IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 17, 2021, accepted February 24, 2021, date of publication February 26, 2021, date of current version March 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3062776

An Effective Hybrid NARX-LSTM Model for Point and Interval PV Power Forecasting

MOHAMED MASSAOUDI[1,2], (Student Member, IEEE), INES CHIHI[3], LILIA SIDHOM[3], MOHAMED TRABELSI[4], (Senior Member, IEEE), SHADY S. REFAAT[1], (Senior Member, IEEE), HAITHAM ABU-RUB[1], (Fellow, IEEE), AND FAKHREDDINE S. OUESLATI[2]

[1]Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Doha, Qatar
[2]Laboratoire matériaux molécules et applications (LMMA) à l'IPEST, Carthage University, Tunis 2036, Tunisia
[3]Laboratory of Energy Applications and Renewable Energy Efficiency (LAPER), El Manar University, Tunis 1068, Tunisia
[4]Department of Electronic and Communications Engineering, Kuwait College of Science and Technology, Safat 13002, Kuwait

Corresponding author: Mohamed Massaoudi (mohamed.massaoudi@qatar.tamu.edu)

This work was supported in part by the National Priorities Research Program (NPRP) under Grant NPRP10-0101-170082 through the Qatar National Research Fund (a member of Qatar Foundation), and in part by the IBERDROLA QSTP LLC. The Open Access Funding was provided by the Qatar National Library.

ABSTRACT This paper proposes an effective Photovoltaic (PV) Power Forecasting (PVPF) technique based on hierarchical learning combining Nonlinear Auto-Regressive Neural Networks with exogenous input (NARXNN) with Long Short-Term Memory (LSTM) model. First, the NARXNN model acquires the data to generate a residual error vector. Then, the stacked LSTM model, optimized by Tabu search algorithm, uses the residual error correction associated with the original data to produce a point and interval PVPF. The performance of the proposed PVPF technique was investigated using two real datasets with different scales and locations. The comparative analysis of the NARX-LSTM with twelve existing benchmarks confirms its superiority in terms of accuracy measures. In summary, the proposed NARX-LSTM technique has the following major achievements: 1) Improves the prediction performance of the original LSTM and NARXNN models; 2) Evaluates the uncertainties associated with point forecasts with high accuracy; 3) Provides a high generalization capability for PV systems with different scales. Numerical results of the comparison of the proposed NARX-LSTM method with two real-world PV systems in Australia and USA demonstrate its improved prediction accuracy, outperforming the benchmark approaches with an overall normalized Rooted Mean Squared Error (nRMSE) of 1.98% and 1.33% respectively.

INDEX TERMS Long short-term memory (LSTM), photovoltaic power forecasting, nonlinear auto-regressive neural networks with exogenous input (NARXNN), Tabu Search Algorithm (TSA).

$\mathfrak{D}_1$	Data set for the 1st case study
$\mathfrak{D}_2$	Data set for the 2nd case study
P	Station pressure (bar)
A	Altimeter indication
ACE	Average Coverage Error
Adam	Adaptive Moment Estimation optimizer
AE	Autoencoder
Cc	Cloud coverage
CNN	Convolutional Neural Network
DIr	Diffuse horizontal radiation(W/m^2)
Dp	Dew point
EMD	Empirical Mode Decomposition
HO	Hyperparameter Optimization
HT	Hilbert Transformation
Ir	Horizontal radiation (W/m^2)
LSH	Local Sensitive Hashing
LSTM	Long-Short Term Memory
MAPE	Mean Absolute Percentage Error
MARS	Multivariate Adaptive Regression Spline
MSE	Mean Square Error
NARXNN	Non-linear Auto-regressive exogenous Network
nRMSE	normalized Rooted Mean Squared Error
PICP	Prediction Interval Coverage Probability
PINAW	Prediction Interval Normalized Average Width
PINC	Prediction Interval Nominal Confidence

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

PM	Persistence Model
PVPF	Photovoltaic power forecasting
Rh	Relative humidity (%)
RNN	Recurrent Neural Network
SCA	Sine Cosine Algorithm
SD	Standard derivation
SVM	Support Vector Machine
T	Ambient temperature (°C)
TCM	Time Correlation Modification
TSA	Tabu Search Algorithm
V	Visibility
Wd	Wind direction (Â°)
WGPR	Weighted Gaussian Process Regression
Ws	Wind speed (m/s)

I. INTRODUCTION

Energy transition towards renewables is a global trend in the twenty-first century. Solar energy is leading this transition due to its massive resources' potential [1]. Photovoltaic (PV) energy provides high availability and long durability to stakeholders. However, PV energy has several limitations in terms of low power stability and poor power quality. PV plants are continuously disturbed by weather conditions such as cloud cover, wind speed, and temperature variation [2]. Furthermore, the key stimulus for energy generation, namely, the irradiation, is only available in the daytime. Therefore, forecasting models have been widely employed for PV systems to estimate the generated PV power from one side and load demand from the other side to ensure a smart demand response and effective energy management [3], [4]. During the last few years, Time Series Forecasting (TSF) becomes a dynamic research area supported by the exponential use of Big Data in all the research fields due to the explosive development of information and communication technologies and significant hardware improvement [5]. Therefore, the research community has been focusing on the development of effective PV Power Forecasting (PVPF) techniques to stabilize and secure the grid operation [6].

So far, Machine Learning (ML) techniques have been proposed to interpret the behavior of the selected feature patterns that continuously change over time to reconstitute a clear vision about future values [7]. In simple words, these techniques analyze the past (input data) to estimate the future (system behavior). PVPF consists of direct and indirect PVPF [8]. On one side, direct PVPF analyses the historical weather data to predict PV power production [2], [8]. On the other side, indirect PVPF uses a dual-stage methodology. In the first stage, the solar irradiance is predicted as the most dependent PV power feature. The second stage implies the determination of the PV power based on mathematical correlations [2]. According to the authors' work in [9], it has been proved that the second strategy is more accurate, versatile, cheaper, and less computationally demanding. In [10], the authors reported that hybrid models can achieve competitive results compared to the-state-of-the-art techniques by combining two or more single ML models.

In view of the latest progress on time series forecasting, the time horizon is a fundamental taxonomy for PVPF applications. Typically, there are three major types for forecasting horizons, namely, short, medium, and long-term forecasting [5], [11]. Short-term forecasting includes an hourly to weekly estimation to effectively manage sudden dispatching and cope with operational outliers. Medium-term PVPF is valid from a week up to months ahead which is mainly used for preventive maintenance, planning, and asset utilization management purposes. The last type is the long-term forecasting for a year up to several years ahead [12]. Long-term PVPF is used for budget planning and feasibility studies of large-scale PV projects. It has been reported from the literature that the most popular and useful form of PVPF is the short-term PVPF [12].

Moreover, with the increasing penetration of PV power, two types of PVPF are introduced: point and interval forecasting. Point forecasting, named also single-valued forecasting, predicts single values of PV power, while interval forecasting quantifies the uncertainties associated with PVPF. Despite the huge research work dedicated to enhance the accuracy of point PVPF techniques, the increased unpredictability of weather conditions represents a significant limitation for these techniques [13]. A detailed comparison of the recent related works in PVPF is reported in Table 1.

For point forecasting, the authors in [27] succeeded to predict the PV power with high accuracy using a Recurrent Neural Network (RNN) technique where a historical database is used in the training process. However, it has been concluded that the vanishing gradient limits the forecasting time horizon of the RNN due to the gradient explosion and disappearance [5]. It is worth mentioning that RNN suffers from a high computational burden in the training phase. Moreover, these methods require a complete database for the training where a poor-quality database has a major impact on the output accuracy. Long-Short Term Memory (LSTM) algorithm is commonly employed as an elegant variation of the RNN model to overcome the information loss in depth. Deep Learning (DL) models were proposed to extract significant information from data representations to effectively track the feature patterns and improve the forecasting system capabilities. In [28], the authors proved that LSTM outperforms the existing benchmarks in terms of lower Rooted Mean Square Error (RMSE) and Mean Absolute Error (MAE). Unlike the poor effectiveness of conventional statistical models in tracking nonlinear variations, Nonlinear Auto-Regressive with Exogenous inputs Neural Network (NARXNN) is an improved version of RNN to manage the stochastic weather parameters [29]. In [30], a Dual-stage Attention mechanism-based RNN (DA-RNN) technique was investigated. The DA-RNN mechanism concept pays attention to the input series using NARXNN and then LSTM investigates time instances. The DA-RNN model achieves a MAE $= 7.14\ 10^{-2}\%$ and an RMSE $= 1.97\ 10^{-2}\%$. However, the NARXNN could miss the interpretation for local spatial attention in the first-level attention with the high variation of weather conditions

TABLE 1. A detailed comparison of PVPF methods with previous research work.

Method	Ref.	Year	Country	Lowest error	Advantages	Limitations
LSH	[14]	2018	USA, China	nRMSE=4.23%	Simple and Accurate	High sensibility to weather variations
WGPR	[15]	2018	Singapore	nRMSE=8.34%	Efficient for high-dimensional data	Require large datasets
Ensemble method	[16]	2018	New England	nRMSE=7.356%	Simple with less parameter requirements	Require a high data quality
CNN	[17]	2019	Taiwan	RMSE=140.90W	High generalization ability and robustness	Heavy computational effort
MARS	[18]	2019	China	RMSE=0.30 kWp	Cope with complex weather patterns	Difficult to implement
Copula function-LSTM	[19]	2019	USA	MAPE=6.65%	Alleviate the vanishing gradient problems	Difficult to implement
EMD-SCA-ELM	[20]	2020	India	MAPE=1.88 %	High accuracy and end-to-end design	Requires heavy tuning
Self-attention	[21]	2020	USA	RMSE= 0.071kW	High effectiveness and accuracy	Difficult parameter optimization
Deep ConvNets	[22]	2020	USA	nRMSE=21.2%	Do not require TSF models and weather data	Complex nonlinear modeling
AE-LSTM-PM	[23]	2020	USA	nRMSE=8.39%	High effectiveness with TSF	High computational burden
SVM	[24]	2020	Australia	RMSE=0.186 kW	High universality potential and automatic tuning	High Complexity
LSTM-TCM	[25]	2020	Australia	nRMSE=6.29%	High generalization ability	Heavy training and high complexity
Residual network	[26]	2020	Australia	MSE=0.027 kW	High accuracy	Prone to overfitting

in sequential processing. Furthermore, the encoder-decoder architecture is computationally extensive due to the continuous calculation of the pre-weighted input obtained from the encoder or an additional RNN in every iteration of the sequential architecture.

For interval forecasting, quantile regression and bootstrap method were proposed for PV power point uncertainty quantification [31]. In that work, the authors examined the usefulness of wavelet transform and radial basis function neural network to generate uncertainties with PV forecasts. From the results reported in [31], it has been noticed that multi-step multivariate PVPF was not reported. In [32], the authors proposed an improved interval PVPF using the Bayesian theory. The proposed model in [32] employs correlated weather scenario generation and ensemble models. The work presented in [33] introduced a Convolutional Neural Network (CNN)-based wavelet transform for point and interval forecasting. The proposed model can generate interval and point PVPF results for different experimental conditions. Despite the inherent potential of LSTM model for time series point forecasting, LSTM is usually losing sight of significance in interval forecasting. In order to tackle the above problems, this paper aims to further supplement the existing studies by proposing a highly accurate model for point and interval forecasting using a virtue of prepossessing blocks. The performance of the proposed model is demonstrated at different seasons of the year, different PV power locations, and several forecasting horizons. The main contributions of this paper are summarized as follows:

1) A hybrid method for nonlinear time series forecasting composed of LSTM and NARXNN models is comprehensively explained. Despite the large popularity of LSTM model, the high variability of environmental conditions restricts the LSTM model from achieving accurate forecasts. The proposed technique not only enhances the LSTM performance but also supplements the proposed integrated point forecasts by the uncertainties qualification to reduce the operational risks in PV systems.
2) A novel PVPF paradigm is deeply analyzed. The evaluation of the new approach is performed using two real datasets. Small-scale and large-scale PV plants are employed to demonstrate the model performance generalization for point and interval forecasting with single and multi-step prediction.
3) A comparative study with twelve benchmarks has been conducted to reveal the high reliability of the proposed NARX-LSTM model as a competitive model in capturing time dependencies with a high exactitude. Additionally, the proposed approach improves the prediction performance of the original LSTM and NARXNN models.

The remainder of the paper is organized as follows: Section II proposed a comprehensive presentation of the components

employed in the simulation process with a mathematical and architectural overview. In Section III, a case study is provided and the simulation results are presented and interpreted to validate the proposed model. Finally, Section IV concludes the paper.

II. FORECASTING MODELS AND PROPOSED ARCHITECTURE

In this section, the NARXNN Network and LSTM neural network model and Tabu Search algorithm are first presented. Then, the proposed NARX-LSTM model is comprehensively explained.

A. NARX NEURAL NETWORK

NARXNN is introduced as an improved form of Nonlinear Auto-Regressive Neural Network (NARNN). NARNN uses a single delayed feedback loop of the output regressor. On the other side, NARXNN involves two tapped-delay lines from the input-output signals [34]. The exogenous input values are integrated into the parametric equation of NARXNN as [35]:

$$y(n) = f[x(n); \bar{u}(n); \bar{y}(n-1)] \tag{1}$$

$$y(n) = f[x(n), \ldots, x(n-d_E+1); \hat{x}(n), \ldots, \hat{x}(n-d_y+1)] \tag{2}$$

where $\bar{u}(n) \in \mathbb{R}$ and $\bar{y}(n) \in \mathbb{R}$ denote the model input and output at the discrete timestep n, $d_E \geq 1$ and $d_y \geq 1$ represent the input and output memory orders respectively with $\{d_E, d_y\} \in \mathbb{N}^*$. The feedback loop improves the sensitivity of the NARXNN predictor to the historical data. Fig. 1 illustrates the underlining working mechanism of the NARXNN algorithm.

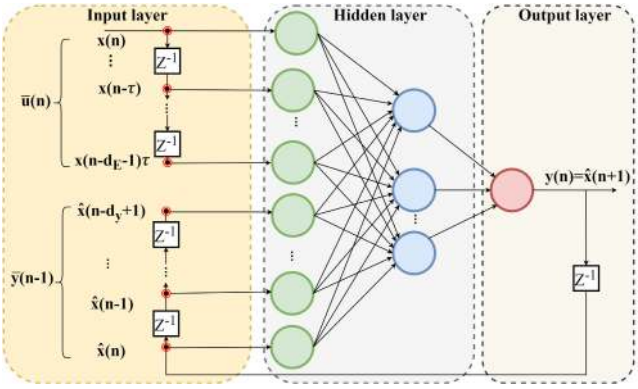


FIGURE 1. Three-layered NARXNN with $d_E \geq 1$ delayed inputs and $d_y \geq 1$ delayed outputs with z^{-1} presenting the unit time delay.

As shown in Fig.1, the NARXNN algorithm consists of a two-layer feed-forward network, with a linear transfer function in the output layer and a sigmoid function (σ) in the hidden layer calculated as [36]:

$$\sigma(x) = \frac{1}{1 + exp(-\sigma)} \tag{3}$$

This network has the specific feature of involving tapped delay lines to store previous values of $u(n)$ and $y(n)$ sequences. The output of the NARXNN, $y(n)$, is fed back to the input of the network (with a delay). Two different modes for the training of NARXNN model are depicted [35]:

- Series-Parallel (SP) Mode: This method takes the feedback delayed information from the real values.
- Parallel (P) mode: where the estimated outputs are set for the output's regressor.

However, for better accuracy and effective training, the NARX-SP feedback makes use of an on open-loop then the parallel feedback is switched in the evaluation part to a closed-loop. In [36], the simulations proved that the NARXNN provides better accuracy in discovering the behavior of the sequential output compared to conventional methods such as Feed-forward and Elman Networks [37]. The high performance is since the input vectors are inserted through two tapped-delay lines from the input-output signals. These delays present a jump-ahead connection in the time-unfolded network to provide the ability for the gradient descent to propagates back in a shorter path.

Nevertheless, the elimination of output memory for NARXNN significantly reduces its computational resources [35]. Moreover, NARXNN as part of the RNN faces, to a certain degree, the vanishing gradient problem. RNN after a specific input number stops learning and negatively affects the prediction accuracy. This challenge appears when the gradient descent shrinks in long-range dependencies which causes a fading memory issue [35].

B. LONG SHORT-TERM MEMORY

The LSTM network has been introduced to tackle the vanishing gradient problem in RNN architecture causing training failures [5]. The LSTM model fully exploits the long dependencies through a distinguished gate control and specialized memory mechanism. The whole mechanism is conducted by a series of LSTM cells presented in Fig. 2.

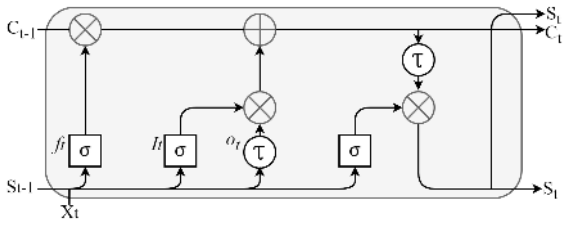


FIGURE 2. The internal chain structure of an LSTM unit: ⊗ and ⊕ present the pointwise scalar multiplication and the sum function respectively.

According to Fig. 2, the exploding and the vanishing gradient effects are mitigated by employing four gates in the LSTM architecture, specifically the input gate I_t, the forget gate f_t, the output gate o_t, and the self-recurrent unit. These gates are responsible for managing the interactions among the memory units where the gate selection is achieved according to hyperbolic tangent ($tanh$) function, Sigmoid function, or matrix multiplication. The input gate verifies the ability of the input signal in modifying the state of the

memory cell. Additionally, the f_t gate discards the irrelevant status that misleads the forecasting process and keeps only the important information to be forwarded to the hidden layers. The f_t values are between 0 and 1 where a higher value means that the information is of utmost importance and a result of 0 leads to completely discarding the information. Here, contrary to the input gate, the output gate checks its effect on the state of other memory cells. The LSTM gates, hidden outputs, and cell states are given as follows [5]:

$$I_t = \sigma(W_I x_t + U_i h_{t-1} + b_I) \tag{4}$$
$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{5}$$
$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{6}$$
$$\widetilde{c} = f_t \odot c_{t-1} \tag{7}$$
$$h_t = o_t \odot tanh(c_t) \tag{8}$$

where x_t and c_t denote the input sample at time t and the memory unit respectively. (b_f, b_I, b_o),(W_f, W_I, W_o), and (U_f, U_I, U_o) stands for the biases, recurrent weights, and input weights for each gate respectively. The function $\odot$ is the corresponding multiplication of the elements. h_{t-1} is the hidden layer for the respective gates x in the current timestamp. To improve the cognition using a DL approach, the recurrent connections could be a chain of stacked LSTM fully connected to form a sequence base model. Let $x = [x_1, \ldots, x_t]$ be the input sequence for LSTM cells with $x_t \in \mathbb{R}^k$. Here, K denotes the dimensional vector of the original data set at the t^{th} timestep.

Firstly, the h_{t-1}, c_{t-1}, and x_t pass the input information to The LSTM unit. The LSTM gates interact with the inputs to generate an action based on a logic function. After passing by f_t, a new cell state c_t is built. x_t and h_{t-1} move to the forget layer to quantify the importance of the information between 0 and 1. At this stage, the f_t gate takes a decision whether if the information has to be stored, maintained, or removed. Then, the forget gate will update the cell state c_t with the new important information based on the proportion of the information occupied by the actual and the previous cell state. The final hidden layer of the LSTM is computed to obtain the remaining state value.

C. TABU SEARCH ALGORITHM

The identification of the most suitable configuration for DL models is essential to meet the optimal performance for a specific dataset [38]. To solve complex optimization problems, TSA is widely adopted as a global stepwise metaheuristic algorithm. The key idea of the TS algorithm consists of employing a tabu table to memorize the movements that happened in the previous iterations in order to avoid cycles. This is conducted by blocking the overlook of the traced movements registered in the tabu table [39]. The optimal solution in the neighborhood is chosen. The mechanism of TSA is illustrated in Fig. 3:

Referring to the flowchart in Fig. 3, define $\mathbb{A} = \{a_1, \ldots, a_M\}$ where M is the cardinality of $\mathbb{A}$. Let a set $S(a_q), q \in \{1, \ldots, M\}$, a fixed subset $\mathbb{A}\backslash a_q$. a_q denote the symbol-neighborhood of a_q. define $w_v(a_q), v = 1, \ldots, N$ as

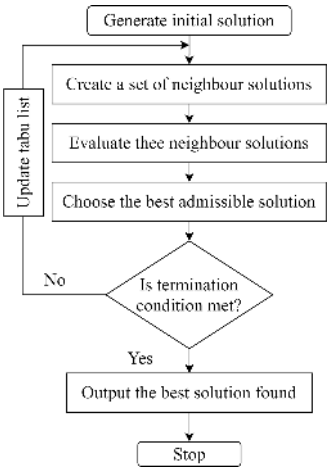


FIGURE 3. Flowchart of the standard TSA.

the vth symbol-neighbor of a_q. Assuming $x^m = [x_1^m, x_{n_t}^m]$ with $x_i^m \in \mathbb{A}$ a solution candidate in the mth iteration. The (u, v)th neighborhood vector $z^m(u, v)$ of x^m, $u = 1, n_t$, $v = 1, \ldots, N$ based on Euclidean distance can be written as [40]:

$$z^m(u, v) = [z_1^m(u, v), z_2^m(u, v) \ldots, \ldots, z_{n_t}^m(u, v)] \tag{9}$$

The (u, v)th neighbor satisfies the following condition [40]:

$$z_i^m(u, v) = \begin{cases} x_i^m & for \quad i \neq u \\ w_v(x_u^m) & for \quad i = u \end{cases} \tag{10}$$

In the solution space, $n_t N$ vectors are generated from a defined vector in a specific coordinate. Next, the TSA passes to the best vector among the neighborhood while satisfying the non-move back towards the previous schemes. The Tabu matrix (T) of size $n_t M \times N$ and t coordinates representing the tabu values of moves is computed as:

$$T = \begin{pmatrix} t_{1,1} & t_{1,(n_t-1)M+1} & \cdots & t_{1,n_t M} \\ \vdots & \vdots & \ddots & \vdots \\ t_{N,1} & t_{N,(n_t-1)M+1} & \cdots & t_{N,n_t M} \end{pmatrix} \tag{11}$$

Define g^m the vector with the maximum-likelihood performance cost generate until the mth operation of TSA. The shift (u_1, v_1) is accepted if one of the following conditions is respected [40]:

$$\varphi(z^m(u_1, v_1)) \leq \varphi(g^m) \tag{12}$$
$$T(u_1 - 1)M + q, v_1 = 0 \tag{13}$$

where the term q is defined as $a_q = x_{u1m}, a_q \in \mathbb{A}$. The next move is calculated as [40]:

$$(u_2, v_2) = \underset{u,v;u\neq u_1,v\neq v_1}{arg} \varphi(z^m(u, v)) \tag{14}$$

The computing work is repeated until the optimal solution is met. TSA is advantageous in terms of adaptability, robustness, and accuracy. Particularly, the strong neighborhood local search ability of TSA and its super-fast searching speed is found to be very suitable in case of high dimensions of the

hyperparameters [39]. This justifies its usefulness for LSTM tuning operation.

D. PROPOSED NARX-LSTM ARCHITECTURE

The proposed model (NARX-LSTM) is a vigorous combination of NARXNN and LSTM models. NARX-LSTM model merges the properties of individual predictors to achieve better results. In the proposed system architecture, NARXNN is associated with embedded memories that make jump-ahead connections in the time-unfolded network. The embedded memory of NARXNN consists of simple tapped delayed values to neurons. The NARXNN is used hereinafter to calculate the residual error correction. The residual error correction is implemented to reduce the sensitivity of the network to time dependencies. We assume that the additional tapped time delays from NARXNN improve the accuracy of the prediction engine using the error Hankel matrix to increase the weight of the residual error correction. Let $E_n = [e_1, \ldots, e_n]$ is the error vector between the real values $Y_t = [y_1, \ldots, y_n]$ and the forecasted values of NARXNN $\hat{Y}_t = [\hat{y}_1, \ldots, \hat{y}_n]$. The residual error is calculated as follows:

$$E_n = Y_t - \hat{Y}_t = y_i t - \sum_{i=1}^{n} w_i F_{it}, \quad (t = 1, \ldots, n) \quad (15)$$

where F and w denote the nonlinear mapping function of NARXNN and the corresponding weight value. A Hankel matrix is built from the expansion of the error vector into a multi-dimensional data matrix in order to capture the dynamic change of the model variation as:

$$E_n^* = [e_1^*, \ldots, e_L^*] = \begin{pmatrix} e_1 & e_2 & \cdots & e_L \\ \vdots & \vdots & \ddots & \vdots \\ e_K & e_{K+1} & \cdots & e_{K+L-1} \end{pmatrix} \quad (16)$$

where $L \in [2 \leqslant L \leqslant L(n/2)]$ is the window length and $K = n - L + 1$ is the number of overlapping segments. Next, two types of transformations are computed for the Hankel error matrix: MinMax transformation and Hilbert transformation (HT) [41], [42]. The MinMax transformation is calculated as follows [42]:

$$x_n = \frac{x_r - x_{min}}{x_{max} - x_{min}} \quad (17)$$

where x_n denotes the normalized value, x_r is the real value. Here, x_{min} and x_{max} are the minimum and maximum values. Let $f(x)$ be the error function and $H()$ the Hilbert transform operator. The HT $g(x)$ of $f(x)$ is defined and altered as follows [41]:

$$\begin{aligned} g(y) = H(f(x)) &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(x)\mathrm{d}x}{x - y} \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(x - y)\mathrm{d}x}{x} \end{aligned} \quad (18)$$

The HT can be reformulated as a convolution by:

$$g(y) = \frac{1}{\pi X} f(x) \quad (19)$$

A fast algorithm based on convolution theorem is employed to develop the HT as [41]:

$$g(y) = ifft[fft(\frac{1}{\pi x}) fft(f(x)))] \quad (20)$$

where $fft()$ and $ifft()$ denote the fast Fourier transform and the inverse fast Fourier transform respectively satisfying [41]:.

$$fft(\frac{1}{\pi x}) = -jsgn(freq) \quad (21)$$

where $freq$ and sgn denote the frequency and the sign function respectively. From Eq.20, the product $fft(\frac{1}{\pi x}) fft(f(x)))$ presents special filtering of Fourier transform and the input signal. This operation transforms the frequency components by computing a phase shift of -90 ° for positive frequency and 90° for negative frequency. The calculation of the inverse Fourier transform on the product leads to generate the HT of the error vector. HT is characterized by a good multi-resolution for signals analysis. The proposed NARX-LSTM model is built from two stages: In the first stage, the NARXNN receives the weather information to primarily predict the PV power $\hat{Y} = [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n]$. The temporary PV power forecasts from NARXNN are used to calculate the error values and compute the vector error correction. The vector error correction is calculated by the association of the MinMax transform and HT of the error values vector generated by NARXNN. The vector error correction function (V) is fed with the original database to LSTM model optimized by TSA as:

$$X(t) = x_i(t) + V[y(t); x_i(t)] \quad (22)$$

By adopting this methodology, the LSTM network acquires the learning potential of NARXNN to improve the pattern recognition of the forecasting system. The LSTM gates use the processed data as follows:

$$i_t' = \sigma(W_I(x_t + V[y_t; x_t]) + U_I h_{t-1} + b_I) \quad (23)$$
$$f_t' = \sigma(W_f(x_t + V[y_t; x_t]) + U_f h_{t-1} + b_f) \quad (24)$$
$$o_t' = \sigma(W_o(x_t + V[y_t; x_t]) + U_o h_{t-1} + b_o) \quad (25)$$
$$\widetilde{c'} = f_t \odot c_{t-1}' \quad (26)$$
$$h_t' = o_t' \odot tanh(c_t') \quad (27)$$

where (i_t', f_t', o_t') denote the improved input, forget, and output gates respectively. $\widetilde{c'}$ and c' denote the cell state and the memory unit for NARX-LSTM respectively. The reason behind the selection of NARXNN is due to its successful implementation in handling the latching problem and nonlinear system identification. On the other hand, LSTM memories decrease the vanishing gradient. Adding to LSTM architecture, a general class of regression model with time delay has been given the main importance in this investigation. The forecasting system could be stabilized or destabilized by certain stochastic inputs. Hence, it is significant to consider stochastic effects on the steadiness property of the delayed information. For a better understanding of the proposed model, the data processing is resumed in four stages as shown in Fig. 4.

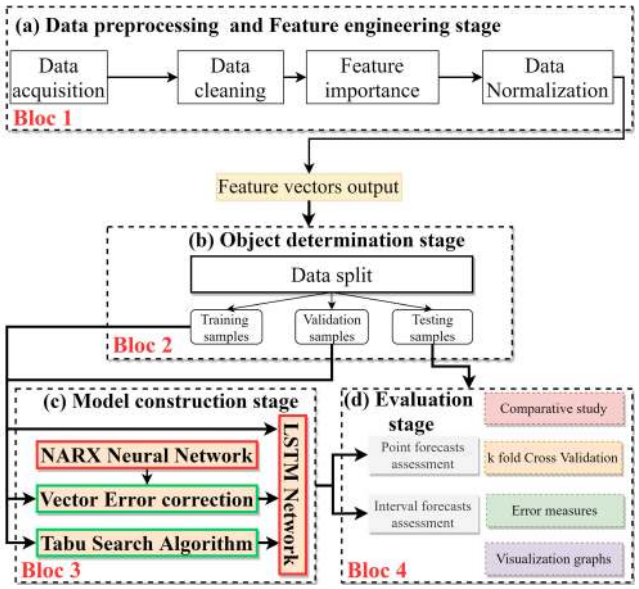


FIGURE 4. Proposed NARX-LSTM architecture for PVPF.

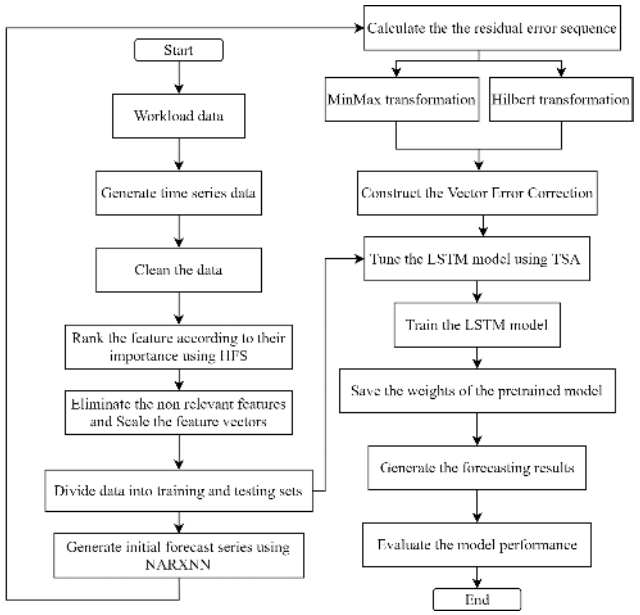


FIGURE 5. Flowchart of the proposed forecast method.

In the first stage, data preprocessing and feature engineering is computed starting from data acquisition to store useful information about the state of the PV system. Then, data cleaning from missing and invalid samples and feature importance evaluation are conducted to avoid its adverse impact on the forecasting system accuracy. Finally, the data is normalized with a magnitude range of [0, 1] using Min-Max method [42]. This normalization prevents the features from getting affected by the bad influence of the outliers and data scale while increasing the convergence speed and the performance of the model. In the second stage, a typical data separation procedure between inputs and outputs from the training and testing folds is computed in the object determination stage. Then, the training data is fed firstly to NARXNN model to generate the PV power, then the output is used to calculate the vector error correction to be fed to LSTM model optimized by TSA. The output of the model construction stage is the PV power point forecasts and its associated uncertainties. From this explanation, NARX-LSTM could be described as an association of two module types: (1) NARXNN that receives the sequence of external inputs as well as the recurrent output layer state; (2) LSTM cells that receive corrective information with the original features to generate the final outputs. Thus, the residual information fed from the NARXNN output is reused for feeding LSTM model as an additional source of information apart from the standard training data. The flowchart of the proposed model is shown in Fig. 5.

III. CASE STUDY AND SIMULATION RESULTS

In this section, the feature engineering process is conducted based on the Australian weather parameters. Then, the simulation results are presented. Finally, a comparative study is deployed to assess the performance of the proposed model. The objective is to demonstrate the capabilities of NARX-LSTM for tackling TSF with high accuracy.

A. FEATURE ENGINEERING

To evaluate the performance of the proposed NARX-LSTM model using a real dataset, the experimental study employs two PV power plants from different locations in order to validate the generalization capabilities of the proposed model.

$\mathfrak{D}_1$: the public data from Desert Knowledge Alice Springs Center (DKASC) in Central Australia (23.7618° S, 133.8749° E) is used for the first case study [43]. Starting in September 2008, the DKASC consists of a demonstration facility of 38 sites to build a high confidence level of PV technologies with different manufactures and stakeholders [43]. These sites consist of real-life monitoring of PV technologies from various types, models, and configurations as presented in Table 2-3. The specific characteristics of the two years from DKASC's public-facing repository were deeply analyzed to target the PV power output from DKASC online portal (DKASC 2019). The inputs of our system comprise ambient temperature Celsius (T in (°C)), Wind direction (Wd in Â°), Horizontal Radiation (Ir in (W/m^2)), Diffuse horizontal radiation (DIr in (W/m^2)), and Relative humidity (Rh in (%)) while the output is the active power (kW). It is worth noting that the yearly lagged PV power has been added as an additional input for the first case study. The yearly lagged PV power represents the historical PV power at the same time of the previous year.

$\mathfrak{D}_2$: Urbana-Champaign solar farm-Flyover (UCF) data set from USA (40°06'07.3°N 88°13'37.5°W) is used for the second case study [44]. The weather data is collected from the National Oceanographic and Atmospheric Administration (NOAA). Starting from its first year of operation, The UCF

TABLE 2. The related characteristics of the PV plant.

System specification	Characteristics
$\mathfrak{D}_1$	
Array rating	191.74 kW
Average of Powering	141house
Location	Alice Springs, Australia
PV technology	crystalline Silicon, CdTe/CIGS
Array area	$4 * 38.37m^2$
Type of tracker	Fixed: Ground Mount, Single Axis, Dual Axis
Inverter size/type	46 kW, SMA/Sunny Mini Central 6000A
$\mathfrak{D}_2$	
Array rating	5.87 MWp
Location	South of Windsor Road
Array area	20.8 acres
Average of powering	2% for the Urbana campus
Number of panels	18867 panels
Carbon reduction	6000 metric tons/year

TABLE 3. Data types.

Case study	Data	timestep	Training year	Testing year
Case study I	$\mathfrak{D}_1$	5min	2017	2018
Case study II	$\mathfrak{D}_2$	1hour	2016	2017

plant approximately generates 7.28 megawatt-hours (MWh). This latter is considered the largest solar array installation for Urbana campus. The used data comprises T, Rh, Cloud coverage (Cc), visibility, Wind speed (Ws), station Pressure (P), Altimeter indication (A), and PV power. More details about the characteristics of UCF is found in Table 2-3.

It is worth noting that the database is cleaned from missing values and smoothed from instrumental malfunction measures. Furthermore, the data rows containing the low irradiance measurements have been eliminated to enhance the prediction potential of the proposed method. The input parameters have a direct relation with the predicted PV power output. However, this relationship is not equally partitioned. Various models have been proposed to measure the diversity of feature importance. In this study, the idea is to permit one parameter and estimate the increase of forecasting error in each case by the probability value (P-value) calculation. The domain knowledge is investigated using a combination of Elastic Net and Extreme Gradient Boosting (XGB) evaluation models to enhance the features selection reliability. Table 4 presents the numerical measures of P values for feature inputs.

According to Table 4, the irradiation from $\mathfrak{D}_1$ and the temperature and relative humidity from $\mathfrak{D}_2$ have the highest impact on PV power estimation. For the sake of simplicity, the time series data representation will include all the

TABLE 4. Feature importance analysis with $\mathfrak{D}_1$ and $\mathfrak{D}_2$.

Features	XGB weights	Elastic Net weights
$\mathfrak{D}_1$		
Horizontal radiation	1.9460 ± 0.0070	1.9917 ± 0.0075
Temperature	0.0274 ± 0.0005	0.0175 ± 0.0002
Lagged PV power	0.0212 ± 0.0002	0.0080 ± 0.0001
Relative humidity	0.0071 ± 0.0003	0.0017 ± 0.0001
Diffuse horizontal radiation	0.0064 ± 0.0000	0.0048 ± 0.0001
Wind direction	0.0046 ± 0.0001	0.0014 ± 0.0000
$\mathfrak{D}_2$		
Relative humidity	0.4104 ± 0.0127	0.1641 ± 0.0054
Temperature	0.3739 ± 0.0074	1.6558 ± 0.0229
Station pressure	0.0910 ± 0.0043	0.0545 ± 0.0033
Wind speed	0.0785 ± 0.0042	0.0243 ± 0.0027
Altimeter indication	0.0348 ± 0.0019	0.0013 ± 0.0001
Cloud coverage	0.0342 ± 0.0029	0.0000 ± 0.0000
Dew point	0.0274 ± 0.0021	0.9074 ± 0.0091
Visibility	0.0030 ± 0.0004	0.0054 ± 0.0015

examined features as inputs associated with the date features (minute/hour/day/month/year) to forecast the future PV power. The date features are converted using one-hot encoder to fit the model requirements. The two data sets were split into training and testing folds. A full year $\{2017_{\mathfrak{D}_1}, 2016_{\mathfrak{D}_2}\}$ is dedicated to model training and the rest of $\{2018_{\mathfrak{D}_1}, 2017_{\mathfrak{D}_2}\}$ data is employed for testing purposes.

B. PERFORMANCE EVALUATION OF THE PROPOSED FORECAST SYSTEM

The evaluation of the proposed NARX-LSTM model is performed to statistically confirm its high performance using real data sets. Due to the relative effectiveness of score metrics under certain testing conditions, multiple statistical error measures were employed for a reliable forecasting performance assessment. In this paper, the evaluation procedure is twofold:

- Point forecast assessment is quantitatively computed where a series of error metrics were employed. The selected performance evaluation measures for point forecasting include normalized RMSE (nRMSE), normalized MAE (nMAE), and squared coefficient of determination (R^2) [23], [31]. These percentage error measures are employed due to their scale-independent propriety and their efficiency. Thus, these measures are also used in this paper for comparative purposes with other benchmarks from different scaled databases.
- Interval forecast criteria of PV power include the Average Coverage Error (ACE), Prediction Interval Normalized Average Width (PINAW), Prediction Interval Nominal Confidence (PINC) [13]. The PINAW index gives information about the model sharpness.

Furthermore, the normalized pinball Loss function (L) is employed for its wide applicability in the interval assessment [45].

In this paper, a normalization is adopted in order to make the error metrics scale-free when comparing the proposed model with other data sets. The normalization is conducted by the mathematical division by the PV capacity for each PV plant. The error equations between the actual PV power (y_i) and the forecasted power ($\hat{y}_i$) for n timesteps are given as [31], [45]:

$$nMAE = 100(\%)\frac{1}{n}\frac{\sum_{i=0}^{n-1}|y_i - \hat{y}_i|}{y_c} \tag{28}$$

$$nRMSE = 100(\%)\sqrt{\frac{1}{n}\frac{\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2}{y_c}} \tag{29}$$

$$R^2 = 1 - \frac{\sum_{i=0}^{n-1}(\hat{y}_i - y_i)^2}{\sum_{i=0}^{n-1}(\bar{y}_i - y_i)^2}, \bar{y} = \sum_{i=0}^{n-1} y_i \tag{30}$$

$$PICP = \frac{1}{n}\sum_{i=0}^{n-1}\zeta, \zeta = \begin{cases} 1, T_i \in [L_i, U_i] \\ 0, T_i \notin [L_i, U_i] \end{cases} \tag{31}$$

$$PINAW = \frac{1}{nR}\sum_{i=0}^{n-1}(U_i - L_i) \tag{32}$$

$$ACE = PICP - PINC \tag{33}$$

$$L_{m,t}(q_{m,t}, p_t) = \begin{cases} (1 - \frac{m}{100})(q_{m,t} - p_t), & p_t < q_{m,t} \\ \frac{m}{100}(p_t - q_{m,t}), & p_t \geqslant q_{m,t} \end{cases} \tag{34}$$

where n and y_c denote the total number of samples and the capacity of the PV plant, i.e., 211.259 kW and 4733.25 kW for $\mathfrak{D}_1$ and $\mathfrak{D}_2$ respectively. q, $q_{m,t}$ and p_t denote the quantile, mth normalized actual quantile, and normalized predicted quantile at time t. In this paper, Averaging the quantiles' $L_{q,t}$ values is adopted to obtain the overall pinball loss score for a specific forecasting horizon. Moreover, the evaluation of the proposed model is conducted using 10-fold cross-validation for better reliability of the evaluation procedure.

TABLE 5. Experimental environment.

Experimental Environment	Proprieties
Operating system	Windows 10
Processor	Intel Core i7 9th Gen
Graphics card	NVIDIA GeForce GTX 1650
Programming language	Python 3.6.7
Deep learning framework	Keras 2.4.3
Platform	Tensorflow 2.3.0
Supplementary tools	Sklearn 0.23.2

The TSA is implemented using hyperactive library [48]. This algorithm uses a defined search space to identify the best hyperparameters for a given data set. In order to avoid the verbose computational burden, extensive experiments were conducted to minimize the search space. It has been found that the configuration of 20 NARXNN neurons, a delay value of 4, a Levenberg-Marquardt optimization algorithm, and three LSTM layers generates the best results. Then, LSTM units, dropout values, activation, and optimizer functions are found using TSA optimization. In order to verify the model generalization for a PV farm with a higher scale, the automatic search mechanism is conducted only for $\mathfrak{D}_1$. The search space for TSA algorithm and the optimization results are found in Table 6.

TABLE 6. Search space HO of the proposed model.

Model	Configuration	Search Space	TSA
LSTM	LSTM layer1	[64, 512]	512
	LSTM layer2	[36, 256]	256
	LSTM layer3	[36, 256]	36
	Dropout 1	[0.1, 0.5]	0.1
	Dropout 2	[0.1,0.5]	0.1
	Dropout 2	[0.1,0.5]	0.1
	activation	[Sigmoid,Softplus ,Selu,Elu,Softmax]	Softplus
	optimizer	[adam,SGD,RMSprop]	adam
Time (min)			189.52
Time/iteration (min)			18.95

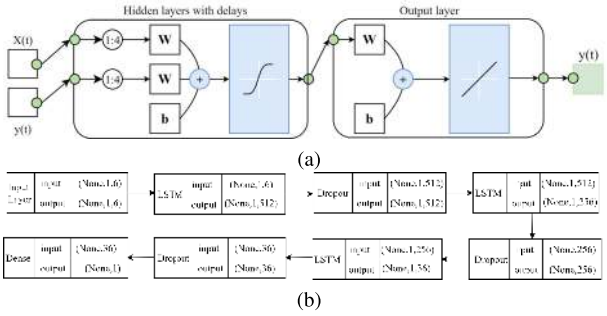
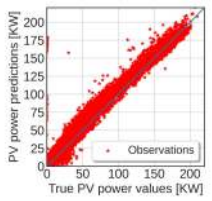
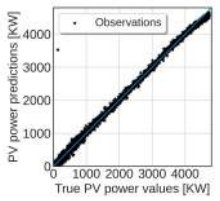


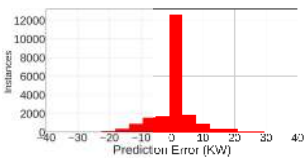
FIGURE 6. Final architecture of the NARX-LSTM model after TSA optimization with (a)NARXNN module(b) LSTM module.



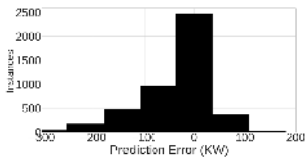
(a) Scatter plot with $\mathfrak{D}_1$.



(b) Scatter plot with $\mathfrak{D}_2$.



(c) Error distribution with $\mathfrak{D}_1$



(d) Error distribution with $\mathfrak{D}_2$

FIGURE 7. Scatter plots ((a), (b)) and error distributions ((c), (d)) of PV measured power and forecasted power for two PV datasets $\mathfrak{D}_1$ and $\mathfrak{D}_2$.

The LSTM model uses 100 iterations with an early stopping function to avoid overfitting problems. To get better visibility of the model performance, the scatter plots and the error distributions are illustrated in Fig. 7.

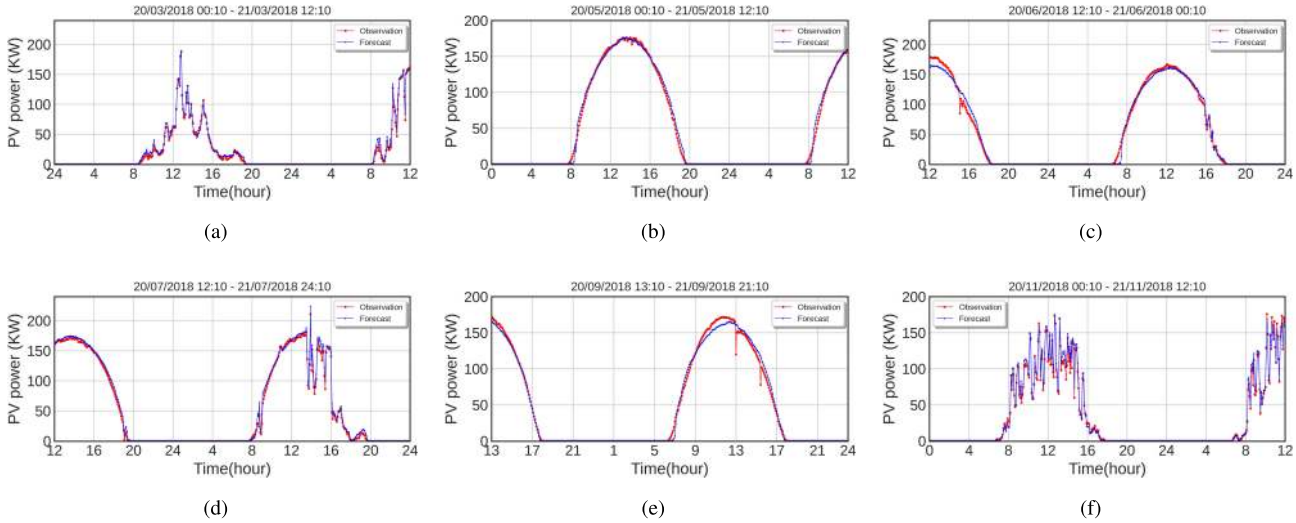


FIGURE 8. Multiple tests of NARX-LSTM model for 5-min horizon forecasting on $\mathfrak{D}_1$ under different weather conditions.(a) March 20, Rainy. (b) May 20, sunny. (c) June 20, sunny. (d) July 20, sunny to cloudy.(e) September 20, Sunny. (f) November 20, foggy to cloudy.

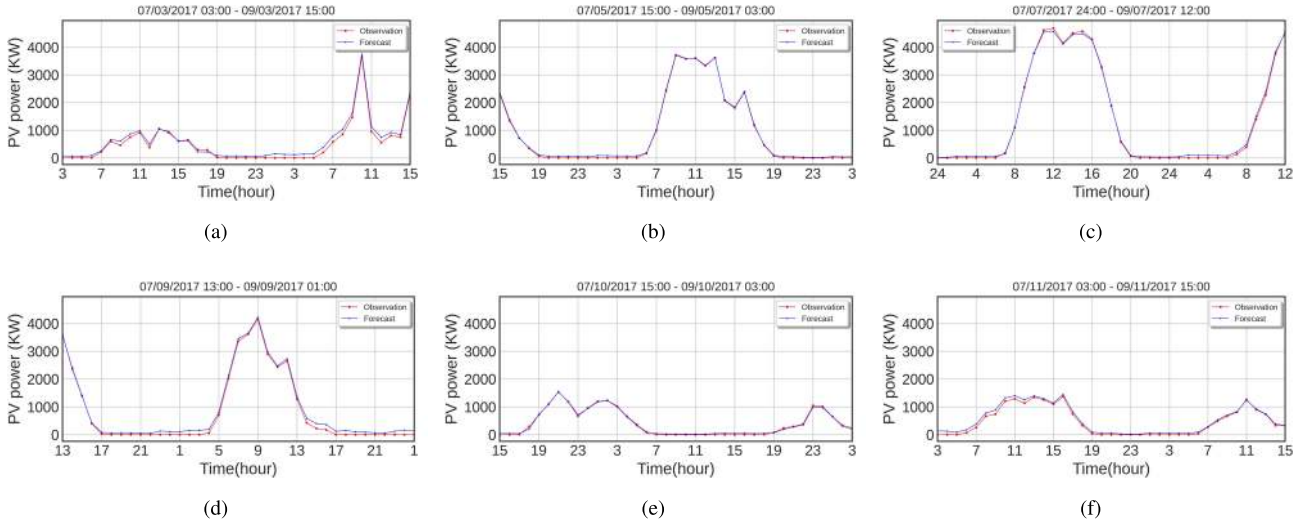


FIGURE 9. Multiple tests of NARX-LSTM model for 1-hour horizon forecasting on $\mathfrak{D}_2$ under different weather conditions:(a)March 7,Rainy. (b) May 7, sunny to cloudy. (c) July 7, sunny. (d) September 7, sunny to cloudy.(e) October 20, Rainy. (f) November 20, foggy to cloudy.

C. POINT PV POWER FORECASTING

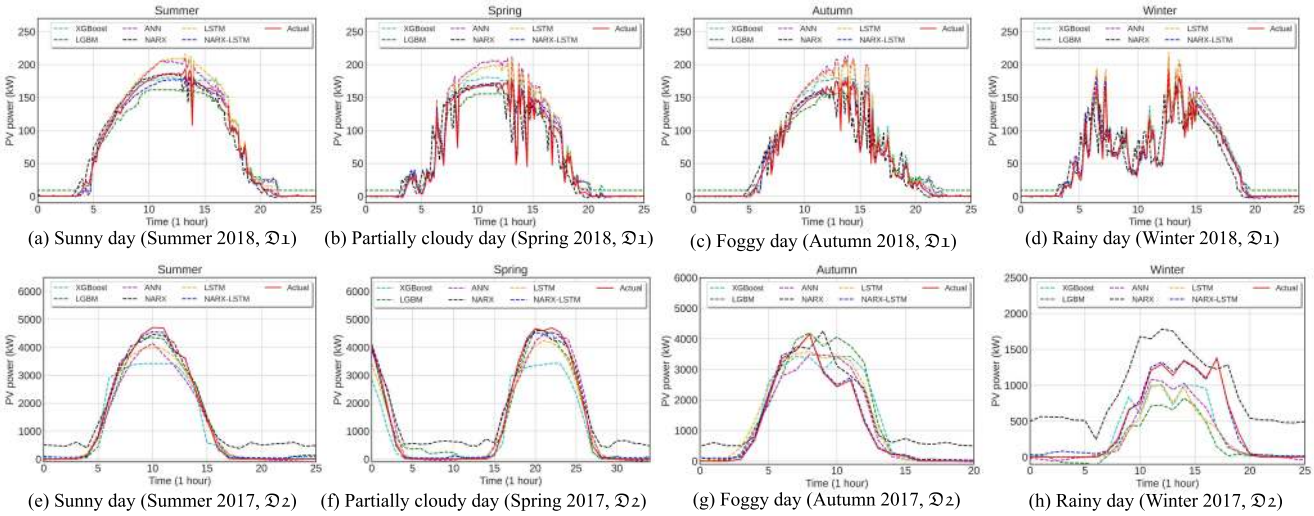
This section tackles the numerical validation of the model performance using two data sets from different locations (USA and Australia) with different speeds of data acquisition. The feasibility of NARX-LSTM model is conducted using Python programming language. The NARX-LSTM model is implemented using Keras and Sckit-learn packages [46], [47]. The model is evaluated on single-step and multi-step ahead respectively. The experimental environment is described in Table 5.

From Table 6, it can be seen that the LSTM layers are fixed at 512, 256, and 36 units for the three LSTM levels respectively. The dropout values are fixed at 0.1 for all the aforementioned levels. The optimizer function is adam. The most suitable activation function, Softplus, is calculated as follows:

$$f(x) = log(1 + exp(x)) \tag{35}$$

It bears noting that the TSA is computationally demanding with a conversion speed of 18.95 min per iteration. This low convergence speed could be explained by the fact that the high-resolution data set for a small-timestep of 5 minutes causes an exhaustive calculation due to a large number of investigated patterns. The final architecture of the proposed model is shown in Fig. 6.

As can be seen from the scatter plots with $\mathfrak{D}_1$ and $\mathfrak{D}_2$ in Fig. 7, the proposed model follows the actual PV power with high accuracy. Despite the low error values, the 1 hour ahead daily PV predictions on $\mathfrak{D}_2$ generates better results. Fig. 8 and Fig. 9 illustrate, for 5 min and 1 hour ahead,



(a) Sunny day (Summer 2018, $\mathfrak{D}_1$) (b) Partially cloudy day (Spring 2018, $\mathfrak{D}_1$) (c) Foggy day (Autumn 2018, $\mathfrak{D}_1$) (d) Rainy day (Winter 2018, $\mathfrak{D}_1$)

(e) Sunny day (Summer 2017, $\mathfrak{D}_2$) (f) Partially cloudy day (Spring 2017, $\mathfrak{D}_2$) (g) Foggy day (Autumn 2017, $\mathfrak{D}_2$) (h) Rainy day (Winter 2017, $\mathfrak{D}_2$)

FIGURE 10. Daily PVPF results for 5-min and 1-hour ahead.

the actual/predicted daily PV power for different weather conditions (sunny, partially cloudy, foggy, rainy, and cloudy days).

According to Fig. 8-9, the estimated power tracks the real one with great exactitude. It can be seen that the results are very promising especially during sunny days. However, the model accuracy is degraded in fast changes of the weather conditions (Fig. 8.(e)). Nevertheless, even with fast fluctuations of the weather parameters, the NARX-LSTM network is able to follow the general shape of the real values. For a quantitative assessment with error metrics and contrast models, Table 7 presents the characteristics of reference models including LSTM, NARX, XGBoost, ANN, and LGBM. It worth noting that Random Search method is adopted to train XGboost, LGBM ANN. Moreover, the selected hyperparameters of the LSTM and NARXNN models are employed in the tuning of the proposed NARX-LSTM model. All the proposed models use the same configuration from Table 5.

To begin with, the comparative study is conducted based on different climate conditions (sunny, partially cloudy, cloudy, and rainy). The point forecasting measures are adopted for the performance comparison. The simulation results are shown in Fig. 10.

According to Fig. 10, it can be clearly observed that the forecasting performance highly depends on a particular season of the year. Specifically, the prediction performance of NARX model decreases significantly during the winter days (Fig. 10. (d) and Fig. 10(h)). According to the simulations, NARX-LSTM has been seen performing best than other individual models. The performance superiority is observed on winter and spring days. With more detailed information comparing the original LSTM with the proposed model, it can be said that the error correction vector has significantly enhanced its performance in capturing the trend of the actual PV power. From Fig. 10, it has been found that

TABLE 7. Hyperparameters settings for reference models.

Base models	Hyperparameter settings
LSTM	The number of LSTM units for the first layer is 512 with a dropout of 0.1; the number of LSTM units for the second layer is 256 with a dropout of 0.1; the number of LSTM units for the third layer is 36 with a dropoutof 0.1; The optimizer function is Adam; The activation is Softplus.
NARXNN	The maximum iteration number is 1000; The neuron number is 20; The delay function is 4.
XGBoost	The number of estimators is 50; The learning rate is 0.001; The tree complexity is 2; The gamma value is 0.5; The max depth is 6.
ANN	The maximum iterations are 1000; The hidden layer sizes is 96, and 30; The activation function is relu, the solver is Adam.
LGBM	The number of estimators is 50; the learning rate is 0.001; The max depth is 6.

TABLE 8. Testing time (second) obtained for different PV power forecast models with two datasets.

Model	Proposed	LSTM	XGB	LGBM	ANN	NARX
$\mathfrak{D}_1$	3.37	2.821	0.806	0.682	0.637	1.578
$\mathfrak{D}_2$	0.904	1.773	0.613	0.636	0.626	0.84

the performance superiority has been attributed to NARX-LSTM. The computational complexity of the testing model is resumed in Table 8.

According to Table 8, the simple ANN offers the least computation time. The proposed model is relatively more computationally demanding than the original LSTM (case I), while on $\mathfrak{D}_2$, the proposed model took less time than the original LSTM. Despite the longer testing calculation of the proposed NARX-LSTM model, it is easily applicable to real-world experiments (especially with the popularization of

TABLE 9. Score performance comparison on $\mathfrak{D}_1$ and $\mathfrak{D}_2$.

		$\mathfrak{D}_1$			$\mathfrak{D}_2$		
Weather	**Model**	$\mathbf{R}^2$ $(\%)\pm SD$	**nRMSE** $(\%)\pm SD$	**nMAE** $(\%)\pm SD$	$\mathbf{R}^2$ $(\%)\pm SD$	**nRMSE** $(\%)\pm SD$	**nMAE** $(\%)\pm SD$
Sunny	NARX-LSTM	99.40	2.39	1.46	99.83	1.54	1.39
	LSTM	95.04	7.98	6.38	95.72	6.74	4.3
	XGB	96.91	6.05	4.8	90.25	9.76	5.65
	LGBM	97.52	3.98	4.35	98.42	4.47	3.04
	ANN	97.01	4.93	5.95	93.81	8.08	5.73
	NARX	97.86	4.67	3.26	92.81	7.92	8.65
Partially cloudy	NARX-LSTM	99.27	2.14	1.23	99.91	1.09	0.96
	LSTM	95.88	5.6	3.82	95.11	7.14	4.64
	XGB	94.53	6.58	4.31	81.5	12.56	8.45
	LGBM	96.51	3.92	4.23	97.84	5.05	4.08
	ANN	96.78	3.28	4.89	96.52	6.66	5.24
	NARX	77.85	11.47	6.15	90.13	10.11	9
Cloudy/foggy	NARX-LSTM	99.29	2.33	1.5	99.62	1.88	1.63
	LSTM	95.59	6.59	4.93	88.67	10.34	6.84
	XGB	95.82	6.15	4.56	83.94	12.6	7.49
	LGBM	96.77	3.95	4.44	84.06	13.19	7.82
	ANN	96.93	4.05	5.3	91.03	9.25	7.19
	NARX	78.24	13.14	8.04	80.38	12.36	10.59
Rainy	NARX-LSTM	97.4	1.07	0.62	99.41	0.82	0.67
	LSTM	95.28	1.62	1.12	57.89	5.09	2.88
	XGB	92.2	1.9	1.14	54.42	5.49	3.16
	LGBM	36.12	4.14	3.92	48.98	7.7	4.58
	ANN	97.09	1.11	0.81	25.84	8.24	6.1
	NARX	68.62	4.15	2.27	87.67	7.46	4.25
Overall	NARX-LSTM	**98.84**$\pm$0.83	**1.98**$\pm$0.53	**1.20**$\pm$0.35	**99.69**$\pm$0.19	**1.33**$\pm$0.41	**1.16**$\pm$0.37
	LSTM	95.45$\pm$0.32	5.45$\pm$2.37	4.06$\pm$1.93	84.35$\pm$15.52	7.33$\pm$1.90	4.67$\pm$1.42
	XGB	94.86$\pm$1.75	5.17$\pm$1.90	3.99$\pm$1.76	77.53$\pm$13.7	10.10$\pm$2.90	6.19$\pm$2.02
	LGBM	81.73$\pm$26.34	4.62$\pm$0.49	3.94$\pm$0.02	82.33$\pm$20.09	7.60$\pm$3.45	4.88$\pm$1.79
	ANN	96.95$\pm$0.11	3.34$\pm$1.41	4.24$\pm$2.01	76.80$\pm$29.49	8.06$\pm$0.92	6.07 $\pm$0.72
	NARX	80.64$\pm$10.66	8.36$\pm$4.01	4.93$\pm$2.29	87.75$\pm$4.63	9.65$\pm$1.83	8.37$\pm$2.26

Note: The best information criterion is in boldface.

cloud computing). Table 9 resumes the score performance measures results on $\mathfrak{D}_1$ and $\mathfrak{D}_2$ respectively.

According to the numerical results presented in Table 9, it can be concluded that the combination NARX-LSTM outperforms other individual models. The NARX-LSTM generates a mean nRMSE $= 1.98\%$ and nRMSE $= 1.33\%$ on $\mathfrak{D}_1$ and $\mathfrak{D}_2$ respectively. The obtained results confirm the high performance of the proposed model in TSF.

As the paper attempts to offer a clear assessment of NARX-LSTM model, a fair comparison is conducted with the recent benchmark models, specifically, Ensemble of Methods (ENS), improved ANN, Grey-Box model (GB), Correlation model, Extreme Learning Machines (ELM), modified LSTM, Differential Polynomial Neural Network (D-PNN), Autoencoders LSTM (AE-LSTM), and Wavelet transform-Radial Basis function neural network- Particle Swarm Optimization (WT-RBFNN-PSO). The comparison of the proposed NARX-LSTM model with the cited models is conducted based on hourly ahead daily PV power forecasting. Thus, the NARX-LSTM performance is only included with $\mathfrak{D}_2$ in this comparative experiment. Table 11 includes the performance of contrast models with NARX-LSTM model based on their mean nRMSE measures as the main performance index between different models.

As can be seen from Table 11, the NARX-LSTM model could achieve the best results based on the overall forecasting performance in point forecasting with one-hour resolution horizon. The mean nRSME of NARX-LSTM is nRMSE $=$ 1.33% while the closest model (WT-RBFNN-PSO) achieves a mean nRMSE $= 1.85\%$. The proposed NARX-LSTM model is advantageous for short-term PVPF in terms of forecasting accuracy. In order to validate the model performance for multistep forecasting, different forecasting horizons were applied for the forecasting settings. Due to the difference of $\mathfrak{D}_1$ and $\mathfrak{D}_2$ resolution, $\mathfrak{D}_1$ is assigned to forecasts one hour, 6 hours, and 12 hours ahead while $\mathfrak{D}_2$ will predict 3 hours, 6 hours, and one day ahead of PV power. The simulation results are shown in Fig. 11, where four random days from the testing set were selected to conduct these experiments.

Regarding the model performance for different forecasting horizons in Fig. 11, it can be said that the general shape of the PV power is followed with good exactitude for 1 hour 3 hour, and 6-hours ahead on $\mathfrak{D}_1$ and $\mathfrak{D}_2$. however, the proposed model can be slightly inaccurate when the forecasting

TABLE 10. Score performance comparison using $\mathfrak{D}_1$ and $\mathfrak{D}_2$.

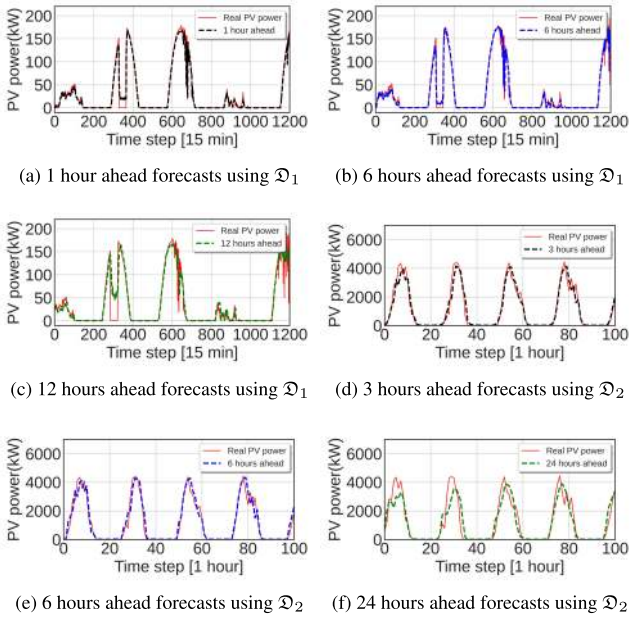
	$\mathfrak{D}_1$							$\mathfrak{D}_2$						
Fold	**Horizon**	**R^2 (%)**		**nRMSE (%)**		**nMAE (%)**		**Horizon**	**R^2 (%)**		**nRMSE (%)**		**nMAE (%)**	
		train	test	train	test	train	test		train	test	train	test	train	test
1	1 hour	99.86	99.83	1.08	1.17	0.61	0.63	3 hours	93.15	93.16	7.47	7.65	3.72	3.75
2		99.88	99.88	0.99	0.98	0.5	0.5		93.89	92.9	7.07	7.61	3.63	3.81
3		99.84	99.75	1.15	1.46	0.63	0.63		93.52	94.53	7.28	6.72	3.78	3.51
4		99.89	99.91	0.94	0.88	0.42	0.42		93.67	94.08	7.21	6.9	3.61	3.63
5		99.9	99.9	0.92	0.88	0.42	0.43		93.95	92.69	7.06	7.46	3.59	3.63
6		99.92	99.92	0.8	0.78	0.37	0.39		93.52	93.5	7.31	7.03	3.89	3.71
7		99.89	99.89	0.93	0.93	0.41	0.42		93.19	94.38	7.46	6.85	3.9	3.61
8		99.87	99.87	1.05	1.03	0.57	0.57		93.86	93.15	7.08	7.51	3.68	3.92
9		99.9	99.9	0.88	0.89	0.42	0.43		93.95	93.38	7.03	7.46	3.48	3.69
10		99.89	99.9	0.94	0.89	0.44	0.44		93.63	93.76	7.2	7.33	3.51	3.69
Mean		99.89	99.88	0.97	0.99	0.49	0.49		93.64	93.56	7.22	7.26	3.68	3.7
SD		0.02	0.05	0.1	0.19	0.09	0.09		0.28	0.59	0.15	0.33	0.14	0.11
1	6 hours	99.57	99.49	1.92	2.08	0.85	0.92	6 hours	93.82	93.39	7.12	7.25	3.98	4.13
2		99.66	99.63	1.71	1.77	0.67	0.69		94.09	93.3	6.96	7.33	3.59	3.7
3		99.55	99.46	1.96	2.14	0.84	0.87		93.51	94.46	7.24	7.11	3.77	3.81
4		99.65	99.61	1.73	1.83	0.69	0.72		93.99	93.93	7.01	7.11	3.53	3.64
5		99.62	99.57	1.79	1.91	0.78	0.82		93.98	93.43	7.06	6.94	3.61	3.43
6		99.52	99.48	2.02	2.1	0.91	0.95		93.49	94.11	7.3	6.99	3.86	3.65
7		99.59	99.57	1.87	1.9	0.82	0.81		93.2	93.02	7.46	7.53	3.92	3.81
8		99.55	99.58	1.95	1.88	0.92	0.87		93.96	92.69	7.05	7.59	3.48	3.74
9		99.62	99.63	1.79	1.78	0.68	0.67		93.68	92.77	7.19	7.71	3.56	3.86
10		99.66	99.6	1.69	1.84	0.65	0.65		93.54	94.15	7.26	7.06	3.88	3.88
mean		99.6	99.57	1.85	1.93	0.79	0.8		93.73	93.53	7.17	7.27	3.72	3.77
SD		0.05	0.06	0.11	0.13	0.1	0.1		0.27	0.58	0.15	0.26	0.17	0.17
1	12 hours	98.89	98.66	3.08	3.38	1.11	1.19	24 hours	83.86	83.56	11.52	11.34	6.37	6.09
2		98.91	98.93	3.06	3.01	1.14	1.12		84.7	82.56	11.17	12.16	6.15	6.92
3		98.84	98.53	3.16	3.55	1.25	1.37		85.6	80.05	10.9	12.27	5.83	6.29
4		98.79	98.9	3.21	3.1	1.11	1.07		83.95	83.11	11.46	11.82	6.01	6.07
5		98.96	98.74	2.99	3.33	1.07	1.17		84.05	83.27	11.38	12.09	6.33	6.73
6		98.79	98.76	3.21	3.29	1.33	1.37		84.88	83.27	11.11	11.51	6.06	6.3
7		98.93	98.68	3.03	3.35	1.09	1.13		81.35	82.96	12.45	10.97	7.21	6.4
8		98.85	98.73	3.14	3.28	1.14	1.18		84.64	82.28	11.21	12.1	6.05	6.61
9		98.7	98.54	3.34	3.5	1.27	1.28		85.01	85.29	11.05	11.24	5.74	5.82
10		98.9	98.73	3.07	3.28	1.12	1.16		85	84.95	11.06	11.39	5.64	5.98
mean		98.86	98.73	3.13	3.31	1.17	1.21		84.31	83.23	11.34	11.69	6.14	6.33
SD		0.07	0.12	0.1	0.15	0.08	0.1		1.11	1.41	0.42	0.43	0.42	0.33

horizon exceeds the six-hour resolution. Table 10quantifies the forecasting performance on $\mathfrak{D}_1$ and $\mathfrak{D}_2$. For the sake of the integrity of the forecasting procedure for multiple steps ahead, ten-fold cross-validation procedure is adopted.

According to Table 10, the performance of NARX-LSTM decreases with the increase of the forecasting horizon. A clear conclusion can be drawn that the single-step predictions are more accurate than multiple steps predictions. However, from the simulation results on hourly daily ahead predictions, it is concretely evident that the proposed NARX-LSTM could generate accurate results with an $R^2 = 99.89\% \pm 0.02$ and $R^2 = 99.69\% \pm 0.19$ for $\mathfrak{D}_1$ and $\mathfrak{D}_2$ respectively. While the worse results are achieved for 12-hours and 24-hours ahead with $R^2 = 98.86 \pm 0.07$ and $R^2 = 84.31\% \pm 1.11$ for $\mathfrak{D}_1$ and $\mathfrak{D}_2$ respectively. As reported in Table 10, it can be said that the forecasting skill of the proposed NARX-LSTM dramatically decreases after 6 hours ahead forecasting threshold. Therefore, the validity of the proposed model on 24-ahead forecasting is not considered in what follows. For the sake of comparing the model forecasting performance with benchmarks based on 3 hours and 6 hours ahead forecasting, several benchmarks were employed including Wavelet Transform-Generalized Regression Neural

TABLE 11. Comparison of 1-hour ahead daily forecasts. The best information criterion is in boldface.

Model	**nRMSE (%)**
ENS [49]	7.16
Improved ANN [50]	3.76
GB [49]	7.36
Correlation model [51]	5.68
ELM [51]	10.14
Modified LSTM [25]	8.83
D-PNN [52]	13.42
AE-LSTM [23]	8.39
WT-RBFNN-PSO [31]	1.85
NARX-LSTM	**1.33**



(a) 1 hour ahead forecasts using $\mathfrak{D}_1$ (b) 6 hours ahead forecasts using $\mathfrak{D}_1$

(c) 12 hours ahead forecasts using $\mathfrak{D}_1$ (d) 3 hours ahead forecasts using $\mathfrak{D}_2$

(e) 6 hours ahead forecasts using $\mathfrak{D}_2$ (f) 24 hours ahead forecasts using $\mathfrak{D}_2$

FIGURE 11. Multistep forecasting tests of NARX-LSTM model using $\mathfrak{D}_1$ and $\mathfrak{D}_2$ datasets for different forecasting horizons.

TABLE 12. Comparative study of 3-hour ahead daily PVPF.

Model	nRMSE (%)	nMAE (%)
3-hour ahead PVPF results.		
WT-GRNN [31]	12.88	7.65
WT-GRNN-PSO [31]	9.89	7.38
WT-RBFNN [31]	12.75	8.12
WT-RBFNN-PSO [31]	**6.85**	5.44
NARX-LSTM ($\mathfrak{D}_2$)	7.26	**3.70**
6-hour ahead PVPF results.		
WT-GRNN [31]	12.38	9.78
WT-GRNN-PSO [31]	12.39	10.02
WT-RBFNN [31]	12.77	10.29
WT-RBFNN-PSO [31]	10.95	9.03
NARX-LSTM ($\mathfrak{D}_1$)	**1.93**	**0.80**
NARX-LSTM ($\mathfrak{D}_2$)	7.27	3.77

Note: The best information criterion is in boldface.

Network (WT-GRNN), Wavelet Transform-Generalized Regression Neural Network-Particle Swarm Optimization (WT-GRNN-PSO), Wavelet transform-Radial Basis function neural network (WT-RBFNN), and WT-RBFNN-PSO models [31]. Table 12 resumes the forecasting performance of NARX-LSTM ($\mathfrak{D}_2$) with the cited benchmarks based on 3-hour ahead and 6-hour ahead multistep forecasting.

Table 12 presents the performance comparison of the proposed model with other benchmarks. In terms of nMAE, the proposed model is favorable with nMAE = 3.70%. while WT-RBFNN-PSO(nRMSE = 6.85%) is outperforming the NARX-LSTM (nRMSE = 7.26%) in terms of lower

TABLE 13. Performance evaluation of interval forecasting.

Model	PINC (%)	PICP (%)	ACE (%)	PINAW (%)
ICNN-QR [55]	90	90.035	0.035	0.047
CNN-QR [55]	90	90.064	0.064	0.079
BELM [54]	90	93.920	3.920	-
NN-QR [55]	90	90.068	0.068	0.081
Persistance [54]	90	94.920	4.920	-
PSOWA [53]	90	93.120	3.120	18.840
IDEWA [53]	90	96.170	6.170	17.680
NARX-LSTM ($\mathfrak{D}_1$)	90	90.024	0.024	0.030
NARX-LSTM ($\mathfrak{D}_2$)	90	90.018	0.018	0.026

nRMSE values. On the other side, according to Table 12, the performance of the proposed model is best with the dominance of NARX-LSTM on $\mathfrak{D}_1$ data set. The NARX-LSTM generates an nRMSE = 1.93 % and nMAE = 0.80% on $\mathfrak{D}_1$, while the reported performance errors on $\mathfrak{D}_2$ are equal to nRMSE = 7.27% and nMAE = 3.77% for 6-hour daily PVPF. The proposed model is found highly accurate with good competitiveness skills for a single-step and multistep point forecasting. Nevertheless, a longer time horizon requires further investigations.

D. INTERVAL FORECASTING OF PV POWER

Interval forecasting is essentially conducted to quantify the uncertainties of the PVPF associated with point forecasting. Interval forecasting is a challenging task due to the stochastic variation of the weather parameters during the days and seasons of the year. The unavoidability of point forecasting errors has significantly intensified the inherent role of interval forecasting. In this paper, prediction intervals are used for quantitative characterizations of PV power forecasts. The loss function ℓ of single-point forecasts in quantile regression is defined as:

$$\ell(\xi|\alpha) = \begin{cases} \alpha\xi_i & \xi \geqslant 0 \\ (\alpha-1)\xi_i, & \xi < 0 \end{cases} \tag{36}$$

where ξ is the quantile value quantified between 0 and 1 and ξ is calculated as:

$$\xi = y_i - f(x_i) \tag{37}$$

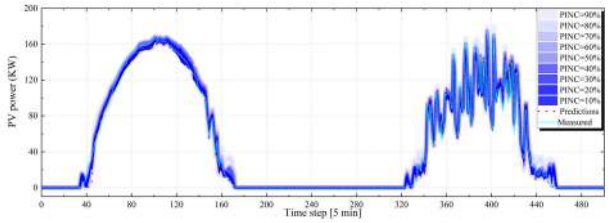
Here $f(x)$ is the forecasted quantile model. The average function for a given data set is computed as:

$$\ell(y, |\alpha) = \frac{1}{N}\sum_{i=1}^{N} \ell(y_i - f(x_i)|\alpha) \tag{38}$$

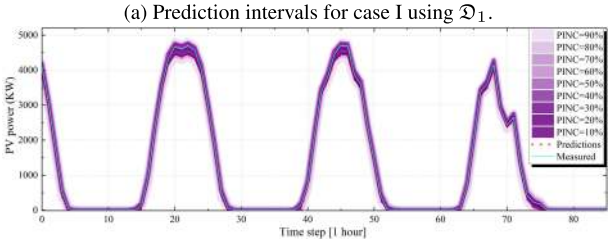
After adding a risk uncertainty-based LSTM model, the forecasting system becomes able to perform quantile regression. The interval forecasting model is implemented in Keras by the modification of the loss function as Eq. 38 to provide

TABLE 14. Results of multi-step interval forecasting for 1-hour, 2-hours, and 3-hours forecasting horizons.

Time(hh:mm)	Actual value(kW)	1-hour ahead	2-hours ahead	3-hours ahead
		$\mathfrak{D}_1$		
07:00	5.224	[3.733,9.727]	[3.406,11.228]	[5.643,17.995]
08:00	8.101	[6.879,11.327]	[4.578,13.646]	[6.229,17.796]
09:00	10.328	[9.574,14.003]	[5.814,16.221]	[7.602,17.533]
		...		
13:00	83.072	[75.757,93.114]	[70.670,88.299]	[72.931,82.184]
14:00	87.347	[80.687,93.388]	[74.637,91.911]	[79.381,86.549]
15:00	91.136	[85.214,96.983]	[82.648,97.055]	[84.173,91.728]
		$\mathfrak{D}_2$		
08:00	57.88	[13.798,334.950]	[0.002,415.508]	[13.798,334.950]
09:00	593.25	[362.070,891.557]	[228.263,1089.999]	[362.070,891.557]
10:00	1214	[920.289,1936.505]	[945.482,2044.318]	[920.289,1936.505]
		...		
17:00	1193.5	[398.825,1204.242]	[317.962,1568.137]	[865.867,1175.649]
18:00	311.5	[285.082,1072.712]	[626.824,1146.699]	[409.353,995.631]
19:00	17.34	[16.044,151.493]	[0.085,100.161]	[2.045,163.101]

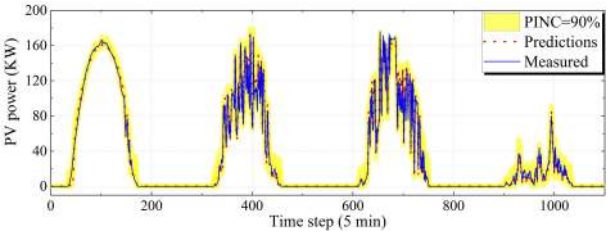


(a) Prediction intervals for case I using $\mathfrak{D}_1$.

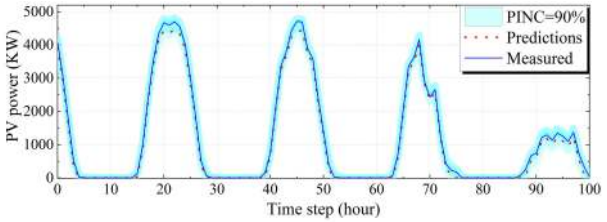


(b) Prediction intervals for case II using $\mathfrak{D}_2$.

FIGURE 12. Prediction interval results for $\mathfrak{D}_1$ and $\mathfrak{D}_2$.



(a) NARX-LSTM at 5 min ahead with PINC=90% on $\mathfrak{D}_1$.



(b) NARX-LSTM at 1-hour ahead with PINC=90% on $\mathfrak{D}_2$

FIGURE 13. 90% Prediction interval results on $\mathfrak{D}_1$ and $\mathfrak{D}_2$.

uncertainties from point PVPF. The experimental results are conducted for two data sets to validate the generalization of the forecasting model. The first case study is conducted for 5 min daily forecasting while the second case study is for hourly daily forecasting. The simulation results on $\mathfrak{D}_1$ and $\mathfrak{D}_2$ are illustrated in Fig. 12.

Regarding Fig. 12, it has been found that for case I-II that the prediction intervals encloses most of the actual values. The interval forecasting yields accurate uncertainty results for 5 min and 1 hour ahead predictions. In Fig. 13, 90% prediction intervals are illustrated where they enclose the actual PV power for different weather conditions.

In order to compare the results of interval forecasting for one step ahead, several recent benchmarks are employed including Particle Swarm Optimization-based Weighted Average (PSOWA), Convolutional Neural Network-Quantile Regression(CNN-QR), Improved Convolutional Neural Network-Quantile Regression (ICNN-QR), Neural Network-Quantile Regression NN-QR, Bagging Extreme Learning Machine (BELM), Persistance model, and Improved Differential Evolution-based Weighted Average (IDEWA) [53]–[55]. The interval error measures for the proposed NARX-LSTM model with the other models are resumed in Table 13.

The error measures for $\mathfrak{D}_1$ and $\mathfrak{D}_2$ are computed in Table 13. Particularly, for PINC = 90%, the NARX-LSTM generates a PICP = 90.024% for 5 min ahead of daily forecasting and PICP = 90.018% for hourly daily forecasting using $\mathfrak{D}_2$ data set. The calculated Pinball loss is equal to 0.16% and 0.12% for $\mathfrak{D}_1$ and $\mathfrak{D}_2$ respectively. Regarding the error measures, a general conclusion can be drawn that the interval forecasting with NARX-LSTM generates a higher performance in terms of lower ACE and PINAW compared to the state-of-the-art. In this paper, the uncertainty analysis takes into account the multistep forecasting validation. The proposed NARX-LSTM model is examined for 1 hour, 2-hours, and 3-hours ahead horizon. The results of multistep validation for interval forecasting are reported in Table 14.

From Table 14, it can be mentioned that the proposed NARX-LSTM model is strongly efficient in interval forecasting where all the true values are enclosed by the generated prediction intervals especially in 1 hour ahead and 2-hour ahead. The expansion of the forecasting horizon to 3 hours ahead leads to losing sight of accuracy in terms that some of the true values exceed the forecasting intervals. For example, at 17:00, from the second case study, the actual PV power is 1193.5kW, while the prediction interval is [865.867,1175.649]. However, the increase of the forecasting horizon decreases the forecasting accuracy. In order to overcome this horizon limitation, the data set can be extended to more related features such as the lucrative inclusion of the zenith and azimuth. By enriching the data representation, the proposed NARX-LSTM can employ the additional information to enhance its prediction accuracy. Furthermore, the computational burden caused by the multiple processing units can limit the potential of the NARX-LSTM model, which requires further investigations. As a conclusion, the NARX-LSTM model proves its suitability for single step, 1-hour, and 2-hour ahead interval forecasting.

IV. CONCLUSION

This paper proposed a new computing framework based on the combination of Nonlinear Auto-Regressive Exogenous Neural Network (NARXNN) and Long Short-Term Memory (LSTM) optimized by Tabu search algorithm. It was demonstrated that the proposed model is strongly able to capture the behavior of weather changes as well as the uncertainties associated with point forecasts to generate accurate short-term PV power forecasting. Furthermore, the performance of the hybrid model was investigated for different horizon scales in PV power forecasting. The presented simulation results have shown that the NARX-LSTM model impressively offers higher efficiency and accuracy compared to the commonly deployed methods in PV power forecasting problems. Moreover, the NARX-LSTM model showed excellent performance and good generalization capabilities from two different locations and utility scales. The proposed model contributes to smart grids in terms of efficient unit commitment and reliable budget planning with a high certitude. Future work will broaden the scope to include a feasibility investigation of the proposed model for other smart grid applications including wind power and load forecasting.

REFERENCES

[1] O. Ellabban, H. Abu-Rub, and F. Blaabjerg, "Renewable energy resources: Current status, future prospects and their enabling technology," *Renew. Sustain. Energy Rev.*, vol. 39, pp. 748–764, Nov. 2014.
[2] M. Massaoudi, I. Chihi, L. Sidhom, M. Trabelsi, and F. S. Oueslati, "Medium and long-term parametric temperature forecasting using real meteorological data," in *Proc. 45th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, vol. 1, Oct. 2019, pp. 2402–2407.
[3] M. Massaoudi, S. S. Refaat, H. Abu-Rub, I. Chihi, and F. S. Wesleti, "A hybrid Bayesian ridge regression-CWT-catboost model for PV power forecasting," in *Proc. IEEE Kansas Power Energy Conf. (KPEC)*, Jul. 2020, pp. 1–5.
[4] M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati, and H. Abu-Rub, "A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for short-term load forecasting," *Energy*, vol. 214, Jan. 2021, Art. no. 118874. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544220319812
[5] Y. Jung, J. Jung, B. Kim, and S. 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," *J. Cleaner Prod.*, vol. 250, Mar. 2020, Art. no. 119476.
[6] 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.
[7] M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, H. Abu-Rub, and F. S. Oueslati, "Short-term electric load forecasting based on data-driven deep learning techniques," in *Proc. 46th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2020, pp. 2565–2570.
[8] M. Massaoudi, I. Chihi, L. Sidhom, M. Trabelsi, S. S. Refaat, and F. S. Oueslati, "Performance evaluation of deep recurrent neural networks architectures: Application to PV power forecasting," in *Proc. 2nd Int. Conf. Smart Grid Renew. Energy (SGRE)*, Nov. 2019, pp. 1–6.
[9] Z. Lei and Y. Yang, "Research on data mining algorithm for regional photovoltaic generation," in *Proc. 3rd EAI Int. Conf. Adv. Hybrid Inf. Process.*, Nanjing, China, vol. 301. Cham, Switzerland: Springer, 2019, pp. 429–438.
[10] M. N. Akhter, S. Mekhilef, H. Mokhlis, and N. Mohamed Shah, "Review on forecasting of photovoltaic power generation based on machine learning and Metaheuristic techniques," *IET Renew. Power Gener.*, vol. 13, no. 7, pp. 1009–1023, May 2019.
[11] G. Cervone, L. Clemente-Harding, S. Alessandrini, and L. Delle Monache, "Short-term photovoltaic power forecasting using artificial neural networks and an analog ensemble," *Renew. Energy*, vol. 108, pp. 274–286, Aug. 2017.
[12] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif, "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization," *Renew. Sustain. Energy Rev.*, vol. 124, May 2020, Art. no. 109792.
[13] D. W. van der Meer, J. Widén, and J. Munkhammar, "Review on probabilistic forecasting of photovoltaic power production and electricity consumption," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 1484–1512, Jan. 2018.
[14] M. Yang and X. Huang, "Ultra-short-term prediction of photovoltaic power based on periodic extraction of PV energy and LSH algorithm," *IEEE Access*, vol. 6, pp. 51200–51205, 2018.
[15] H. Sheng, J. Xiao, Y. Cheng, Q. Ni, and S. Wang, "Short-term solar power forecasting based on weighted Gaussian process regression," *IEEE Trans. Ind. Electron.*, vol. 65, no. 1, pp. 300–308, Jan. 2018.
[16] X. Zhang, Y. Li, S. Lu, H. F. Hamann, B.-M. Hodge, and B. Lehman, "A solar time based analog ensemble method for regional solar power forecasting," *IEEE Trans. Sustain. Energy*, vol. 10, no. 1, pp. 268–279, Jan. 2019.
[17] C.-J. Huang and P.-H. Kuo, "Multiple-input deep convolutional neural network model for short-term photovoltaic power forecasting," *IEEE Access*, vol. 7, pp. 74822–74834, 2019.
[18] X. Zhang, F. Fang, and J. Liu, "Weather-classification-MARS-based photovoltaic power forecasting for energy imbalance market," *IEEE Trans. Ind. Electron.*, vol. 66, no. 11, pp. 8692–8702, Nov. 2019.

[19] S. Han, Y.-H. Qiao, J. Yan, Y.-Q. 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.

[20] M. K. Behera and N. Nayak, "A comparative study on short-term PV power forecasting using decomposition based optimized extreme learning machine algorithm," *Eng. Sci. Technol., Int. J.*, vol. 23, no. 1, pp. 156–167, Feb. 2020.

[21] Y. Ju, J. Li, and G. Sun, "Ultra-short-term photovoltaic power prediction based on self-attention mechanism and multi-task learning," *IEEE Access*, vol. 8, pp. 44821–44829, 2020.

[22] H. Wen, Y. Du, X. Chen, E. Lim, H. Wen, L. Jiang, and W. Xiang, "Deep learning based multistep solar forecasting for PV ramp-rate control using sky images," *IEEE Trans. Ind. Informat.*, vol. 17, no. 2, pp. 1397–1406, Feb. 2021.

[23] Y. Zhang, C. Qin, A. K. Srivastava, C. Jin, and R. K. Sharma, "Data-driven day-ahead PV estimation using autoencoder-LSTM and persistence model," *IEEE Trans. Ind. Appl.*, vol. 56, no. 6, pp. 7185–7192, Nov./Dec. 2020.

[24] M. Pan, C. Li, R. Gao, Y. Huang, H. You, T. Gu, and F. Qin, "Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization," *J. Cleaner Prod.*, vol. 277, Dec. 2020, Art. no. 123948.

[25] F. Wang, Z. Xuan, Z. Zhen, K. Li, T. Wang, and M. Shi, "A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework," *Energy Convers. Manage.*, vol. 212, May 2020, Art. no. 112766.

[26] H. Zang, L. Cheng, T. Ding, K. W. Cheung, Z. Wei, and G. Sun, "Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning," *Int. J. Electr. Power Energy Syst.*, vol. 118, Jun. 2020, Art. no. 105790.

[27] S. Balluff, J. Bendfeld, and S. Krauter, "Meteorological data forecast using RNN," *Int. J. Grid High Perform. Comput.*, vol. 9, no. 1, pp. 61–74, Jan. 2017.

[28] 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.

[29] G. Marcjasz, B. Uniejewski, and R. Weron, "On the importance of the long-term seasonal component in day-ahead electricity price forecasting with NARX neural networks," *Int. J. Forecasting*, vol. 35, no. 4, pp. 1520–1532, Oct. 2019, doi: 10.1016/j.ijforecast.2017.11.009.

[30] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. W. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 2627–2633.

[31] Y. Wen, D. AlHakeem, P. Mandal, S. Chakraborty, Y.-K. Wu, T. Senjyu, S. Paudyal, and T.-L. Tseng, "Performance evaluation of probabilistic methods based on bootstrap and quantile regression to quantify PV power point forecast uncertainty," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 4, pp. 1134–1144, Apr. 2020.

[32] M. Sun, C. Feng, and J. Zhang, "Probabilistic solar power forecasting based on weather scenario generation," *Appl. Energy*, vol. 266, May 2020, Art. no. 114823.

[33] H. Wang, H. Yi, J. Peng, G. Wang, Y. Liu, H. Jiang, and W. Liu, "Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network," *Energy Convers. Manage.*, vol. 153, pp. 409–422, Dec. 2017.

[34] E. Rangel, E. Cadenas, R. Campos-Amezcua, and J. L. Tena, "Enhanced prediction of solar radiation using NARX models with corrected input vectors," *Energies*, vol. 13, no. 10, p. 2576, 2020.

[35] T.-N. Lin, C. L. Giles, B. G. Horne, and S.-Y. Kung, "A delay damage model selection algorithm for NARX neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2719–2730, 1997.

[36] S. A. Khan, V. Thakore, A. Behal, L. Bölöni, and J. J. Hickman, "Comparative analysis of system identification techniques for nonlinear modeling of the neuron–microelectrode junction," *J. Comput. Theor. Nanosci.*, vol. 10, no. 3, pp. 573–580, 2013.

[37] S. J. Abdulkadir and S.-P. Yong, "Empirical analysis of parallel-NARX recurrent network for long-term chaotic financial forecasting," in *Proc. Int. Conf. Comput. Inf. Sci. (ICCOINS)*, Jun. 2014, pp. 1–5.

[38] W. VanDeventer, E. Jamei, G. S. Thirunavukkarasu, M. Seyedmahmoudian, T. K. Soon, B. Horan, S. Mekhilef, and A. Stojcevski, "Short-term PV power forecasting using hybrid GASVM technique," *Renew. Energy*, vol. 140, pp. 367–379, Sep. 2019.

[39] A. M. Mohammed and S. O. Duffuaa, "A tabu search based algorithm for the optimal design of multi-objective multi-product supply chain networks," *Expert Syst. Appl.*, vol. 140, Feb. 2020, Art. no. 112808.

[40] N. Srinidhi, T. Datta, A. Chockalingam, and B. S. Rajan, "Layered tabu search algorithm for large-MIMO detection and a lower bound on ML performance," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 2955–2963, Nov. 2011.

[41] H.-J. Glaeske and V. K. Tuan, "Some applications of the convolution theorem of the Hilbert transform," *Integral Transforms Special Functions*, vol. 3, no. 4, pp. 263–268, Dec. 1995.

[42] M. Massaoudi, S. S. Refaat, H. Abu-Rub, I. Chihi, and F. S. Oueslati, "PLS-CNN-BiLSTM: An end-to-end algorithm-based Savitzky–Golay smoothing and evolution strategy for load forecasting," *Energies*, vol. 13, no. 20, p. 5464, 2020. [Online]. Available: https://www.mdpi.com/1996-1073/13/20/5464

[43] *Data Download | DKA Solar Centre*. Accessed: Sep. 23, 2019. [Online]. Available: http://dkasolarcentre.com.au/locations/alice-springs

[44] G. C. Adele Kuzmiakova and A. McKeehan. *Machine-Learning-for-Solar-Energy-Prediction*. Accessed: Aug. 20, 2020. [Online]. Available: https://github.com/ColasGael/Machine-Learning-for-Solar-Energy-Prediction

[45] M. Sun, C. Feng, and J. Zhang, "Conditional aggregated probabilistic wind power forecasting based on spatio-temporal correlation," *Appl. Energy*, vol. 256, Dec. 2019, Art. no. 113842.

[46] F. Chollet. *Keras-Team/Keras: Deep Learning for Humans*. Accessed: Aug. 20, 2020. [Online]. Available: https://github.com/keras-team/keras

[47] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, and A. Mueller, "Scikit-learn," *GetMobile: Mobile Comput. Commun.*, vol. 19, no. 1, pp. 29–33, 2015.

[48] S. Blanke. (2019). *Hyperactive: A Hyperparameter Optimization and Meta-Learning Toolbox for Machine-/Deep-Learning Models*. [Online]. Available: https://github.com/SimonBlanke

[49] L. Gigoni, A. Betti, E. Crisostomi, A. Franco, M. Tucci, F. Bizzarri, and D. Mucci, "Day-ahead hourly forecasting of power generation from photovoltaic plants," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 831–842, Apr. 2018.

[50] S. Theocharides, G. Makrides, A. Livera, M. Theristis, P. Kaimakis, and G. E. Georghiou, "Day-ahead photovoltaic power production forecasting methodology based on machine learning and statistical post-processing," *Appl. Energy*, vol. 268, Jun. 2020, Art. no. 115023.

[51] W. Yin, Y. Han, H. Zhou, M. Ma, L. Li, and H. Zhu, "A novel non-iterative correction method for short-term photovoltaic power forecasting," *Renew. Energy*, vol. 159, pp. 23–32, Oct. 2020.

[52] L. Zjavka, "Photo-voltaic power daily predictions using expanding PDE sum models of polynomial networks based on operational calculus," *Eng. Appl. Artif. Intell.*, vol. 89, Mar. 2020, Art. no. 103409.

[53] Q. Ni, S. Zhuang, H. Sheng, G. Kang, and J. Xiao, "An ensemble prediction intervals approach for short-term PV power forecasting," *Sol. Energy*, vol. 155, pp. 1072–1083, Oct. 2017.

[54] W. Liu and Y. Xu, "Randomised learning-based hybrid ensemble model for probabilistic forecasting of PV power generation," *IET Gener., Transmiss. Distrib.*, vol. 14, no. 24, pp. 5909–5917, 2020.

[55] Y. Yu, M. Wang, F. Yan, M. Yang, and J. Yang, "Improved convolutional neural network-based quantile regression for regional photovoltaic generation probabilistic forecast," *IET Renew. Power Gener.*, vol. 14, no. 14, pp. 2712–2719, 2020.

MOHAMED MASSAOUDI (Student Member, IEEE) received the degree in energy engineering from the National Engineering School of Monastir, ENIM, Tunisia, in 2018. He is currently pursuing the Ph.D. degree with INSAT, Tunisia, and also a student worker at the Texas A&M University of Qatar. His research interests include machine learning and deep learning techniques for energy management and Big data analytics in smart grid systems. He has served as a Reviewer for some conferences and journals, including the IEEE Transactions on Industrial Informatics and the IEEE Transactions on Power Systems.

INES CHIHI received the Ph.D. degree and University Habilitation from the National Engineering School of Tunis, Tunisia (ENIT), in 2013 and 2019, respectively. She is currently an Associate Professor on Automation and Industrial Computing with the National Engineering School of Bizerta, Tunisia (ENIB), and a member of the Laboratory of Energy Applications and Renewable Energy Efficiency (LAPER), University Tunis El Manar. Her research area is based on intelligent modeling, control, and fault detection, of complex systems with unpredictable behaviors.

LILIA SIDHOM received the Ph.D. degree in automation and industrial computing from the National Institute of Applied Sciences of Lyon (INSA Lyon), Villeurbanne, France, in 2011, and the University Habilitation degree from INSAT, Tunisia, in 2019. She is currently an Associate Professor on Automation and Industrial Computing with the Department of Mechanical Engineering, National Engineering School of Bizerta, Tunisia (ENIB), and a member of the Laboratory of Energy Applications and Renewable Energy Efficiency (LAPER), University Tunis El Manar. She received the automation and industrial computing engineering degree from the National Institute of Applied Sciences and Technology (INSAT), Tunis, Tunisia, in 2007. Her current research interests include advanced control of nonlinear complex and innovative system, soft sensors, smart diagnostic, observer/estimator. Her research applications are concerning smart grid, biomedical, drill string systems, and robots.

MOHAMED TRABELSI (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from INSAT, Tunisia, in 2006, and the M.Sc. degree in automated systems and the Ph.D. degree in energy systems from INSA Lyon, France, in 2006 and 2009, respectively. From October 2009 to August 2018, he has been holding different Research positions at Qatar University and Texas A&M University at Qatar. Since September 2018, he has been with the Kuwait College of Science and Technology as an Associate Professor. His research interests include systems control with applications arising in the contexts of power electronics, energy conversion, renewable energies integration, and smart grids. He has published more than 100 journals and conference papers, and he is an author of two books and two book chapters.

SHADY S. REFAAT (Senior Member, IEEE) received the B.A.Sc., M.A.Sc., and Ph.D. degrees in electrical engineering from Cairo University, Giza, Egypt, in 2002, 2007, and 2013, respectively. He has worked in the industry for more than 12 years, as an Engineering Team Leader, a Senior Electrical Engineer, and an Electrical Design Engineer, on various electrical engineering projects. He is currently an Assistant Research Scientist with the Department of Electrical and Computer Engineering, Texas A&M University at Qatar. He has published more than 95 journals and conference papers. His principal work area focuses on electrical machines, power systems, smart grid, big data, energy management systems, reliability of power grids and electric machinery, fault detection, and condition monitoring and development of fault-tolerant systems. He has also participated in and led several scientific projects over the last eight years. He has successfully realized many potential research projects. He is a member of The Institution of Engineering and Technology (IET) and the Smart Grid Center—Extension in Qatar (SGC-Q).

HAITHAM ABU-RUB (Fellow, IEEE) received two Ph.D. degrees.

He has worked at many universities in many countries, including Poland, Palestine, USA, Germany, and Qatar. Since 2006, he has been with Texas A&M University at Qatar, where he has served as the Chair of the Electrical and Computer Engineering Program for five years and has been currently serving as the Managing Director of Smart Grid Center. His main research interests include electric drives, power electronic converters, renewable energy, and smart grid. He is a recipient of many national and international awards and recognitions. He is a recipient of the American Fulbright Scholarship, the German Alexander von Humboldt Fellowship, and many others. He has published more than 400 journals and conference papers, five books, and six book chapters. He has supervised many research projects on smart grid, power electronics converters, and renewable energy systems.

FAKHREDDINE S. OUESLATI received the Ph.D. degree and Habilitation in Physics to direct research from the Faculty of Sciences of Tunis, University of Tunis El Manar, Tunisia, in 2002 and 2012, respectively. He was a Civil Engineer with the National Engineering School of Tunis (ENIT), University of Tunis El Manar, in 1990. He worked as a Project Engineer at AFH (Tunis), from 1990 to 2002. He worked as an Assistant Professor (2002–2013) at ISSAT, Higher Institute of Applied Sciences and Technology of Gabes, INRST, National Institute of Scientific and Technical Research, Borj Cedria, University of Carthage, Tunisia, and ESTI, Higher School of Technology and Computer, University of Carthage, Tunis, Tunisia. He was an Associate Professor (2013–2018) and a Professor (since 2018) at the National Engineering School of Carthage (ENICarthage), University of Carthage. His research area covers several areas, including materials science, energy systems, pollution, and renewable energies with expertise in multi-component multi-phase convection-diffusion problems. He is currently responsible for the axis of the flow and transport phenomena in the Laboratory of Energy and Heat and Mass Transfer (LETTM), Faculty of Sciences of Tunis (FST), Tunisia. He has several international publications and is a member of several boards of directors/scientists and conference organizing committees. He has been a visiting professor in several universities and international prestigious organizations. He has chaired many international conferences; has been a regular reviewer for journals and international research projects; and an Invited Professor for many institutes.