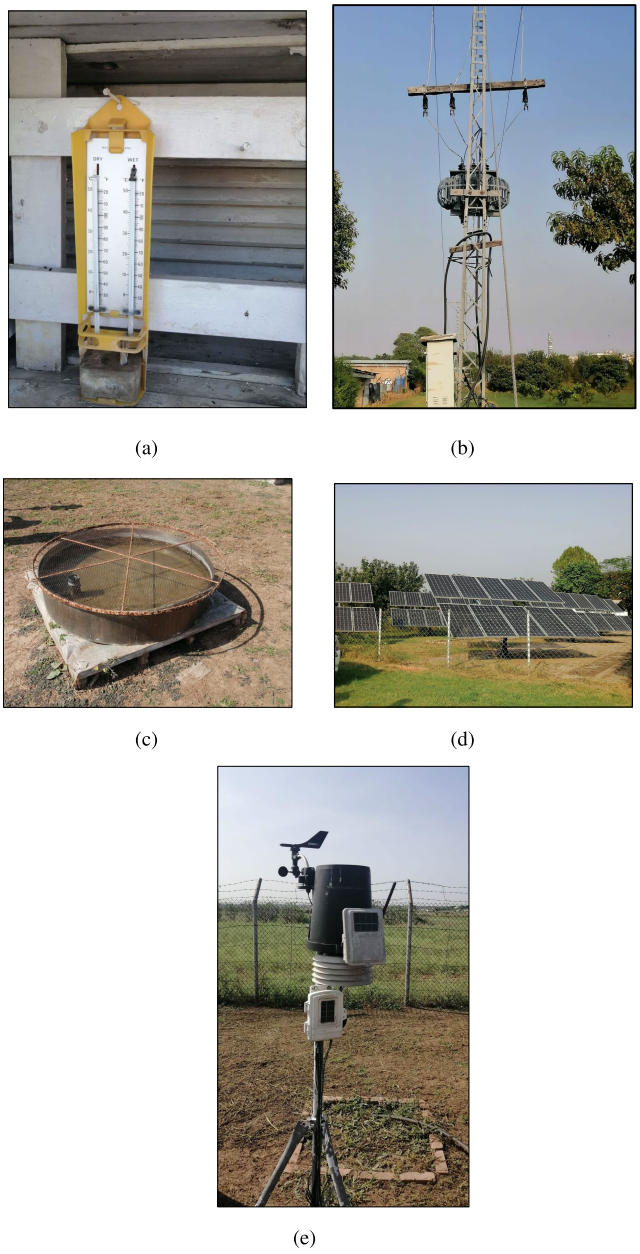


FIGURE 6. Atmospheric data sensors located at CEWRI field station, NARC, Islamabad, Pakistan (a) Temperature Sensor (b) Grid-connected Distribution Transformer (c) Pan Evaporation (d) Solar PV Panels (e) Wind Sensor.

National Agricultural Research Centre (NARC), Islamabad, Pakistan. The recording location has latitude and longitude coordinates of 33.4° north and 73.8° east with an altitude of 1632 feet. The location has multiple temperature, wind, pan evaporation, and solar sensors, as shown in Figure 6. Initially, the methodology includes meteorological data like the maximum temperature, minimum temperature, wind speed, pan evaporation, rainfall, and relative humidity at 0900 hours daily. Additionally, the experimental analysis utilized the relative and average humidity at 1400 hours. All the daily data, including PV power generation data, is considered for prediction from June to December 2020.

TABLE 2. Output IMF combinations (OCs) and its corresponding input combinations (ICs) by algorithm 1.

Output Combination		IC-11		IC-12		IC-13		IC-21		IC-22		IC-23		IC-31		IC-32	
		IMF	R	IMF	R	IMF	R	IMF	R	IMF	R	IMF	R	IMF	R	IMF	R
OC-1	1,6	2,6	0.88	1,4	0.11	3,5	0.08	4,6	0.82	3,5	-0.06	1,2	0.01	3,5	0.79	1,2	-0.14
OC-2	2,3	3,4	0.06	1,2	0.05	5,6	0.03	2,3	0.07	4,5	-0.03	1,6	0.02	2,3	-0.36	1,4	0.17
OC-3	4,5	2,4	-0.49	3,5	0.23	1,6	-0.11	3,4	-0.15	2,5	0.12	1,6	-0.10	1,4	0.20	2,3	-0.14

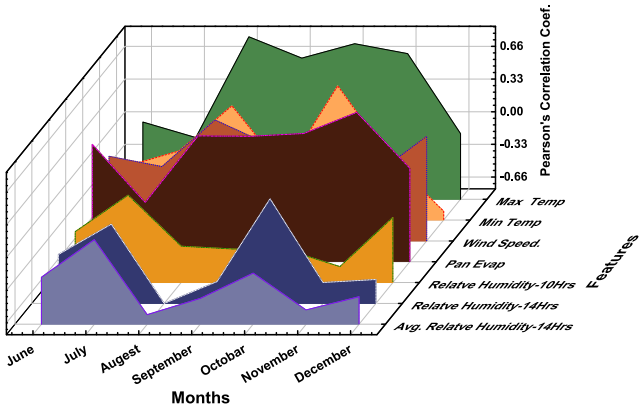


FIGURE 7. Pearson's correlation coefficient value of each atmospheric variable with photo-voltaic power generation data.

B. PIVOTAL ATMOSPHERIC VARIABLES TO PV POWER

PV Power generation can depend on various atmospheric data streams. Maximum or minimum temperature, pan evaporation, humidity, and relative humidity are considered PV power data prediction factors. However, their effect on prediction is not the same for all weather factors. Some factors have more influence over predicting PV power data than others. To determine the most influential data factors, Perason's Correlation Coefficient (PCC) is employed. The lowest values of PCC are utilized to discard the unrelated weather data variables for the input data stream. This data pre-processing not only simplifies the prediction but also mitigates the computational complexity of the prediction process.

$$r_{x,y} = \frac{\sum_i^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^N (x_i - \bar{x})^2 \sum_i^N (y_i - \bar{y})^2}} \tag{2}$$

In equation 2, $r_{x,y}$ denotes the Perason's Correlation Coefficient between y_i PV power generation data and a possible weather data series x_i. Meanwhile, N denotes the length of time series data. $\bar{x}$ and $\bar{y}$ are the averages of x and y, respectively.

Figure 7 displays the PCC values for all the atmospheric variables for 30 days. Among these weather variables, maximum and minimum temperatures have the lowest variations. Both of these data series, including pan evaporation, have the highest PCC value compared to the rest of the weather variables. Therefore, wind and humidity-related factors are removed from the input.

C. EMD

Empirical Mode Decomposition (EMD) typically includes the adaptive decomposition of any complex signal into multiple intrinsic mode functions (IMFs). These IMFs represent the energies related to different intrinsic time scales and separate any event in time and frequency. The decomposition of the time series consists of the following steps.

1) The EMD algorithm's basic step separates each Intrinsic Mode Function (IMF) from the signal. The approach involves finding the minima and maxima to create the upper $E_u(t)$ and lower $E_l(t)$ envelopes of the signal. Next, we calculate the local mean using these two envelopes.

$$\mu = \frac{E_u(t) + E_l(t)}{2}$$

2) The original signal x is subtracted from this local mean μ to produce an oscillating component, y. Assume y as new signal x. and repeat the process from step 1 to acquire new y.

$$y_i = x - \mu$$

3) The process repeats until the stoppage criterion is satisfied. The condition is fulfilled when y does not have more than two extrema of the same sign, ensuring that y is in single oscillatory mode. If the above condition is not satisfied, the process repeats itself.
4) After extraction of all IMFs, the residual signal is obtained after subtracting the sum of all IMFs from the original signal.

$$RES = x - \sum_i^N y_i \tag{3}$$

Each residual represents the overall tendency in the original signal.

D. PERFORMANCE MEASURES

Root Mean Square Error(RMSE) is the most commonly used performance measure to verify the prediction. It calculates the mean difference between the predicted values of an estimator and the original ground values. Mathematically, it represents the standard deviation of differences between data inputs and the regression line called residual. RMSE measures the tightness of these points around the regression line. The more

TABLE 3. Experimental results of the proposed and conventional technique and their corresponding units usage on LSTM and NARX.

	LSTM Network												NARX Network		
	SGDM				ADAM				RMSPROP						
	Three Network	Single Network	Direct Selected Input	Direct Full Input	Three Network	Single Network	Direct Selected Input	Direct Full Input	Three Network	Single Network	Direct Selected Input	Direct Full Input	Three Network	Single Network	Direct Selected Input
RMSE	14.44	16.03	17.50	21.82	**8.17**	8.40	27.11	20.20	8.19	8.74	22.95	20.74	15.45	8.24	25.43
MAE	10.00	8.39	11.01	8.50	3.30	**2.81**	15.61	10.33	4.18	4.84	13.21	11.84	11.43	2.63	16.55
MAPE	6.80	5.78	7.02	5.69	2.72	2.41	9.72	6.92	3.14	3.61	8.26	7.84	6.39	**1.46**	9.79
R-Seq	0.97	0.96	0.96	0.93	**0.99**	0.99	0.90	0.94	0.99	0.99	0.93	0.94	0.96	0.96	0.88
NN-Units	-	-	-		-	-	-		-	-	-		24,43,3	47	11
LSTM-Units	134,24,107	12	8	10	86,84,166	139	84	107	197, 178, 125	73	194	132	-	-	-
Time Cost (msec)	5.56	21.32	65.04	16.58	4.74	24.60	19.40	17.28	8.04	26.74	32.48	56.06	1.71	1.98	2.57

TABLE 4. Experimental results of the proposed and conventional technique and their corresponding units usage on GRU.

	GRU Network											
	SGDM				ADAM				RMSPROP			
	Three Network	Single Network	Direct Selected Input	Direct All Full Input	Three Network	Single Network	Direct Selected Input	Direct All Full Input	Three Network	Single Network	Direct Selected Input	Direct All Full Input
RMSE	33.14	49.34	39.82	33.85	**7.26**	8.19	35.78	18.22	7.54	7.70	20.70	17.06
MAE	24.98	30.72	25.89	23.08	3.18	**2.99**	15.02	10.21	4.05	4.26	11.88	9.25
MAPE	16.04	24.46	18.58	14.98	2.60	**2.13**	10.15	6.71	2.93	3.09	7.92	6.60
R-Seq	0.83	0.61	0.74	0.84	**0.99**	**0.99**	0.86	0.86	0.99	0.99	0.95	0.84
GRU-Units	2,3,3	1	2	3	168,112,144	178	70	51	80,176,149	171	200	172
Time Cost(msec)	6.708	53.932	19.734	19.601	7.347	42.909	19.802	25.174	19.787	57.798	17.873	24.058

the RMSE value near 1, the less the residual values. In other words, it sticks around the line more tightly.

$$RMSE = \sqrt{\frac{\sum_i^N (y_i - x_i)^2}{N}} \tag{4}$$

In equation 4, y_i denotes the actual values, x_i is the predicted value and N denotes the total length of y_i or x_i.

Mean absolute error (MAE) measures the absolute difference between the original value and the predicted value. Mathematically, it computes the non-negative error between the prediction and the original value. It provides the mathematical framework for averaging errors on all the data points.

$$MAE = \frac{\sum_i^N |y_i - x_i|}{N} \tag{5}$$

In equation 5, y_i denotes the actual values, x_i is the predicted value and N denotes the total length of y_i or x_i.

The mean absolute percentage error (MAPE) measures the accuracy of the prediction or forecasting method. It is commonly used as it provides the relative error. Even it is utilized as a loss function in regression training models. The mathematical framework of MAPE is defined as

$$MAPE = \frac{1}{N} \sum_i^N |\frac{y_i - x_i}{y_i}| \tag{6}$$

In equation 6, y_i denotes the original value, x_i is the predicted value and N denotes the total length of y_i or x_i.

Performance measure of R^2 determines the fitness of regression or prediction model data. It calculates the proportion of variation in the dependent variable that can be attributed to the independent variable. The value of R^2 ranges between 0 and 1. The value near zero means the data does not fit well to the regression model while the value near one implies good data fitness to the regression model. However, it does not give any clue about the accuracy of the regression model. Therefore, R^2 is typically utilized with other performance measures. The mathematical framework of the stated method is as follows.

$$R^2 = 1 - \frac{\sum_i^N (y_i - x_i)}{\sum_i^N (y_i - \hat{y})} \tag{7}$$

In equation 7, y_i denotes the original value, x_i is the predicted value and $\hat{y}$ denotes the average over y.

E. EXPERIMENTAL SETUP

The experimental evaluation involves deep learning approaches like LSTM and NARX networks. Three configurations are being considered for the LSTM network to acquire the complete analysis. First, for three IMF combinations of PV power generation, three individual LSTM networks are trained separately for a combination. In the second configuration, a single network is trained, having all three outputs, one for each combination. Lastly, a direct approach uses the maximum and minimum temperature and pan evaporation as inputs. In contrast, the target prediction is power generation PV time series data. Each network is trained using three different optimization functions.

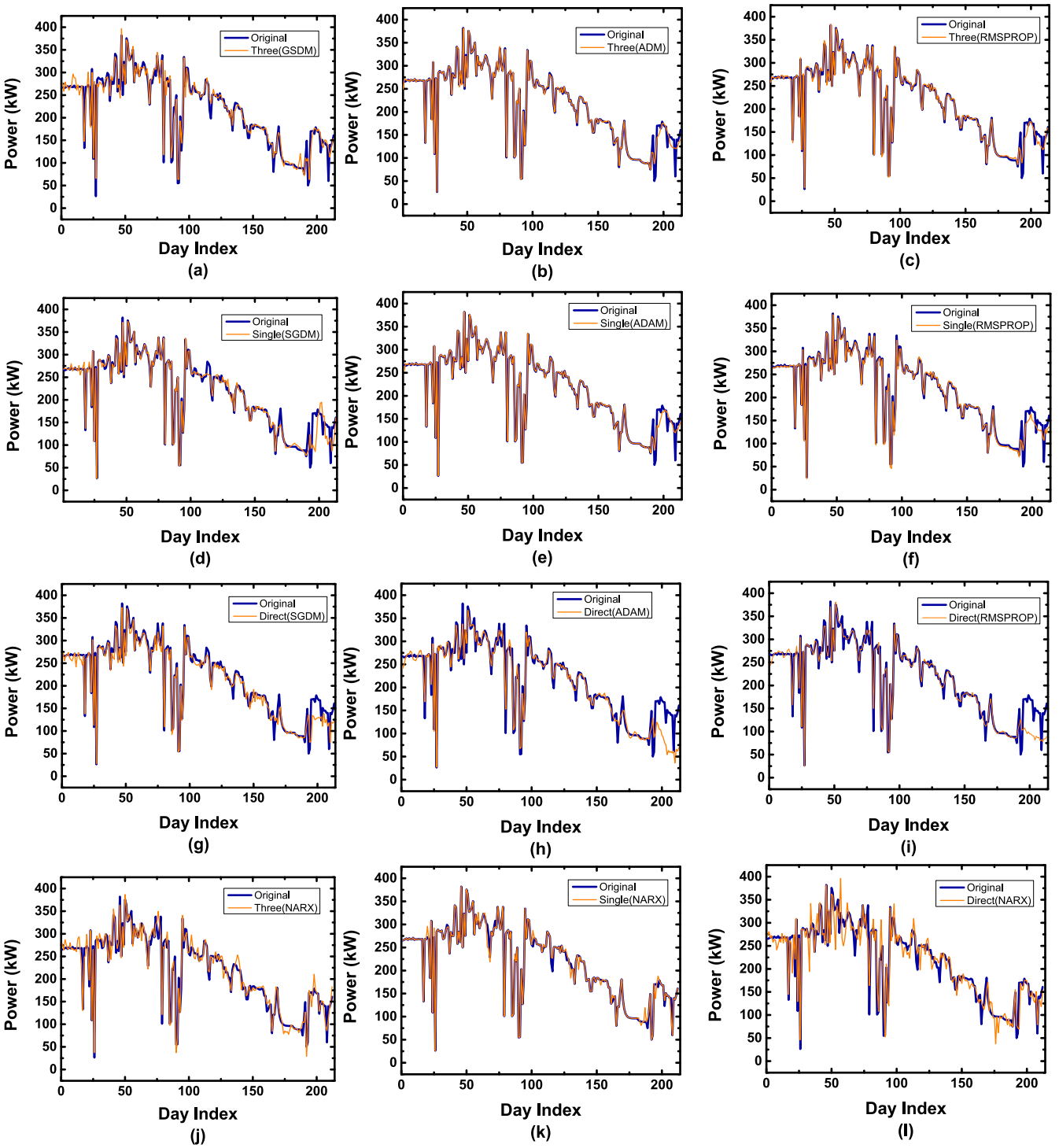
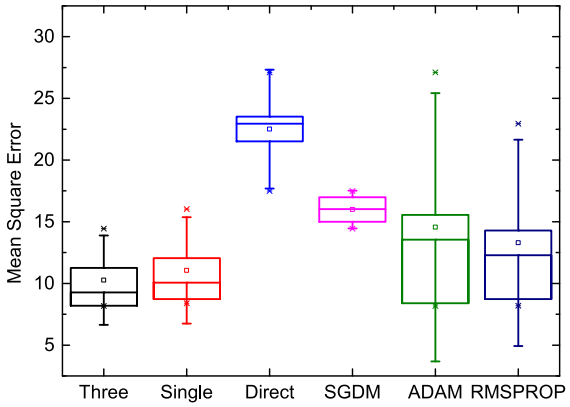
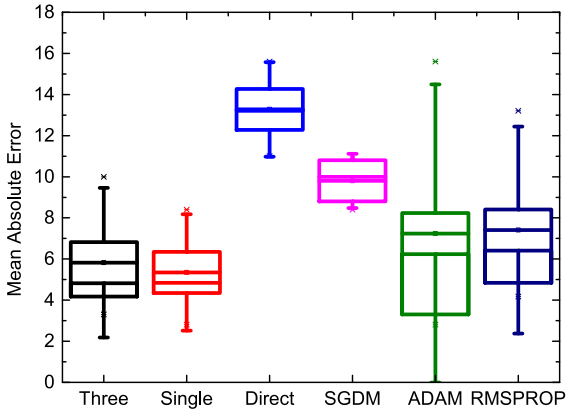


FIGURE 8. Graphs of prediction results of LSTM and NARX networks where 1^{st} row, graphs (a-c) represent three LSTM networks with output, while 2^{nd} row, graphs (d-f) are related to a single LSTM network an output, and 3^{rd} row graphs (g-i) are results of direct approach. The last row graphs (j-l) provides the results of the NARX network.

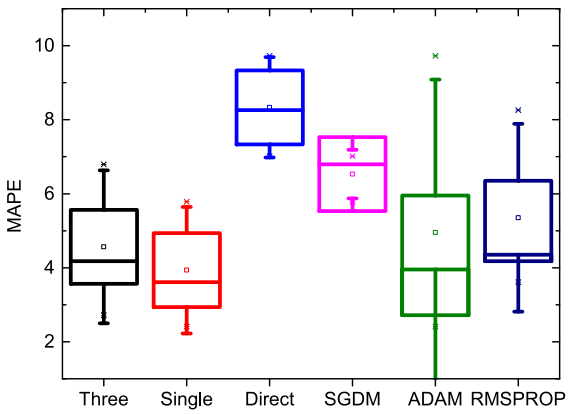
Stochastic Gradient Descent with Momentum (SGDM) [49] optimization is the ubiquitous scheme among these. To update parameters, SGD is an iterative process that involves the minimization of the loss function by taking small steps in the direction of the negative gradient of the loss function. However, simple SGD oscillates with the longest downward step towards achieving the minima of object function. Therefore, an additive term of momentum is



(a) Mean and standard deviation of error (MSE)



(b) Mean and standard deviation of error (MAE)



(c) Mean and standard deviation of error (MAPE)

FIGURE 9. Graphs of errors across various configurations.

included to mitigate the oscillations. The standard update of parameters is

$$\theta_{i+1} = \theta_i - l\nabla L(\theta_i) + m(\theta_i - \theta_{i-1}) \tag{8}$$

In equation 8, θ_i denotes training parameter i while learning rate l should be greater than zero. m denotes the momentum term. While $\nabla L(\theta_i)$ denotes the gradient of the loss function. SGDM employs a single learning rate for all parameter updates. On the contrary, Root Mean Square Propagation (RMSP) utilizes different learning rates for different parameter updates. It computes the moving average of the square of each element of the parameter gradient,

$$u_{i+1} = \alpha_2 u_i - (1 - \alpha_2)[\nabla L(\theta_i)]^2 \tag{9}$$

$$\theta_{i+1} = \theta_i - \frac{l\nabla L(\theta_i)}{\sqrt{u_i} + \varepsilon} \tag{10}$$

In equation 9, α_2 is the decay rate of the moving average. Whereas, in equation 10, ε is the small value for prevention in case of square root value of u_i is zero. Adaptive moment estimation (ADAM) [50] has a comparable updation mechanism to RMSP. Additionally, it includes the moment term as well. Parameter gradients and corresponding squared values are used to calculate the moving averages of each element.

$$m_{i+1} = \alpha_1 m_i + (1 - \alpha_1)\nabla L(\theta_i) \tag{11}$$

$$u_{i+1} = \alpha_2 u_i + (1 - \alpha_2)[\nabla L(\theta_i)]^2 \tag{12}$$

$$\theta_{i+1} = \theta_i - \frac{lm_l}{\sqrt{u_i} + \varepsilon} \tag{13}$$

In equations 11 and 12, α_1 and α_2 are the decay rates of gradients and their squares, respectively. Equation 13 shows the RMSP-like parameter updation procedure.

The experimental evaluation for all proposed methods and direct approaches shows the enhanced performance of overall LSTM variants over the NARX networks. Table 3 exhibits RMSE, MAE, MAPE, and R^2 performance criterion. Figures 8(a), (d), (f) show the SGDM optimization responses for three, single and direct LSTM networks. For three LSTM networks, 14.44, 10, 6.8, and 0.9 are the corresponding values of RMSE, MAE, MAPE, and R^2. Similarly, for a single LSTM network, 16.03, 8.39, 5.78, and 0.96 are values for the same performance measures. In contrast, the direct approach has values of 17.5, 11.01,7.02, and 0.96 for the same performance criteria. For SDGM optimization, three and single network has better overall performance responses as compared to the direct approach. ADAM optimization reports 8.17, 3.30, 2.72, and 0.99 values of RMSE, MAE, MAPE, and R^2, respectively. The single network has 8.40, 2.81, 2.41, and 0.99 values for criteria. In comparison, the direct approach holds the values of 27.11, 15.61, 9.72, and 0.90 for the above-mentioned performance values. The three networks demonstrate overall better performance than the rest of the networks. Comparing Figures 8(b), (e), (h) show the overall better prediction of three networks as compared to the direct and single networks. For RMSPROP optimization, the three networks demonstrate performance values of 8.19, 4.18, 9,72, and 0.99 respectively. In contrast, single network values are 8.74, 4.84, 3.62, and 0.93 respectively. The direct approach has 22.95, 13.21, 8.26, and 0.93 respectively. Figures 8(c), (f), (i) demonstrate the better overall prediction scores of the three networks as compared to the direct or single network scores. Figures 8(j), (k), (l) demonstrate the prediction curves for the three, single and direct NARX networks. The three output NARX network has performance values of 15.45,11.43,6.39, and 0.96 for RMSE, MAE, MAPE, and R^2 respectively.

For a single network, these values are 8.24, 2.63, 1.46, and 0.96 respectively. The direct approach has values of 25.43, 16.55, 9.79, and 0.88 respectively. Overall single NARX networks have better performance scores as compared to the three and direct NARX networks. The LSTM three network with ADAM optimization has the best RMSE value of 8.17, while the LSTM single network with ADAM optimization has the best MAE value of 2.81. Similarly, the single NARX network has the best value of 1.46 and the LSTM three network with ADAM optimization has the best R^2 value of 0.99.

Table 3 demonstrates the Neural Network units utilized for each configuration of the NARX network. It also exhibits the LSTM units for each optimization type and network configuration. Single-layer LSTM shows better overall performance over NARX. However, within the LSTM network, the three network approach performs better than other network configurations but at the cost of more LSTM units.

Experimentation is extended to Gated Recurrent Units (GRUs) implementation in Table 4. The lowest RMSE achieved is 7.26, which belongs to GRU three network implementation with the ADAM optimization. The same optimizer configuration with a single network results in the minimum MAE value of 2.99. While for the same network and optimization configuration, the MAPE and R^2 values are 2.13 and 0.99, respectively. On average, for all experimentations, the time cost of three network configurations is the least as compared to other conventional and proposed methods.

Figure 9 represents the error analysis of the proposed optimizations and configurations. The error and its standard deviation are considered for the three single network configurations across all optimization schemes. The ADAM, RMSPROPS, and SGDM optimization schemes represent the error averages across all configuration network configurations of the single, three networks, and direct approach. The direct approach shows the maximum error in all error measures. Figure 9.a shows the average and standard deviation of MSE. Figure 9.b represents the average and its standard deviation of MAE. And Figure 9.c demonstrates the same for MAPE.

IV. CONCLUSION

Deep learning approaches and a proposed pre-processing reconstruction algorithm, are proposed to improve the prediction of highly irregular, random, and uncertain PV power generation in this paper. The proposed approach employs PCC to select the environment variables based on similarity with PV power data. PCC removed unrelated atmospheric variables. EMD decomposes the time series data of selected atmospheric variables e.g., minimum, maximum temperature, pan evaporation, and PV power generation. Despite using PCC, these decomposed components have no relation to facilitate the prediction of PV power decomposed components. To improve the accuracy of the proposed model a correlation-based signal synthesis (CBSS) algorithm was proposed to establish the combination of these decomposed components related to combinations of PV power based on correlation. Further, LSTM and NARX deep learning approaches are utilized to find the prediction of PV power data for three different network configurations. These were using three different networks for each output, a single network for all output, and a direct approach without preprocessing. For the LSTM network, Adaptive moment estimation (ADAM), Stochastic Gradient Descent with Momentum (SGDM), and Root Mean Square Propagation (RMSP) optimization are utilized.

The results demonstrate under the LSTM approach with ADAM and RMSPROP optimization R^2 value is above 99%, which shows a good fit. Compared to other approaches, these two optimizations significantly differ between RMSE and MAE. These error measures have relatively low values for these model mixtures, which indicate superiority over other models. To meet the supply and demand of power, improve dispatching capacity, and gird planning of the power department, a PV power forecasting method is proposed, which is advantageous to the power department. In future work, further improving prediction accuracy is achievable using optimization for layers and the number of units in LSTM networks. The optimization is possible using different heuristic methods.

TABLE 5. Metrics used in algorithm 1.

Variables	Metrics	Comments
N	Matrix	Matrix containing EMD components of all features
N_f	Matrix	Matrix containing EMD components of feature f
$C_{k,f}$	Matrix	Sum of EMD components of feature f and k denotes the index
k	Index	Represents the index of all binary combinations of EMD components
N_y	Matrix	Matrix containing all possible combinations of EMD components of output feature y
N_c	Matrix	Matrix containing the all possible combinations of EMD components belong to features
$r_{y,k,f}$	Correlation matrix	Correlation coefficient between the additive combination of EMD components y of output feature f_y and k additive combination of EMD components of feature f
S_y	Vector	Containing the summation corresponding to each additive combination of output feature f_y of all selected correlation coefficients of each feature f
y'	Index	Selected indices of the additive combination of EMD components of output feature y
k'	Index	Selected indices of the additive combination of EMD components
$C'_{y',k',f}$	Matrix	Selected sum of EMD components of feature f and indices y' and k'

APPENDIX
CBSS ALGORITHM FOR PV FORECAST

Table 5 is representing all variables and metrics used in algorithm 1.

Algorithm 1 Correlation-Based Signal Synthesis (CBSS) Algorithm

Input: IMFs of weather variables and PV generation data
Output: *Selected output combinations* $C'_{y'_u,k'_v,f}$

$k \leftarrow 1$
for all f *in* N **do**
  for all i *in* N_f **do**
    for all $j \leftarrow i+1$ *in* N_f **do**
      $C_{k,f} \leftarrow$ *Add* c_i^f *and* c_j^f *EMD component of* f
      $k \leftarrow k+1$
    end for
  end for
end for
for all y *in* N_y **do**
  for all f *in* $N-1$ **do**
    for all k *in* N_c **do**
      $r_{y,k,f} \leftarrow$ *Compute correlation between* $C_{k,f}$ *and* C_{y,f_y}
    end for
    $r_{y,k'(1:3),f} \leftarrow \forall k$ *select top three values from* $r_{y,1:N_C,f}$ *such that their additive components ofcombination do not repeat*
  end for
end for
for all y *in* N_y **do**
  $S_y \leftarrow \sum_{f=1}^{N-1} \sum_{l=1}^{3} r_{y,l,f}$
end for
if N_y *is odd* **then**
  $C_{y'(1:quotient(N_y,2))} \leftarrow \forall y$ *select* $quotient(N_y, 2)$ *combinations and their indices with highest values of* S_y *such that their additive components ofcombination do not repeat*
  $Y \leftarrow [1 : quotient(N_y, 2)]$
else
  $C_{y'(1:\frac{N_y}{2})} \leftarrow \forall y$ *select* $\frac{N_y}{2}$ *combinations and their indices with highest values of* S_y *such that their additive components ofcombination do not repeat*
  $Y \leftarrow [1 : \frac{N_y}{2}]$
end if
for all u *in* Y **do**
  for all v *in* 3 **do**
    for all f *in* $N-1$ **do**
      $C'_{y'_u,k'_v,f} \leftarrow C_{k'_v,f}$
    end for
  end for
end for

ACKNOWLEDGMENT

The climate data is provided by the Climate, Energy, and Water Research Institute (CEWRI), NARC, Islamabad, Pakistan, in this work.

REFERENCES

[1] A. Qazi, F. Hussain, N. ABD. Rahim, G. Hardaker, D. Alghazzawi, K. Shaban, and K. Haruna, "Towards sustainable energy: A systematic review of renewable energy sources, technologies, and public opinions," *IEEE Access*, vol. 7, pp. 63837–63851, 2019.

[2] R. Asghar, M. H. Sulaiman, Z. Mustaffa, Z. Ali, and Z. Ullah, "Integration of electric vehicles in smart grids: A review of the advantages and challenges of vehicle-to-grid technology," in *Proc. Int. Conf. IT Ind. Technol. (ICIT)*, Oct. 2022, pp. 1–7.

[3] Y. Lu, Z. A. Khan, M. S. Alvarez-Alvarado, Y. Zhang, Z. Huang, and M. Imran, "A critical review of sustainable energy policies for the promotion of renewable energy sources," *Sustainability*, vol. 12, no. 12, p. 5078, 2020. [Online]. Available: https://www.mdpi.com/2071-1050/12/12/5078

[4] M. Amer and T. U. Daim, "Selection of renewable energy technologies for a developing county: A case of Pakistan," *Energy Sustain. Develop.*, vol. 15, no. 4, pp. 420–435, Dec. 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0973082611000767

[5] B. Shaker, K. Ullah, Z. Ullah, M. Ahsan, M. Ibrar, and M. A. Javed, "Enhancing grid resilience: Leveraging power from flexible load in modern power systems," in *Proc. 18th Int. Conf. Emerg. Technol. (ICET)*, Nov. 2023, pp. 246–251.

[6] K. Padmanathan, U. Govindarajan, V. K. Ramachandaramurthy, and B. Jeevarathinam, "Integrating solar photovoltaic energy conversion systems into industrial and commercial electrical energy utilization—A survey," *J. Ind. Inf. Integr.*, vol. 10, pp. 39–54, Jun. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2452414X17300742

[7] A. Dajuma, S. Yahaya, S. Touré, A. Diedhiou, R. Adamou, A. Konaré, M. Sido, and M. Golba, "Sensitivity of solar photovoltaic panel efficiency to weather and dust over West Africa: Comparative experimental study between Niamey (Niger) and Abidjan (Côte d'Ivoire)," *Comput. Water, Energy, Environ. Eng.*, vol. 5, no. 4, pp. 123–147, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:59360546

[8] W. Ahmed, B. Khan, Z. Ullah, F. Mehmood, S. M. Ali, E. E. Edifor, S. Siraj, and R. Nawaz, "Stochastic adaptive-service level agreement-based energy management model for smart grid and prosumers," *PLoS ONE*, vol. 17, no. 12, Dec. 2022, Art. no. e0278324.

[9] U. K. Das, K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. I. Idris, W. Van Deventer, B. Horan, and A. Stojcevski, "Forecasting of photovoltaic power generation and model optimization: A review," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 912–928, Jan. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032117311620

[10] J. Zhong, L. Liu, Q. Sun, and X. Wang, "Prediction of photovoltaic power generation based on general regression and back propagation neural network," *Energy Proc.*, vol. 152, pp. 1224–1229, Oct. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1876610218307173

[11] D. P. Simatupang and J. Choi, "PV source inverter with voltage compensation for weak grid based on UPQC configuration," in *Proc. IEEE 18th Int. Power Electron. Motion Control Conf. (PEMC)*, Aug. 2018, pp. 421–427.

[12] D. Simatupang and J. Choi, "Integrated photovoltaic inverters based on unified power quality conditioner with voltage compensation for submarine distribution system," *Energies*, vol. 11, no. 11, p. 2927, 2018. [Online]. Available: https://www.mdpi.com/1996-1073/11/11/2927

[13] A. Tuohy, J. Zack, S. E. Haupt, J. Sharp, M. Ahlstrom, S. Dise, E. Grimit, C. Mohrlen, M. Lange, M. G. Casado, J. Black, M. Marquis, and C. Collier, "Solar forecasting: Methods, challenges, and performance," *IEEE Power Energy Mag.*, vol. 13, no. 6, pp. 50–59, Nov. 2015.

[14] Y. Huang, J. Lu, C. Liu, X. Xu, W. Wang, and X. Zhou, "Comparative study of power forecasting methods for PV stations," in *Proc. Int. Conf. Power Syst. Technol.*, Oct. 2010, pp. 1–6.

[15] A. Bracale, P. Caramia, G. Carpinelli, A. R. Di Fazio, and G. Ferruzzi, "A Bayesian method for short-term probabilistic forecasting of photovoltaic generation in smart grid operation and control," *Energies*, vol. 6, no. 2, pp. 733–747, 2013. [Online]. Available: https://www.mdpi.com/1996-1073/6/2/733

[16] M. G. De Giorgi, P. M. Congedo, and M. Malvoni, "Photovoltaic power forecasting using statistical methods: Impact of weather data," *IET Sci., Meas. Technol.*, vol. 8, no. 3, pp. 90–97, 2014. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-smt.2013.0135

[17] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renew. Sustain. Energy Rev.*, vol. 27, pp. 65–76, Nov. 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032113004334

[18] D. Lauria, F. Mottola, and D. Proto, "Caputo derivative applied to very short time photovoltaic power forecasting," *Appl. Energy*, vol. 309, Mar. 2022, Art. no. 118452. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261921016779

[19] X. Luo and D. Zhang, "An adaptive deep learning framework for day-ahead forecasting of photovoltaic power generation," *Sustain. Energy Technol. Assessments*, vol. 52, Aug. 2022, Art. no. 102326. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213138822003782

[20] L. Wang, Y. Liu, T. Li, X. Xie, and C. Chang, "Short-term PV power prediction based on optimized VMD and LSTM," *IEEE Access*, vol. 8, pp. 165849–165862, 2020.

[21] S. Muzaffar and A. Afshari, "Short-term load forecasts using LSTM networks," *Energy Proc.*, vol. 158, pp. 2922–2927, Feb. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1876610219310008

[22] C. Chen, S. Duan, T. Cai, and B. Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," *Sol. Energy*, vol. 85, no. 11, pp. 2856–2870, Nov. 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038092X11003008

[23] X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM," *Energy*, vol. 148, pp. 461–468, Apr. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544218302056

[24] C. N. Obiora, A. Ali, and A. N. Hasan, "Forecasting hourly solar irradiance using long short-term memory (LSTM) network," in *Proc. 11th Int. Renew. Energy Congr. (IREC)*, Oct. 2020, pp. 1–6.

[25] D. Chandola, H. Gupta, V. A. Tikkiwal, and M. K. Bohra, "Multi-step ahead forecasting of global solar radiation for arid zones using deep learning," *Proc. Comput. Sci.*, vol. 167, pp. 626–635, Jan. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S187705092030795X

[26] M. Aslam, J.-M. Lee, H.-S. Kim, S.-J. Lee, and S. Hong, "Deep learning models for long-term solar radiation forecasting considering microgrid installation: A comparative study," *Energies*, vol. 13, no. 1, p. 147, 2020. [Online]. Available: https://www.mdpi.com/1996-1073/13/1/147

[27] B. P. Mukhoty, V. Maurya, and S. K. Shukla, "Sequence to sequence deep learning models for solar irradiation forecasting," in *Proc. IEEE Milan PowerTech*, Jun. 2019, pp. 1–6.

[28] S. Han, Y. Qiao, J. Yan, Y. Liu, L. Li, and Z. Wang, "Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network," *Appl. Energy*, vol. 239, pp. 181–191, Apr. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261919302065

[29] M. W. Ahmad, M. Mourshed, and Y. Rezgui, "Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression," *Energy*, vol. 164, pp. 465–474, Dec. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544218317432

[30] P. Kumari and D. Toshniwal, "Real-time estimation of COVID-19 cases using machine learning and mathematical models—The case of India," in *Proc. IEEE 15th Int. Conf. Ind. Inf. Syst. (ICIIS)*, Nov. 2020, pp. 369–374.

[31] O. Laib, M. T. Khadir, and L. Mihaylova, "Toward efficient energy systems based on natural gas consumption prediction with LSTM recurrent neural networks," *Energy*, vol. 177, pp. 530–542, Jun. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544219307054

[32] H. Zang, L. Liu, L. Sun, L. Cheng, Z. Wei, and G. Sun, "Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations," *Renew. Energy*, vol. 160, pp. 26–41, Nov. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960148120308557

[33] M. Bouzerdoum, A. Mellit, and A. Massi Pavan, "A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant," *Sol. Energy*, vol. 98, pp. 226–235, Dec. 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038092X13004039

[34] M. Moreira, P. Balestrassi, A. Paiva, P. Ribeiro, and B. Bonatto, "Design of experiments using artificial neural network ensemble for photovoltaic generation forecasting," *Renew. Sustain. Energy Rev.*, vol. 135, Jan. 2021, Art. no. 110450. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032120307371

[35] P. Li, K. Zhou, X. Lu, and S. Yang, "A hybrid deep learning model for short-term PV power forecasting," *Appl. Energy*, vol. 259, Feb. 2020, Art. no. 114216. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261919319038

[36] P. Kumari and D. Toshniwal, "Long short term memory–convolutional neural network based deep hybrid approach for solar irradiance forecasting," *Appl. Energy*, vol. 295, Aug. 2021, Art. no. 117061. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261921005158

[37] M. Gao, J. Li, F. Hong, and D. Long, "Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using LSTM," *Energy*, vol. 187, Nov. 2019, Art. no. 115838. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544219315105

[38] D. Lee and K. Kim, "Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information," *Energies*, vol. 12, no. 2, p. 215, Jan. 2019. [Online]. Available: https://www.mdpi.com/1996-1073/12/2/215

[39] Z. Pang, F. Niu, and Z. O'Neill, "Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons," *Renew. Energy*, vol. 156, pp. 279–289, Aug. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960148120305747

[40] S. Ghimire, R. C. Deo, N. Raj, and J. Mi, "Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms," *Appl. Energy*, vol. 253, Nov. 2019, Art. no. 113541. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261919312152

[41] J. Qu, Z. Qian, and Y. Pei, "Day-ahead hourly photovoltaic power forecasting using attention-based CNN-LSTM neural network embedded with multiple relevant and target variables prediction pattern," *Energy*, vol. 232, Oct. 2021, Art. no. 120996. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544221012445

[42] F. Rodríguez, F. Martín, L. Fontán, and A. Galarza, "Ensemble of machine learning and spatiotemporal parameters to forecast very short-term solar irradiation to compute photovoltaic generators' output power," *Energy*, vol. 229, Aug. 2021, Art. no. 120647. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544221008963

[43] A. Moreno-Munoz, J. J. G. de la Rosa, R. Posadillo, and F. Bellido, "Very short term forecasting of solar radiation," in *Proc. 33rd IEEE Photovoltaic Spec. Conf.*, May 2008, pp. 1–5.

[44] P. Aggarwal, "An exploration of the potential of day-ahead photovoltaic power forecasting models using deep learning neural networks," in *Proc. Int. Conf. Data Sci. Netw. Secur. (ICDSNS)*, Jul. 2023, pp. 1–8.

[45] V. H. Wentz, J. N. Maciel, J. J. Gimenez Ledesma, and O. H. Ando Junior, "Solar irradiance forecasting to short-term PV power: Accuracy comparison of ann and LSTM models," *Energies*, vol. 15, no. 7, p. 2457, 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/7/2457

[46] D. Cannizzaro, A. Aliberti, L. Bottaccioli, E. Macii, A. Acquaviva, and E. Patti, "Solar radiation forecasting based on convolutional neural network and ensemble learning," *Expert Syst. Appl.*, vol. 181, Nov. 2021, Art. no. 115167. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421006060

[47] H. K. Ahn and N. Park, "Deep RNN-based photovoltaic power short-term forecast using power IoT sensors," *Energies*, vol. 14, no. 2, p. 436, Jan. 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/2/436

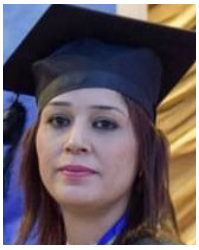
[48] S. Kumar and I. Chong, "Correlation analysis to identify the effective data in machine learning: Prediction of depressive disorder and emotion states," *Int. J. Environ. Res. Public Health*, vol. 15, no. 12, p. 2907, 2018. [Online]. Available: https://www.mdpi.com/1660-4601/15/12/2907

[49] K. P. Murphy, *Machine Learning : A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2013. [Online]. Available: https://www.amazon.com/Machine-Learning-Probabilistic-Perspective-Computation/dp/0262018020/ref=sr_1_2?ie=UTF8&qid=1336857747&sr=8-2

[50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, Y. Bengio and Y. LeCun, Eds. San Diego, CA, USA, May 2015. [Online]. Available: http://arxiv.org/abs/1412.6980

M. DILSHAD SABIR received the Graduate degree in computer engineering from COMSATS University Islamabad and the M.S. and Ph.D. degrees in CE from the National University of Sciences and Technology, Islamabad, in 2012 and 2022, respectively. He is currently a Faculty Member of COMSATS University Islamabad. His research interests include the Internet of Things (IoT), deep learning, correlation pattern recognition, convolution neural networks, and network security.

KAMRAN HAFEEZ received the M.S. degree in electrical power engineering from UET Peshawar, in 2008, and the Ph.D. degree in electrical engineering from COMSATS University Islamabad, in 2020. He is currently an Assistant Professor with the ECE Department, COMSATS University Islamabad. His research interests include power converters, HVDC systems, and electric power systems.

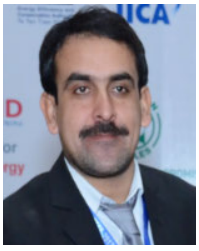


SAMERA BATOOL received the M.S. degree in computer science from the Arid Institute of Information Technology, the M.C.S. degree from NUML University, Islamabad, and the Ph.D. degree from NUST College of Electrical and Mechanical Engineering. Prior to her Ph.D. studies. She is currently serves as an Assistant Professor of Computer Science with the Department of Computer Science, COMSATS University Islamabad. She has published in several reputable journals with high impact factors. Her research interests include encompass healthcare data analytics, machine learning and deep learning-based human action recognition, and sensor data analytics for healthcare applications.

GHANI AKBAR received the Ph.D. degree from the Integrated Water Resource Management—Agricultural Engineering Discipline, Australia. He is currently a Principal Scientific Officer (PSO) with the Climate, Energy and Water Research Institute (CEWRI), National Agricultural Research Centre (NARC), under the Pakistan Agricultural Research Council (PARC), an apex research organization in Islamabad, Pakistan. He has more than 23 years of research experience in project planning, development, implementation related to conservation/regenerative agriculture, soil, water conservation, high-efficiency irrigation systems, solar water pumping, irrigation management in cereal crops, watershed management, soil salinity management, simulation modeling, capacity building, field demonstration, and dissemination of research results. He has published more than 65 research articles in peer-reviewed national and international/local journals. He was among the pioneer team members, who played a key role in introducing the Permanent Raised Bed Farming System/Conservation Agriculture in Pakistan. He is a member of the Research Evaluation Committee of several Ph.D. and M.Sc. students in Pakistan. He has local and international research collaborations. He has been placed in the Directory of Productive Scientists of Pakistan with the Engineering Discipline as per the Pakistan Science Foundation Ranking, since 2015. He was awarded the PARSA Best Scientist Award, in 2021, of PARC in the field of Natural Resource (Water) Management Based on scientific achievements, in 2015 and 2020. He is also a Registered Member of the Pakistan Engineering Council (PEC), linkages with the World Congress on Conservation Agriculture (WCCS), World Congress of Soil Science (WCSS), Irrigation Australia, and International Commission on Irrigation and Drainage (ICID). He is a reviewer of several national/international journals.

LAIQ KHAN received the B.Sc. degree (Hons.) in electrical engineering from Khyber Pakhtunkhwa University of Engineering and Technology, Peshawar, Pakistan, in 1996, and the integrated M.S. and Ph.D. degrees in power system dynamics and control from the University of Strathclyde, Glasgow, U.K., in 2003. Before the Ph.D. degree, he was with Siemens Pakistan as a Field Engineer for two years. He was an Assistant Professor with the Faculty of Electronic Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Swabi, Pakistan, until 2008. Then, he joined the Faculty of Electrical Engineering, COMSATS University Islamabad, Abbottabad Campus, Pakistan, and an Associate Professor. He was also a Professor with Islamic University Madinah, Saudi Arabia, for three years. Currently, he is a Professor in power system dynamics and control with COMSATS University Islamabad, Islamabad Campus, Pakistan. He has published more than a 100 publications in highly reputable international conferences and peer-reviewed impact factor journals. His research interests include power system stability and control using PSSs, FACTS controllers, HVDC, robust control theory, intelligent control systems, nonlinear adaptive intelligent control, adaptive predictive intelligent control fault tolerant control power system planning, advanced optimization techniques nonlinear control of renewable WECS and photovoltaic systems, microgrid and smartgrid, and hybrid electric vehicles.

GHULAM HAFEEZ received the B.Sc. degree in electrical engineering from the University of Engineering and Technology Peshawar, Pakistan, and the M.S. and Ph.D. degrees in electrical engineering from COMSATS University Islamabad, Islamabad, Pakistan. He was a Lecturer with the Department of Electrical Engineering, University of Engineering & Technology Mardan. He was a Lecturer with the University of Wah, Wah Cantt, Pakistan. He was also a Research Associate with COMSATS University Islamabad, where his research focus was computational intelligence, forecast process, energy management, operation of electricity market, and electric vehicles in smart power grids. He is currently a Lifetime Charted Engineer from Pakistan Engineering Council. He is also the Manager of University-Industry Linkages/Research Operations and Development in the Directorate of ORIC, University of Engineering & Technology Mardan. He has authored or coauthored peer-reviewed research papers in reputed international journals and conferences. His research interests include sustainable and smart energy, cities and societies, smart grids, applications of deep learning and blockchain in smart power grids, and stochastic techniques for power usage optimization in smart power grids.



ZAHID ULLAH (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering from UET Peshawar, in 2014, and the M.S. degree in electrical engineering from COMSATS University Islamabad, Abbottabad Campus, Abbottabad, Pakistan, in 2017. He is currently pursuing the Ph.D. degree in electrical engineering with Politecnico di Milano, Italy. He was a Lecturer with UMT Lahore, Pakistan, from 2017 to 2020. He has published various articles in reputed journals and IEEE conference proceedings. His research interests include smart grids, energy management, renewable energy systems, ICTs for power systems, V2G, and machine and deep learning.

Open Access funding provided by 'Politecnico di Milano' within the CRUI CARE Agreement