Energy Conversion and Management: X 22 (2024) 100535

Contents lists available at ScienceDirect

Energy Conversion and Management: X

journal homepage: www.sciencedirect.com/journal/energy-conversion-and-management-x

A comprehensive framework for effective long-short term solar yield forecasting

Biplob Ray [a], Dimuth Lasantha [b], Vijayalaxmi Beeravalli [b], Adnan Anwar [c], Md Nurun Nabi [a,*], Hanmin Sheng [d], Fazlur Rashid [e,f], S.M. Muyeen [g]

[a] *Central Queensland University, Melbourne, Australia*
[b] *Central Queensland University, Rockhampton, Australia*
[c] *Centre for Cyber Resilience and Trust (CREST), School of IT, Deakin University, Australia*
[d] *School of Automation Engineering, University of Science and Technology, Chengdu, China*
[e] *Rajshahi University of Engineering & Technology, Bangladesh*
[f] *Missouri University of Science and Technology, Rolla, MO 65401, USA*
[g] *Qatar University, Doha, Qatar*

ARTICLE INFO

Keywords:
Prediction
Forecasting
Photovoltaic
Machine Learning
Framework

ABSTRACT

Due to the variability of Photovoltaic (PV) output, a forecasting framework is essential for grid connected PV plants to ensure a stable and uninterrupted power supply. Among existing prediction and forecasting algorithms, only some have attempted to provide a holistic framework for short and long-term forecasting of PV yield together using automated input feature selections and data cleaning features. Furthermore, it has been identified that many existing algorithms only predicted PV output instead of forecasting in future times; therefore, their reported accuracy needs to be upheld in forecasting scenarios. This paper has proposed a framework to streamline solar yield forecasting for both the short and long term to ensure effective integration of PV plant output with the main grid. The proposed framework has used a novel combination of XGBoost (eXtreme Gradient Boosting), time series seasonal decomposition and rolling LSTM (Long- and Short-Term Memory) model to address the need for a comprehensive forecasting framework in hourly, daily and yearly periods. Based on our experiment result, the developed framework has performed in 98% – 95% prediction accuracy with less than 0.15% normalized Root Mean Squire error (nRMSE). The framework has performed in 89%- 87% forecasting accuracy with less than 0.45% nRMSE. Both the prediction and forecasting performance of the proposed model have outperformed many benchmarks forecasting frameworks, including Long short-term memory (LSTM) based recurrent neural network (RNN), Full RNN (FRNN), Neural Network Ensemble (NNE), Neural Network with AdaBoost, and many more as detailed in our comparative study section.

Introduction

Renewable energy opened the opportunity to create sustainable energy sources without impacting the environment. Solar energy is one of the most promising renewable energy sources due to its reasonable production cost and environmental friendliness. However, due to the high dependency on the climatology parameters [1], the PV plants produce variable energy output to meet ever increasing demand. Some forecasting methods [1,2] exist in the literature to predict influential climatology parameters of PV plants' output. Others have investigated performance parameters and PV module improvement in electrical characteristics and thermodynamics [37,38]. The study on climatology and performance parameters and prediction help to understand correlations between environment-oriented conditions, PV output patterns, and variability patterns.

However, accurate forecasting of PV energy output can be very

Abbreviations: ANN, Artificial neural network; ARENA, Australian Renewable Energy Agency; ASEFS, Australian Solar Energy Forecasting System; CNN, Convolutional neural network; DNI, Direct normal irradiance; DHI, Diffuse horizontal irradiance; GHI, Global horizontal irradiance; LSTM, Long and short-term memory; NN, Neural network; PV, Photovoltaic; RNN, Recurrent neural network; MAPE, Mean Absolute Percentage Error; RMSE, Root Mean Squire Error; nRMSE, normalised RMSE; S-L, Short-Long; XGBoost, eXtreme Gradient Boosting; AEMO, Australian Energy Market Operator; TYM, Typical Meteorological Year; FRNN, Full RNN.

* Corresponding author.
E-mail address: m.nabi@cqu.edu.au (M. Nurun Nabi).

https://doi.org/10.1016/j.ecmx.2024.100535
Received 5 September 2023; Received in revised form 19 December 2023; Accepted 20 January 2024
Available online 24 January 2024

useful for managing and planning PV plants and connected grids. Many short-term solar yield forecasting models [3–6] exist in the literature. For example, the Australian Solar Energy Forecasting System (ASEFS) [3] was developed by the Australian Energy Market Operator (AEMO) for forecasting solar generation of 5 min to 7 days timeframes. Few other existing works [7,8] have developed long term forecasting models. But due to the complexity and high variability of solar yield, the long-term PV plants energy forecasting requires long historic data [4,7] to achieve the required accuracy. However, the longer duration of historic data, like 15–25 years, is difficult to extract, synthesize, clean and process. A review of existing models revealed that many proposed forecasting models have used the same definition for prediction and forecasting, leading to failure to adhere to the correct forecasting definition like the one adapted [9] and detailed in Section "Research Methodology". In the literature, the short-term prediction shows a great success rate of 98 %, but less detail exists about forecasting (or time step ahead prediction) success rate. While short term forecasting can be very useful for day-to-day maintenance, load balancing and decision making but long-term forecasting is required for planning strategic decisions for PV plants connected to power grids. Furthermore, the literature is lacking a comprehensive framework to streamline process from raw weather data to predict/forecast solar yield of PV stations.

Therefore, a comprehensive Short-Long (S-L) term PV energy yield forecasting framework can be helpful to understand the variability of power dispatching and market condition towards profitability. To address the gap, this paper has proposed a comprehensive framework using hybrid machine learning techniques for both Short-Long (S-L) term solar yield forecasting. The paper has made novel contributions to the field below.

- Proposed a framework to streamline both short and long-term solar yield forecasting is crucial for the integration of photovoltaic (PV) plant output with the main power grid. This approach ensures that the variable nature of solar power can be more accurately predicted and managed, facilitating a stable and reliable integration with existing power systems.
- XGBoost (eXtreme Gradient Boosting), time series seasonal decomposition, and rolling LSTM (Long- and Short-Term Memory) models to leverage the strengths of each method. This combination is particularly effective in addressing the complex and dynamic nature of solar energy generation, improving the accuracy of solar yield predictions.
- Prediction, long and short term, using a combination of Wavelet-based time series decomposition and LSTM
- Forecasting, long and short-term, using a combination of robust local mean decomposition and bidirectional (rolling) LSTM.
- Comparative study of the proposed models against existing models in the context of prediction versus forecasting. This high level of accuracy, coupled with a very low error rate, underscores the framework's effectiveness in predicting and forecasting solar yields.

The proposed hybrid forecasting framework has a holistic solar yield prediction and forecasting model for long-short duration. It is evident from the comparative study that the proposed model performs reasonably well compared to similar models in the scale of accuracy and error rate.

The rest of the paper is organized as follows: Section "Literature Review" presents a literature review followed by research methodology, proposed forecasting framework and model selection in Section "Research Methodology", Section "Proposed Framework" and Section "Forecasting model Selection" respectively. Section "Data Preparation" details the data preparation where dataset, feature selection and noise reduction are detailed in Section "Dataset", "Selection of features", and "Noise Elimination" respectively. Section "Experimental Results, Comparative Study, and Discussion" detailed experimental results in section "Experimental Results" which is followed by the comparative study and discussion in section "Comparative Study and Discussion". Finally, the closing remarks are presented in Section "Conclusion".

Literature review

Current studies on solar energy forecasting can be classified into three main categories: climatology, short term and long-term PV output forecasting. Out of the three, most of the current work investigated short term (like, hourly, daily, monthly) energy output and climatology forecasting [1,2]. Very little work investigated long term (like, yearly) forecasting of PV energy output.

Under climatology, solar irradiation was predicted in the paper [2] which has compared and tested three forecasting methods: smart persistence, artificial neural network (multilayer Perceptron) and random forest. Using datasets from three locations of France and Spain, the paper [2] has predicted three components of solar irradiation: global horizontal, beam normal and diffuse horizontal. Their comparisons among five different forecasting algorithms show that random forest performs are better than the other algorithms. The paper [2] also found that BNI and DHI are more complicated to predict than GHI. Jennifer et al. [1] presented an analysis based on the National Solar Radiation Database (NSRDB) to provide a comprehensive 15-year climatology across the United States. The analysis was done on a continental, seasonal and regional scale to understand optimal solar panels orientation. Ray et al. [10] have analyzed large-scale PV stations output data to understand the correlation among the variability of PV production at different locations in Australia. Understanding of PV output based on variation of climate condition is somewhat useful for the transmission system operator(s) for future planning of reserve margin for the system. Few others in the literature have proposed effective prediction models using machine learning and intelligent techniques for forecasting influential climatology parameters like solar irradiance for PV output [11–13].

Within existing literature related to short term forecasting, Raza et al. [4] proposed a framework for day ahead forecast of PV power output. The authors in [4] have used a Neural Network Ensemble (NNE) which is based on Swarm Optimization to get an accurate result in various complex network scenarios. The input parameters in [4] were selected using wavelet transformed historical power which is used to forecast PV outputs for three different types of days: clear, partially cloudy and cloudy which are categorized based on the clearness index. The performance of the framework proposed in [4] is tested using one-year long training set data from seven solar PV sites of the University of Queensland, Australia. The authors [4] have used one year's historic data for training and claims that proposed method outperform some existing algorithm on MAPE metric which by itself is not enough to present accuracy and performance level of the framework. As mentioned previously, AEMO [3] uses a short time forecasting model developed by Commonwealth Scientific and Industrial Research Organization (CSIRO) in 2016 which uses statistical approaches, like decision tree, random forest, etc. David et al. [14] have proposed a methodology to generate dayahead power output forecasts for two PV plants. The paper [14] has investigated the day-ahead power output (PO) from photovoltaic (PV) power plants from the National Oceanic and Atmospheric Administration, and the Canadian Meteorological Centre.

Sheng et al. [15] have proposed a day ahead forecasting model which can adapt to the climate characteristics of different regions or periods which is a new dimension towards adaptive PV forecasting. However, the proposed model [15] has a high error rate which requires improvements. Few other articles [5,6,16,17] in the literature also proposed similar day ahead as well as monthly PV output forecasting techniques. For example, Fouzi et al. [5] and Abdelkader D. et al. [5] have proposed forecasting model using deep learning with 98 % and 94 % accuracy respectively. In [5], Abdelkader D. et al. have used deep learning with Variational AutoEncoder (VAE) where Fouzi et al. in [5] have used LSTM [18]. However, both of the papers did not present data processing or feature selection details. Furthermore, the result is not validated in future time step (t + 1) as per forecasting definition in Section "Research Methodology". Ray et al. [7] have proposed a data driven yearly PV forecasting model using monthly dataset. Still, the data demanding element of the model makes it infeasible to be used by the industry.

Some researchers have used machine learning techniques to model physical problems to improve reliability of renewable systems. For example, in 2022, Wang et al. [32] proposed a multi-domain physics-informed neural network (mPINN) to address heat conduction and natural convection challenges, focusing on temperature gradient discontinuities. They utilized multiple neural networks and innovative training methods, finding that uniform residual points and joint training yielded the most accurate results. This approach proved effective in various heat transfer scenarios, showcasing mPINN's potential in solving complex physical problems. Most recently, Ghalambaz et al. [33,34] have used deep learning to understand energy storage and natural convection heat transfer mechanisms.

Understanding the performance parameters and their impacts on PV systems in a wide range of outdoor conditions is important to create a practical forecasting model. Some studies have evaluated performance parameters, measurement techniques, and their modeling [37–40] to understand the characteristics of PV systems in actual implementation outside of standard test conditions (STCs). For example, Erdem et al. have looked into the cooling effect on performance characteristics of silicon solar cells used for PV systems. They have performed numerical and experimental studies, revealing that cooling applications are important to improve the efficiency and maximum power output of the photovoltaic modules [40]. These studies are important for understanding the influence of various electrical and thermodynamic performance characteristics on PV output.

As detailed in the above literature discussion, although there exists a weekly and monthly [19] PV output prediction model using hourly datasets, to our knowledge, there is a lack of prediction models proposed in the literature that used hourly datasets for longer-term forecasting like a few years. Furthermore, deep learning-based models are proven effective for accurate prediction for datasets with variable nature [20]. However, more work must be done to present a novel framework that will take advantage of both traditional and deep learning algorithms for Short-Long (S-L) solar yield forecasting using hourly dataset. The framework must be capable of automated feature selection and data cleaning for forecasting models. To our knowledge, there is a lack of existing work which efficiently derives a novel framework to process raw PV data for both short- and long-term forecasting for diverse climate conditions.

Therefore, the proposed framework has integrated supervised learning to clean and preprocess data for forecast using statistical methods and deep learning methods for different time variation like hourly, daily, monthly and yearly.

The following section has detailed research methodology which is followed by proposed framework and model selection respectively.

Materials and methods

Research Methodology

The research methodology used in the paper is illustrated in Fig. 1 which started with Typical Meteorological Year (TYM) data from the location of the PV plants. Then, the TYM data are processed using the System Advisor Model (SAM) [21] to generate energy output and to prepare the dataset as detailed in Section "Dataset".

The hourly data extracted from SAM has many meteorological features with relevant energy output that are filled with noisy data as detailed in Section "Dataset". The proposed framework process the noisy dataset using noise reduction and feature selection techniques with a combination of XGBoost and Correlation Index (CI) as detailed in section "Selection of features" and section "Noise Elimination". This research used a novel noise reduction technique after feature selection where the feature selection process selects appropriate features for forecasting model, as detailed in section "Selection of features". In contrast, the noise reduction technique reduces the noise of selected features, so your proposed forecasting models are trained accurately with a high success rate. The proposed framework in Section "Proposed Framework" will forecast both short and long-time duration, as illustrated in Fig. 1, using an hourly time series dataset with variable patterns. Therefore, the dataset is vigorously tested with several algorithms to select and prepare appropriate forecasting models for the proposed framework, as detailed in Section "Forecasting model Selection". Finally, the performance evaluation is done to evaluate performance and to validate, as detailed in section "Experimental Results, Comparative Study, and Discussion". The performance evaluation allowed us to cross-validate and compare our proposed model with other existing forecasting models. It is worth noting that, in the literature most have used prediction and forecasting synonymously. Still, this paper has adapted prediction and foretasting Definition 1 and Definition 2 below from [9] to ensure our proposed model ensures a clear distinction between prediction and forecasting.

Definition 1. (*Prediction*) Let D be the dataset of time duration t_1-t_n with input x where $x \subset D$ and output y where $y \subset D$. The prediction model P use training set $x_{Tr}|y_{Tr}$ where $(x_{Tr} \subset x)|(y_{Tr} \subset y)$ and testing set $x_{Test}|y_{Test}$ where $(x_{Test} \subset x)|(y_{Test} \subset y)$ to train and test itself. Given available input values xt_{n+1} of time dimension t_{n+1}, the trained prediction model P_T can predict output values yt_{n+1} of same time dimension.

Remark 1. The prediction model P_T can predict output of past time domain.

Definition 2. (*Forecasting*) Let D be the dataset of time duration t_1-t_n with input x where $x \subset D$ and output y where $y \subset D$. The forecasting model F use training set $x_{Tr}|y_{Tr}$ where $(x_{Tr} \subset x)|(y_{Tr} \subset y)$ and testing set $x_{Test}|y_{Test}$ where $(x_{Test} \subset x)|(y_{Test} \subset y)$ to train and test itself. The trained forecasting model F_T can forecast output values yt_n or yt_{n+1}.

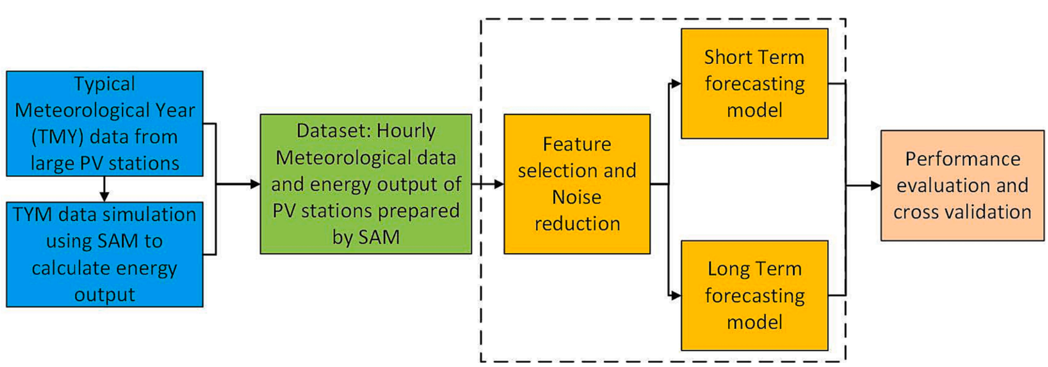

Fig. 1. Research methodology used to develop the proposed framework.

Remark 2. The forecasting model F_T can forecast output of present t_n or future time domain t_{n+1} using input values of past time dimension t_1-t_n.

Proposed framework

Grounded on the work presented in Section "Research Methodology", 4, and Section "Proposed Framework", A Short-Long (SL) forecasting framework is proposed in this paper as illustrated in Fig. 2. The proposed SL frameworks takes processed and cleaned historic data to train the developed machine learning models. The dataset consists of hourly weather features and relevant energy output as detailed in section "Selection of features". The machine learning model of the framework learns from the historic data to prepare forecasting of energy output in future time. The framework in Fig. 2 has presented forecasting in two main groups: short term forecasting, hourly and daily time duration, and long-term forecasting, yearly time duration.

The short-term forecasting is using rolling LSTM model to forecast hourly energy for a day-ahead (24 h) of time whereas the long-term forecasting is using seasonal time series decomposition model to forecast yearly energy output as illustrated in Fig. 2 and detailed in Section "Forecasting model Selection". Our experiment shows that the proposed framework can forecast with 84 %-95 % confidence which is promising considering is it forecasting in future not prediction [9]. The proposed framework can extend the forecasting in a longer time duration, like two days or two years, but the error rate increases (Root Mean Squire Error (RMSE)) and performs degrades (success rate) sharply for a longer time dimension. The experimental and validation result of this proposed framework is presented in section "Experimental Results, Comparative Study, and Discussion".

Forecasting model selection

This paper has evaluated various neural network and statistical machine learning models to perform both short- and long-term forecasting. For example, as the literature shows, combining LSTM with wavelet transform [35] is particularly effective for predicting solar radiation because it addresses climatic data's non-linear and variable nature. Wavelet transforms the time-series data into components that capture frequency and time information, which helps isolate features like seasonal patterns and anomalies. The LSTM network then leverages these decomposed components for prediction, utilizing its strengths in handling sequential data and long-term dependencies. At the same time, integrating robust local mean decomposition with a bidirectional LSTM offers a compelling approach [36] to solar irradiance forecasting. The robust local mean decomposition effectively preprocesses the data, extracting critical features and trends that might otherwise be obscured in raw time-series data.

However, based on the complexity of modeling, data variability, experimental accuracy, and objectives for developing a comprehensive Long-Short Term prediction model, this research has selected Long short-term memory (LSTM) rolling method for forecasting short duration, like hourly and daily. This short-term forecasting ensures regular maintenance operation, like balancing energy output with the grid, can be proactively planned and auctioned. As presented in Equation (1), the rolling-based LSTM adds forecasted values, like O_{n+1}, in the training set to retrain the forecasting model for forecasting the next value O_{n+2}.

$$\left.\begin{array}{ll}\text{Step1: } [I_1|O_1,\ I_2|O_2,\ \ldots\ldots,\ I_n|O_n] & \text{forecasting } O_{n+1} \\ \text{Step2: } [I_2|O_2,\ \ldots\ldots,\ I_n|O_n,\ O_{n+1}] & \text{forecasting } O_{n+2}\end{array}\right\} \tag{1}$$

As the output features of our dataset have a dependency on prior time steps data, like weather, the LSTM one-step rolling-forecast worked well for hourly and multi-step for daily (24 h) forecasting with great accuracy.

However, for long duration forecasting, like years, time series decomposition model performs better. The time series decomposition can automatically decompose time series data to prepare an abstract forecasting model using additive decomposition technique as derived in Equation (2).

$$y_t = S_t + T_t + L_t + N_t \tag{2}$$

The derivation of Equation (2) used annotation y_t for energy at time period t with seasonal component S_t, trend-cycle component T_t, average value L_t and random variation (also called noise) N_t in the data series. The long duration forecasting model has considered a season consisting of three months. In the additive decomposition technique of Equation (2), the level (L_t) and seasonal component (S_t) remain steady, whereas trend (T_t) and noise (N_t) differ for a season.

This approach is justified as it enhances the LSTM's ability to interpret complex, time-variant patterns in weather data, leading to more accurate solar radiation predictions.

The proposed framework in Section "Proposed Framework" uses these selected models for prediction and forecasting.

Data preparation

This section has detailed dataset, feature selection, and noise reduction in sections "Dataset", "Selection of features", and "Noise Elimination", respectively, to prepare input for the proposed forecasting framework.

Dataset

The proposed forecasting models are trained using historical data of

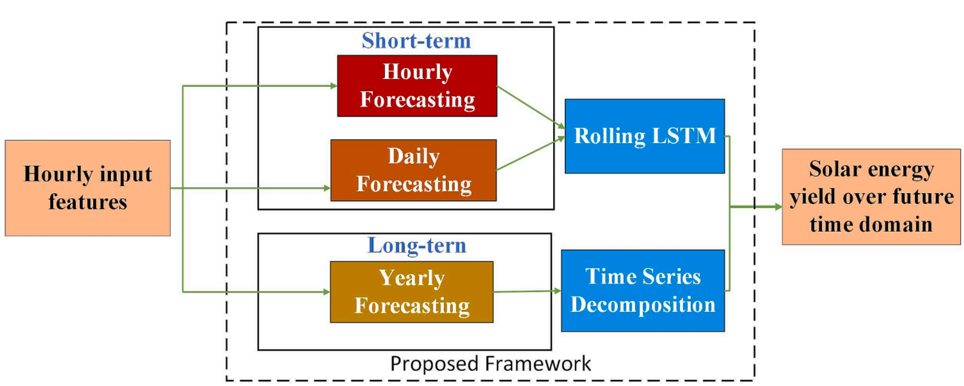

Fig. 2. Proposed forecasting framework.

10 years, from 2005 to 2015, TYM data of large-scale PV stations around Australia from our energy partner [22]. The experiment has considered PV plant data from four different Australian states with different climate profiles, as detailed below.

- **Western Australia:** Albany: Latitude: −34.9333 and Longitude: 117.783.
- **Tasmania: Eddystone:** Latitude: −41 and Longitude:148.3.
- **Victoria: Mildura: Latitude:** −34.2333 and Longitude:142.083.
- **South Australia:** Mt-Gambier: Latitude: −37.75 and Longitude: 140.767.

The selection of diverse climate profile dataset is to ensure the proposed model has generalized implication for all large-scale PV stations. The TYM files (.tm2 file extension) are used as meteorological input data in System Advisor Model (SAM), which is a software developed by the National Renewable Energy Laboratory (NREL) [21], for modelling energy output using pre-existing physical PV systems. The pre-exiting physical PV model of SAM is fine-tuned based on the PV cell inverters and AC lines specification of the PV plants of the data collection points mentioned in the list above.

Based on adjusted physical model specifications of the PV plant and meteorological data from TMY files, the SAM calculates relevant energy production in various time intervals, like hourly, daily and monthly, which can be extracted to an external data storage or in Excel spreadsheet. The initial dataset prepared from SAM's output included hourly energy output and nine meteorological parameters: DNI, DHI, GHI, wind speed, wind direction, humidity, dry bulb, wet bulb, and dew point temperature. Statistical analysis using Python's Panda library [23], written for the Python programming language, is used for data manipulation and analysis to understand the data. For visual analysis, the pandas' group-by method is used to group all data by mean value in specific years, months, days, and hours.

Fig. 3 illustrates yearly grouped mean energy values (kWh) of four selected locations in Australia as listed above. The PV station in Victoria (Mildura) has produced the highest yearly average, 516 (kWh), whereas Albany (WA) produced lowest yearly average 346 (kWh) energy during 2000–2015 as illustrated in Fig. 3. It is noticeable that all states had lesser energy output in 2001 which could be the impact of El Nino˜ in Australia.

The initial analysis of meteorological values helped us to eliminate wind speed, wind direction and humidity parameters as they have very little or no relation with energy output patterns over the years.

To illustrate the influence of meteorological parameters on energy production, Fig. 4 plotted yearly meteorological data with Celsius unit against energy output of Albany PV station where time (years), meteorological parameters (Celsius) and average yearly energy are plotted on x-axis, y1-axis (left) and y2-axis respectively. Based on the yearly data illustrated in Fig. 4, the dry bulb relates with energy output amount and patterns better than the other two parameters.

Fig. 5 illustrates yearly meteorological data (with W/m^2 unit) against the energy output of Mildura PV station where time (years), meteorological parameters (W/m2), and energy are plotted on the x-axis, y1-axis (left), and y2-axis respectively. All (GHI, DNI, and DHI) meteorological parameters (W/m2) have followed historic energy output patterns equally, but DNI has the highest similarity with historic energy output amount, as illustrated in Fig. 5.

The analysis of the rest of the PV stations data showed similar relationships between meteorological and energy output. It is clear from Figs. 3, 4 and 5 that the weather data, therefore energy output, are highly variable. A comprehensive forecasting framework to cover required energy forecasting in short and long-time intervals would be useful for informed operation and maintenance decisions for PV station managers. This analysis gives us the confidence to choose the initial six data parameters for further analysis using feature selection detailed in section "Selection of features". The hourly raw data of selected six data parameters from SAM are processed using feature selection and noise reduction techniques, as detailed in section "Selection of features" and section "Noise Elimination".

Selection of features

To identify the best features for the forecasting model, the features from section "Dataset" are ranked in Fig. 6 based on Feature Importance (FI) and Correlation Index (CI) [10] score. The feature importance method of XGBoost is fast and efficient method to obtain scores of input versus output for each input feature after the boosted trees constructed by gradient boosting. The CI score establishes the degree of the relationship between input and output features [10]. The FI and the CI produce a score between 0 and 1, indicating the value or importance of each input feature for predicting or forecasting output, where a higher score means a more important feature. The condition is formulated as shown in Equation (3) for choosing input features for the forecasting model in our proposed framework.

$$CI(I_F \equiv O_F) > 0.5 and FI(I_F \equiv O_F) > 0.1 \quad (3)$$

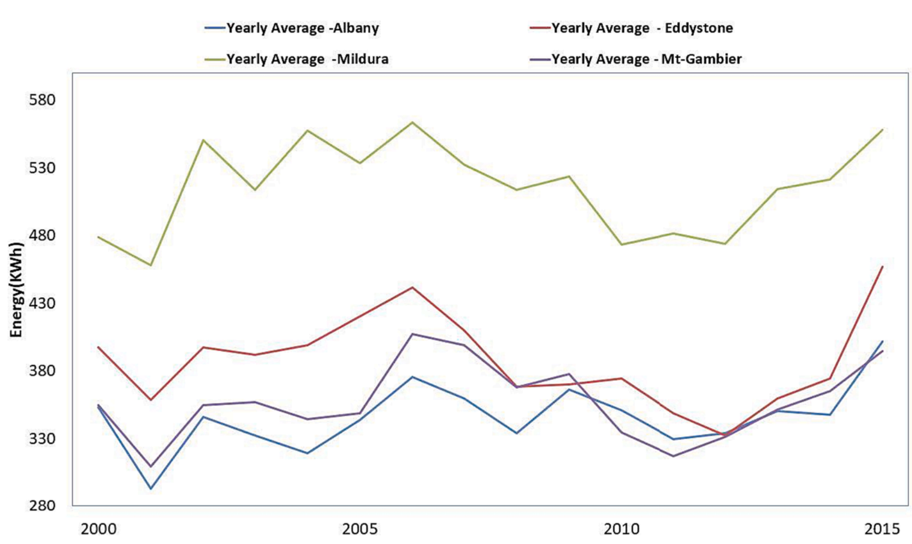

Fig. 3. Visualization of yearly grouped mean data.

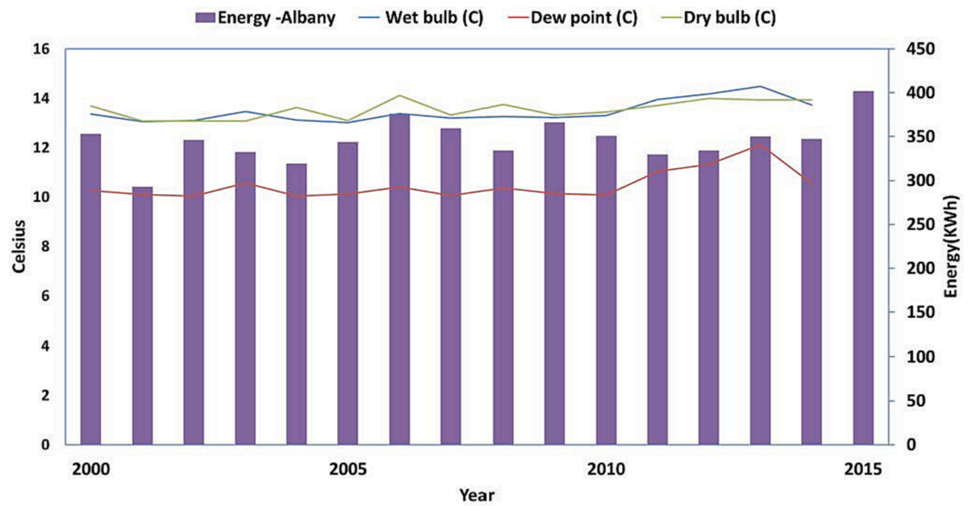

Fig. 4. Historic meteorological data (Celsius) vs energy output of Western Australia.

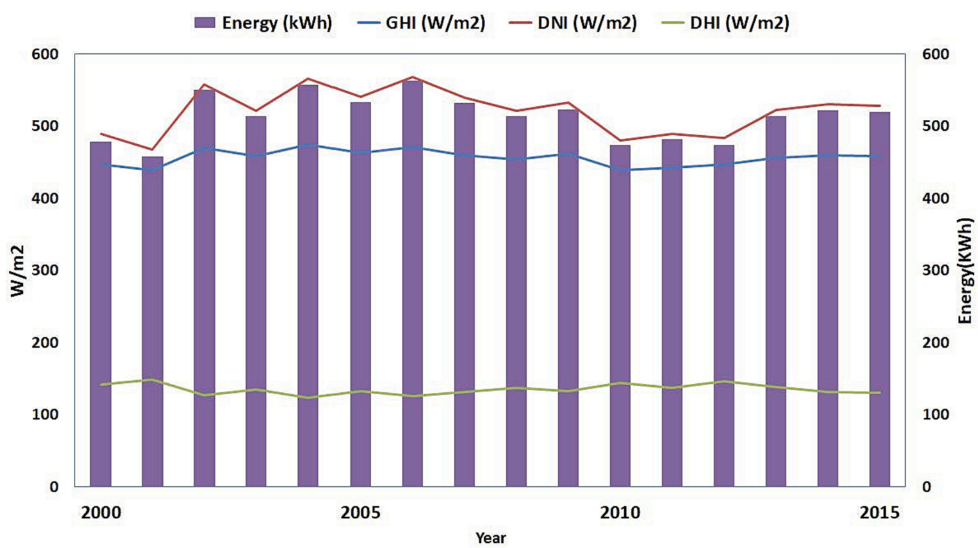

Fig. 5. Historic meteorological data (W/m^2) vs energy output of Victoria, Australia.

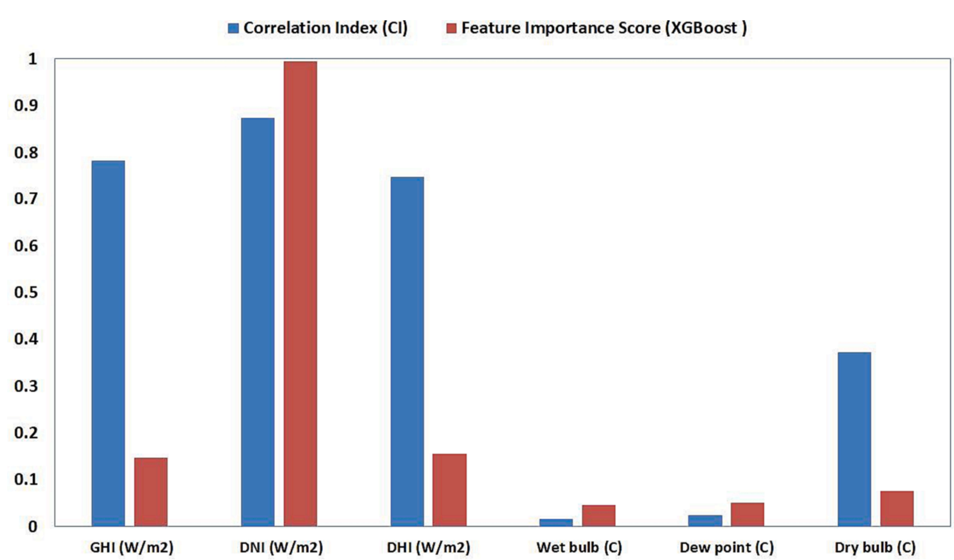

Fig. 6. Selection score by XGBoost (-) and CI (-).

It is evident from the illustrated in Fig. 6 that the GHI, DNI and DHI input features have met conditions derived in Equation (3), therefore, these three input features are used as input in our prediction models with high confidence. It is noticeable that DNI has 94 % score (average of CI and FI) in relation to energy, therefore, the DNI ranked highest among all.

Noise elimination

The hourly dataset from SAM has 25 % noise in Energy (kWh) column like negative values. To prepare the dataset for proper training and testing with a high forecasting success rate, a data cleaning process is very critical. A simple data cleaning method, such as removing the entire row, cannot be implemented due to the high sensitivity of timestamp sequence and high influence of variable input parameters to the output value.

This paper has devised a novel noise elimination algorithm to replace those noise data based on the method proposed in 1. The Algorithm in 1 utilized Extreme Gradient Boosting (XGBoost) [24], which was effectively used with the noisy values by detecting the relationship between characteristics and transient stability based on Phasor Measurement Units (PMU). The learning (L) objective function [17] of XGBoost at iteration t can be expressed as Equation (4) where yi represents training dataset. The l is the loss function to measure the difference between prediction (yi) and target dataset which uses Classification and Regression Trees (CART) function f (x). The term Ω represents penalizing function which balances the complexity of the model.

$$L^{(t)} = {}_{i=1}^{n} l(y_i, y_j^{(t-1)} + f_t(x)) + \Omega(f_t) \tag{4}$$

As presented in Algorithm 1, the noise elimination process uses XGBoost prediction or DNI value to replace noisy entry.

Algorithm 1: Data processing

Algorithm 1: Data processing

Algorithm
for $k = 1 : K$ **do**
  Prepare the training dataset for XGBoost;
  Train the XGBoost model;
  Predict values for negative values in the Energy (kWh) column;
  if $Energy(k) < 0$ **then**
    if $DNI == 0$ **then**
      Replace the noisy data with 0;
    else if *Predicted energy of the row ¿ 0* **then**
      Replace the noisy data with predicted value;
    else
      Replace the noisy data with related DNI value;
    Save the data as CSV format;
  else
    No need to replace the value;
end

As the CI value between DNI and Energy is 94 %, the negative energy values are replaced with DNI value in the first instance, if DBI == 0 as presented in Algorithm 1. This strategy helped us eliminate 5 % noise. For the rest of the noise, the algorithm has trained XGBoost algorithm using input (DNI, DHI, GHI, dry bulb, wet bulb and dew point) and output (Energy) values. The trained XGBoost model helped us to predict the remaining 15 % negative energy values with 91 % accuracy, as detailed in the process of Algorithm 1. Finally, the algorithm has replaced the noisy data with related DNI values for the rest of the 5 % where the XGBoost predicted energy is negative.

As illustrated in Fig. 6, the process has used both the XGBoost and the correlation score of DNI to select the best features.

Experimental Results, comparative Study, and discussion

This section presents experimental results in section "Experimental Results" which is followed by the comparative study and discussion in section "Comparative Study and Discussion".

Experimental results

The experiment result is presented in two sub sub-sections below. In the experiment, each model is evaluated with fivefold cross validation that creates five training and testing sets randomly.

1. Short-term model: The short-term model uses LSTM deep learning technique with the rolling forecast. The proposed framework uses an hourly rolling window to 24 timestamps ahead of time to ensure hourly and uses daily forecasting. The developed short term forecasting model is designed with four LSTM layers and uses Adam [25] optimizer. The model takes the historical hourly dataset as input but uses monthly segment to train the model and test its prediction accuracy before forecasting. Fig. 7 has overlapped hourly historic and predicted data of 7 days (a week) of Feb 2015 from Mildura, Victoria where the X-axis presented the time as days and the Y-axis presented Energy as kWh. The prediction result of Feb 2015 in Fig. 7 has achieved 99 % accuracy and 0.067 normalized RMSE. The rest of the datasets, detailed in section "Dataset", have achieved similar accuracy for prediction using this short-term model.

The model loss and accuracy for the short-term forecasting is illustrated in Fig. 8 and Fig. 9. In total, 100 iterations were used to build the model, which started with a loss of 0.07 and ended with a loss of 0.0038 as illustrated in Fig. 8. The model started with an accuracy of 0.32 and ended with 0.90 after 100 Epoch as illustrated in Fig. 9.

The trained rolling LSTM model is then used to forecast solar yield of 24 h for a specific month of the following year. For example, the model

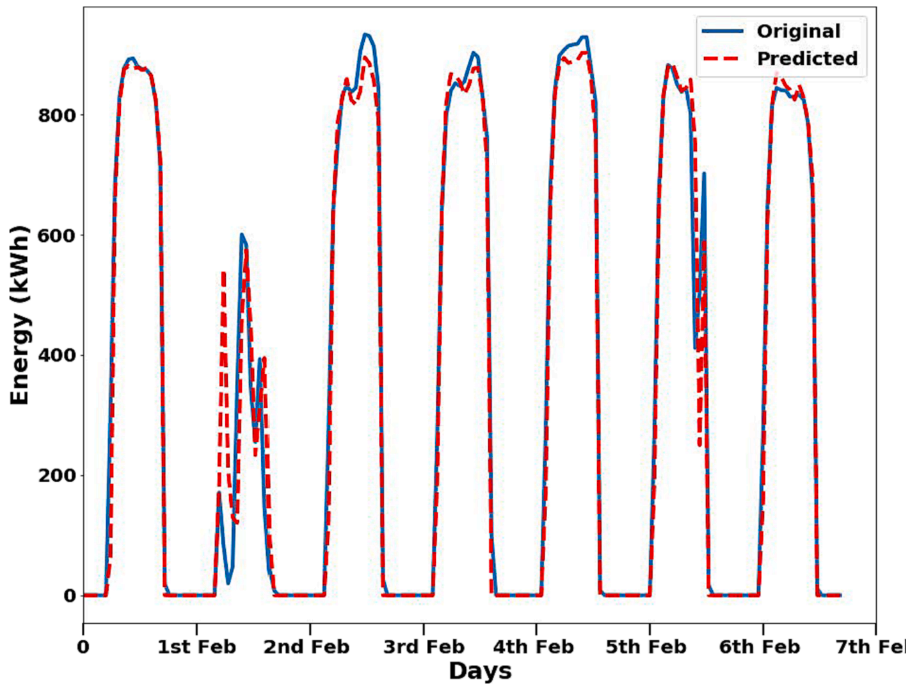

Fig. 7. Prediction result of a week using short-term model.

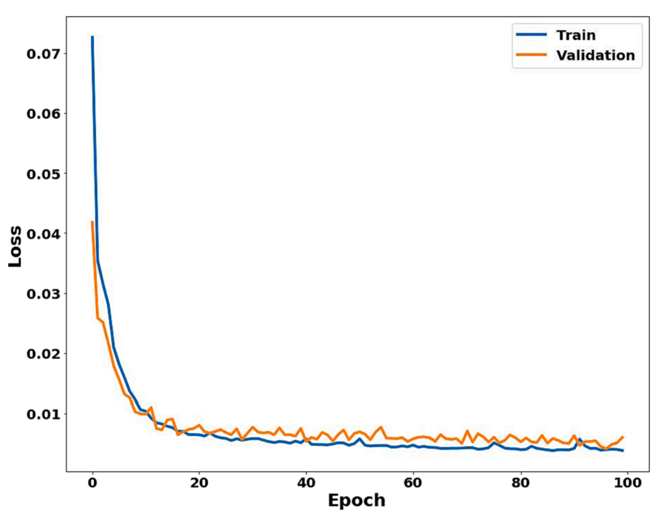

Fig. 8. Model Loss (-) graph of short-term prediction model.

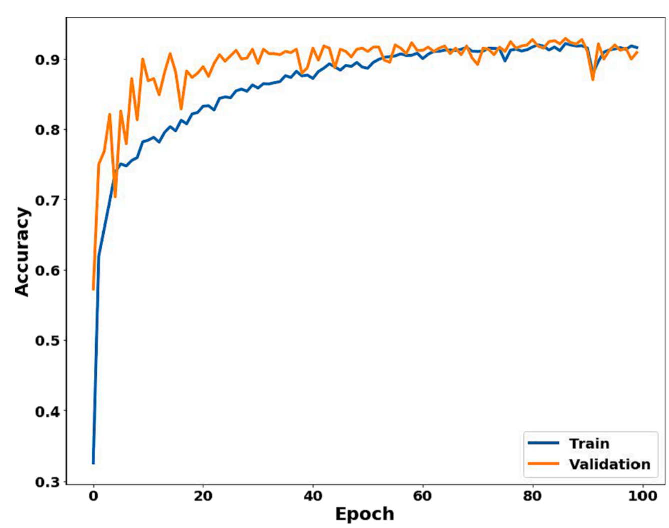

Fig. 9. Model accuracy (-) graph of short-term prediction model.

has used data from February 2015 to forecast the energy yield of 1st February 2016. The Fig. 10 has plotted one month's (February 2015) historic data (blue colored line) with the forecasted data (yellow colored line) of the first day of the following year 2016 (1st February 2016) from Mildura, Victoria where the X-axis presented the time as the day of the month (February) and the Y-axis presented Energy as kWh. The forecasting in Fig. 10 has achieved a 90 % success rate with 0.068 normalized RMSE.

To get a generalized performance indicator of the proposed short-term forecasting model, this paper has experimented using all four datasets detailed in section "Dataset". On average, the trained LSTM model can predict with 98 % accuracy for each month across all our historical data. In contrast, the forecasting has archived an 89.5 % success rate, which varies from 87 % – 92 %, as presented in Table 1. The average Mean Absolute Percentage Error (MAPE) is 5.59 % for the short-term forecasting model with normalized RMSE from 0.11 to 0.14, as presented in Table 1.

2. Long-term model: The long-term forecasting is using a time-series additive model where energy is considered as a function of time. Like the short-term model, the long-term model has taken an hourly dataset as an input as illustrated in the proposed framework in Fig. 2. The long-term forecasting model processes the hourly data to segment as a monthly task, so the features are consecutive to a month of the year. The long-term model uses a quarterly window that means three consecutive months of a year is one quarter. The three months were selected based on patterns of every season with the help of the seasonal component of the time series decomposition model. For example, in Victoria, Australia November to January is considered as summer. Therefore, to predict a year, the long-term model needs to go forward three-time steps ahead of time. Similar to short-term models, the long-term model is also trained using historic data for prediction with 93 % accuracy. Fig. 11 has presented five years prediction results of Mt-Gambier, South Australia, from 2009 to 2014, using a long-term

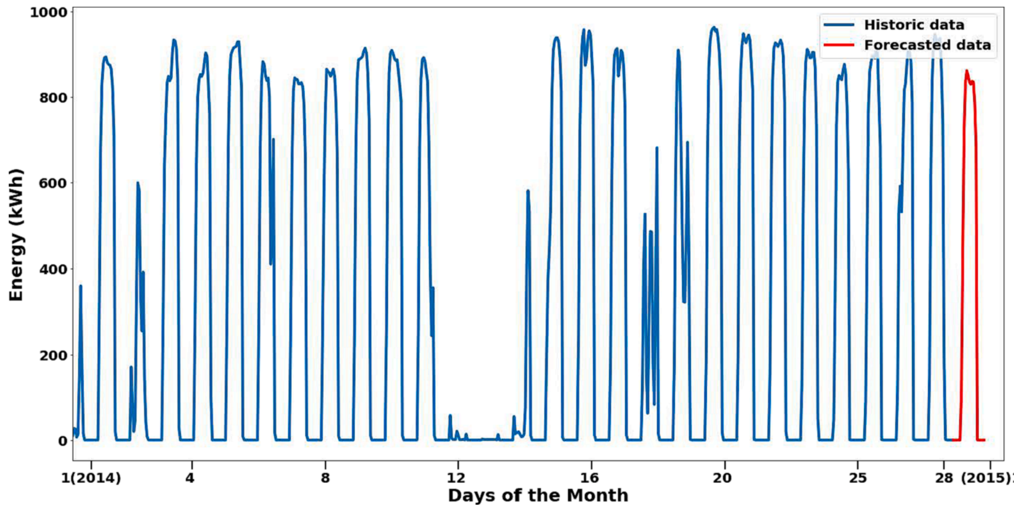

Fig. 10. Historic data (February 2015) with forecasted data (February 2016) using short-term model.

Table 1
Validation metric of short-term forecasting.

Location	Accuracy	MAPE	RMSE	nRMSE
MT Gambier	89	3.01	12.02	0.12
Albany	92	1.9	10.64	0.11
Eddystone	87	4.01	13.05	0.14
Mildura	90	2.17	12.30	0.13

forecasting model where the X-axis presented time as a quarter of a year and the Y-axis presented Energy in kWh.

The trained long-term model is used for forecasting five years ahead of time as presented in Fig. 12 where X-axis presented time as years and Y-axis presented Energy as kWh. This experiment using, Mt-Gambier, South Australia, data has achieved 84 % forecasting accuracy when set with 20-time steps to forecast five years ahead of time. To get a generalized performance indicator of the proposed long-term forecasting model, this paper has performed experiments using all four datasets detailed in section "Dataset". On average, the trained long-term model can predict with 95 % accuracy for five years across all of our historic data whereas the forecasting has achieved an 87 % success rate for one year but 80 % (various from 79 % – 80 %) with a Mean Absolute Percentage Error (MAPE) of 5 % for five years as presented in Table 2.

During prediction, the model has calculated trend component, Tt, over the seasons using seasonal patterns (St) as derived in Equation (2) and illustrated in Fig. 13. The residual of the model shows that the long-term model has performed very well with repeated seasonal patterns but could improve to pick up unusual seasonal pattern changes.

section "Comparative Study and Discussion" has presented a detailed comparative study and discussion between proposed framework and existing similar works.

Comparative study and discussion

It is evident from the experiment that feature selection and noise elimination are two key important ingredients of an effective prediction or forecasting model that uses intermittent dataset with high fluctuation. This finding is supported by recently published literature in similar domain [31]. Therefore, the proposed framework has used multiple data analysis and machine learning based hybrid model to get best features and effective noise elimination which are crucial to get prediction/forecasting outcome appropriate for practical application. Based on the experiment result in section "Experimental Results", the rolling method works well for short term forecasting with steaming or live data collected in short time frequency, like minutes and hours, due to the walk-forward technique for prediction/forecasting as proven in Section

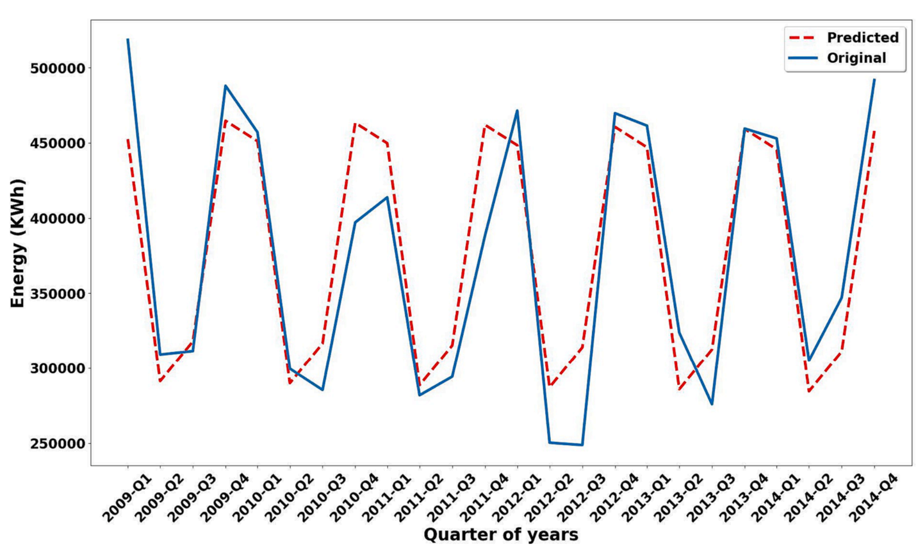

Fig. 11. Quarterly -Yearly Prediction result.

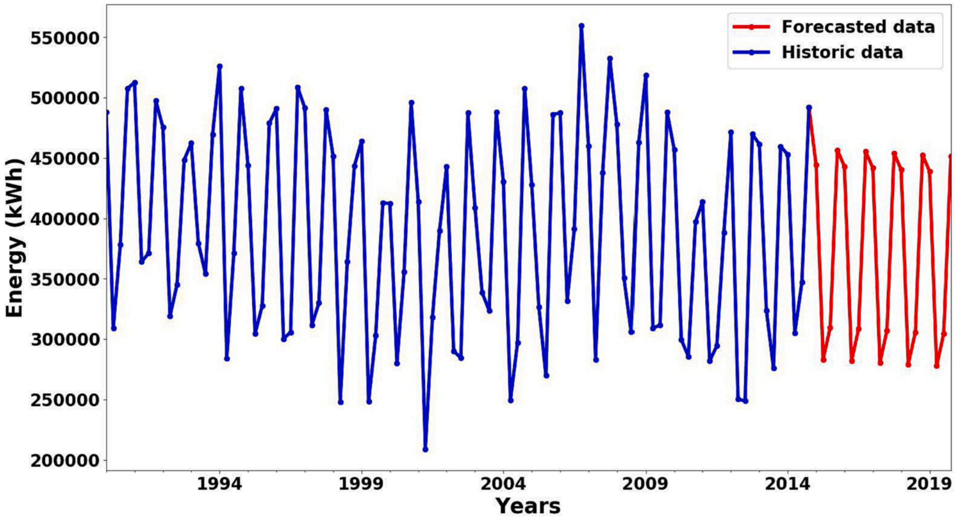

Fig. 12. Yearly forecasting result of Mt-Gambier-SA.

Table 2
Validation metric of long-term forecasting.

Location	Accuracy	MAPE	RMSE	nRMSE
MT Gambier	83	6.54	16.65	0.12
Albany	81	5.15	17.45	0.16
Eddystone	81	6.12	17.05	0.16
Mildura	80	4.56	18.7	0.17

"Forecasting model Selection" and section "Experimental Results". Whereas, due to the variability of the data distribution, it is important to get seasonal patterns grouped together to get effective long-term patterns that will result in better prediction/forecasting, as detailed in Section "Forecasting model Selection" and section "Experimental Results".

In our knowledge, no existing solar yield forecasting framework can predict and forecast for both short and long time. Therefore, it wasn't very easy to prepare a comparative study. Furthermore, much of the existing work has used prediction and forecasting interchangeably with incomplete performance results, so validation results are not always comparable. Keeping all these difficulties in mind, this paper has compared the proposed models with some existing short- and long-term prediction/forecasting models in Table 3 that has considered four important performance parameters, Accuracy (R^2), MAPE, RMSE, and normalized RMSE (nRMSE). Most short-term models in existing literature only presented MAPE and RMSE values as performance indicators. The proposed short-term prediction model has the lowest RMSE and MAPE, whereas the forecasting model has the second-best MAPE and third best RMSE value. It is important to note that all the short-term prediction/forecasting models with lower RMSE than the proposed model in Table 3 have used one specific dataset to validate their model, whereas the proposed model used four datasets with different distributions to validate and then averaged the RMSE for generalization. Both prediction and forecasting of the long-term model have achieved higher accuracy than existing models, as presented in Table 3.

As result detailed in section "Experimental Results", the proposed model's prediction accuracy is higher and error margin is lower than the forecasting accuracy and the error margin. The prediction accuracy and error does not vary when prediction time length increases given similar data distribution, but the forecasting accuracy varies with higher forecasting time steps (duration). For example, the forecasting accuracy will be low, and error will be high for five years ahead of time compared to one year ahead of time. Therefore, based on our experiment result, the forecasting accuracy (R^2) value of a PV yield forecasting model is always dependent on Model design (Md), future time (ti) length, and data distribution (FD) as derived in Equation (5). The accuracy of the forecasting model has an inverse linear relationship with future time (ti) length therefore the accuracy gets lower when the time (ti) value increases which is evident from one year (87 %) and five-year long-term forecasting accuracy value (80 %).

$$F_{R2} = Mf(t_i) + f(F_D) + M_d \tag{5}$$

The prediction accuracy depends only on model design (M_d), target

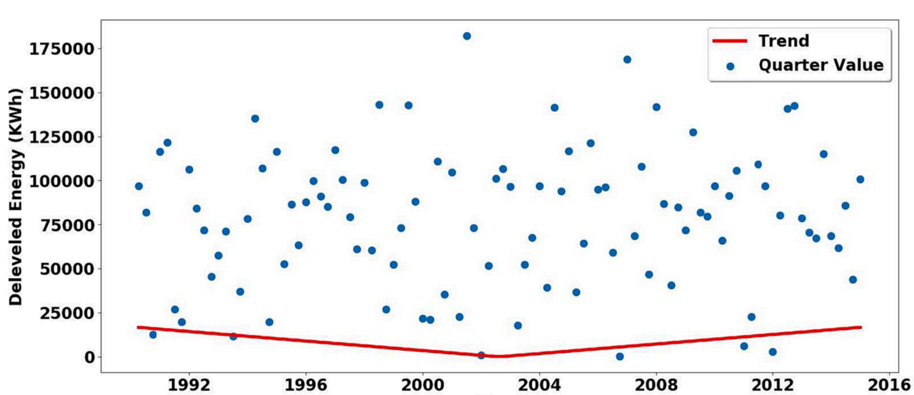

Fig. 13. Seasonal (Quarterly) energy trend pattern.

Table 3
Results of comparative study.

Models	Success Rate	MAPE	RMSE	nRMSE
Short Term Forecasting/Prediction				
Random forest (RF) [2]	–	26.57	88.62	0.19
Neural Network Ensemble(NNE) [4]	91	9.35	–	–
ASEFS(AEMO) - Persistent [3]	–	13.9	18.4	–
North American Mesoscale (NAM) - Persistent[14]	–	9	14.0	0.30
Neural Network with AdaBoost [7]	–	22.74	44.21	–
Weighted k-nearest neighbour (WkNNs) [16]	–	7.7	10.37	–
Radial basis function neural network (RBFNNs) [16]	–	9.14	11.83	–
Regression neural network (PCA-GWO-GRNN) [26]	–	4.96	12.23	–
Genetic algorithm-based SVM (GASVM) [27]	–	1.70	11.22	–
Convolutional Neural Network (CNN) -PVPNet [28]	–	10.94	16.31	–
Deep learning - LSTM - Prediction [29]	99.2	12.45	16	–
Deep learning - LSTM - Prediction [5]	98	8.93	18	–
Variational AutoEncoder (VAE) model- Prediction [5]	98	8.93	18	–
Long Term Forecasting/Prediction				
Deep learning hybrid (LSTM-CNN) [7]	90	2.83	3.89	0.05
Full recurrent neural network (FRNN) [30]	74	–	–	0.49
Proposed Framework				
Short-term Prediction	98	0.566	5.96	0.061
Short-term Forecasting	89.5	2.77	12	0.10
Long-term Prediction	95	0.95	10.5	0.15
Long-term Forecasting	87	5	12.45	0.45

dataset (T_D) and data distribution (P_D) as derived in Equation (6). The data distribution impacts on both prediction and forecasting accuracy that is evident from the results presented in Table 1 and Table 2 where accuracy value varies with the change of dataset.

$$P_{R2} = Mf(t_i) + f(P_D) + M_d \tag{6}$$

Based on the experiment, adapted definitions [9] of prediction and forecasting, it is apparent that prediction and forecasting should be considered two separate categories when presenting the model, framework and their results.

Furthermore, the paper has performed a *t*-test to calculate t-value and p-value on success rate between the proposed model and existing models. For long-term prediction, the t-value $= -1.4474$ whereas the p-value $= 0.098969$ which is significant at p ¡ 0.10. For short-term prediction, the t-value $= -1.16248$ whereas the p-value $= 0.144592$ which is significant at p ¡ 0.15.

Compared to the single purpose prediction/forecasting model, the proposed framework takes higher computational resources and longer to complete the initial training. However, due to lack of comprehensive prediction/forecasting framework like the proposed one, it is not possible to compare the complexity of the proposed framework in numeric scale. The proposed techniques in the framework are validated using diverse climate data from Oceania region, therefore, there can be a need for further training when a new data distribution is introduced to the proposed framework. Finally, the noise eliminations technique may need further refining if the degree of changes in data distribution and noise increase significantly.

Conclusion

A robust solar yield forecasting model can be useful for the energy sector to achieve the best out of the solar renewable energy. This paper has proposed a long-short term solar yield prediction and forecasting model with high accuracy. The research methodology of the proposed framework has used a systematic process to automate feature selection for the proposed models training and testing. A novel noise elimination technique is utilized to process the raw data as a direct input to the proposed framework. To establish generalized error and success rate of the proposed prediction and forecasting models, the paper has used robust validation parameters and datasets of diverse climate conditions. The validation parameters of the proposed techniques and models are compared with similar existing ones in Table 3. It is evident from the comparison table; the proposed framework is effective to understand future solar yield with less error compared to most of the existing models. Due to its higher robustness, adaptability and generalize nature, the proposed framework can be a very useful tool to enable proactive operation and maintenance for grid connect PV systems. The critical findings of this study are below.

- A robust optimization techniques-based method, like XGBoost, is critical to eliminate noise and select appropriate features from large and complicated datasets with high variability in distribution.
- The combination of LSTM with wavelet transform is particularly effective for predicting solar radiation because it addresses climatic data's non-linear and variable nature.
- Integrating robust local mean decomposition with a bidirectional LSTM offers a compelling approach to solar irradiance forecasting.

The proposed noise elimination and feature selection process are valid when the outdoor performance parameters of the PV system are the same. The proposed framework was tested with a dataset from the same geographical area. Therefore, the result will be valid when a similar data distribution is used with the proposed techniques. The experiments have considered a wide range of high-performing machine learning techniques for our experiment but not all existing techniques. In the future, the proposed framework will be tested using diverse data distribution from different geographical regions to make the framework more generalized. It will be very interesting to compare the result of continuous learning and transfer learning with the proposed framework to improve its practical acceptability. Finally, it will be critical to understand the influence of physical, electrical, and thermodynamic parameters on the performance of the frameworks.

CRediT authorship contribution statement

Biplob Ray: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Dimuth Lasantha:** Writing – review & editing. **Vijayalaxmi Beeravalli:** Writing – review & editing. **Adnan Anwar:** Writing – review & editing. **Md Nurun Nabi:** Writing – review & editing. **Hanmin Sheng:** Writing – review & editing. **Fazlur Rashid:** Writing – review & editing. **S.M. Muyeen:** Writing – review & editing.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgment

We want to acknowledge the Centre for Machine Learning, Networking and Education Technology (CML-NET)-CQU for their funding to purchase the dataset. Central Queensland University in Australia provided funding for this study.

references

[1] Kafka JL, Miller MA. A climatology of solar irradiance and its controls across the united states: Implications for solar panel orientation. Renew Energy 2019;135:897–907.
[2] Benali L, Notton G, Fouilloy A, Voyant C, Dizene R. Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. Renew Energy 2019;132:871–84.
[3] Commonwealth Scientific and Industrial Research Organisation (CSIRO), "Australian Solar Energy Forecasting System (ASEFS)," online, accessed on 16th December 2022, May 2016, https://arena.gov.au/assets/2016/07/Aus-Solar-Energy-Forecasting- System-Final-Report.pdf.
[4] Raza MQ, Mithulananthan N, Li J, Lee KY, Gooi HB. An ensemble framework for day-ahead forecast of pv output power in smart grids. IEEE Trans Ind Inf 2019;15(8):4624–34.
[5] F. Harrou, F. Kadri, and Y. Sun, "Forecasting of photovoltaic solar power production using lstm approach," in Advanced Statistical Modeling, Forecasting, and Fault Detection in Renewable Energy Systems, F. Harrou and Y. Sun, Eds. Rijeka: IntechOpen, 2020, ch. 1. [Online]. Available: https://doi.org/10.5772/intechopen.91248.
[6] A. Dairi, F. Harrou, Y. Sun, and S. Khadraoui, "Short-term forecasting of photovoltaic solar power production using variational auto-encoder driven deep learning approach," Applied Sciences, vol. 10, no. 23, 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/23/8400.
[7] B. Ray, R. Shah, M. R. Islam, and S. Islam, "A new data driven long- term solar yield analysis model of photovoltaic power plants," IEEE Access, vol. 8, pp. 136 223–136 233, 2020.
[8] Alanazi M, Alanazi A, Khodaei A. Long-term solar generation forecasting. In: 2016 IEEE/PES Transmission and Distribution Confer- Ence and Exposition (TD); 2016. p. 1–5.
[9] Maleki S, Ray B, Hagh MT. Hybrid framework for predicting and forecasting state of health of lithium-ion batteries in electric vehicles. Sustain Energy Grids Networks 2022;30:100603.
[10] Ray B, Shah R. Performance assessment of prospective pv systems in queensland and new south wales of australia. In: 2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia); 2019. p. 200–5.
[11] Zagouras A, Pedro HT, Coimbra CF. On the role of lagged exogenous variables and spatio–temporal correlations in improving the accuracy of solar forecasting methods. Renew Energy 2015;78:203–18.
[12] Li J, Ward JK. Irradiance forecasting for the photovoltaic systems. In: Proceedings of 2014 International Conference on Modelling, Identification Control; 2014. p. 346–50.
[13] Li J, Ward JK, Tong J, Collins L, Platt G. Machine learning for solar irradiance forecasting of photovoltaic system. Renew Energy 2016;90:542–53.
[14] Larson DP, Nonnenmacher L, Coimbra CF. Day-ahead forecasting of solar power output from photovoltaic plants in the american southwest. Renew Energy 2016;91:11–20.
[15] Sheng H, Ray B, Chen K, Cheng Y. Solar power forecasting based on domain adaptive learning. IEEE Access 2020;8:198580–90.
[16] Zhang Y, Beaudin M, Taheri R, Zareipour H, Wood D. Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators. IEEE Trans Smart Grid 2015;6(5):2253–62.
[17] Rana M, Koprinska I, Agelidis VG. Forecasting solar power generated by grid connected pv systems using ensembles of neural networks. In: 2015 International Joint Conference on Neural Networks (IJCNN); 2015. p. 1–8.
[18] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 11 1997.
[19] Bhatti BA, Broadwater R. Solar photovoltaic output prediction using jackknife regression. In: 2018 North American Power Symposium (NAPS); 2018. p. 1–6.
[20] Zhang D, Han X, Deng C. Review on the research and practice of deep learning and reinforcement learning in smart grids. CSEE J Power Energy Syst 2018;4(3):362–70.
[21] N. Blair, A. Dobos, J. Freeman, T. Neises, and M. Wagner, "System Advisor Model," online, accessed on June 2021, February 2014, https://www.nrel.gov/docs/fy14osti/61019.pdf.
[22] Exemplary Energy Pty Ltd., "Australian Solar and Weather and Climate and microclimate Data for Building Energy Simulation," online, accessed on October 2021, February 2014, http://www.exemplary.com.au/.
[23] NumFOCUS, Inc., "Pandas documentation," online, accessed on October 2021, February 2023, https://pandas.pydata.org/.
[24] Chen T, Guestrin C. Xgboost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '16. New York, NY, USA: Association for Computing Machinery; 2016. p. 785–94.
[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: http://arxiv.org/abs/1412.6980.
[26] Ge L, Xian Y, Yan J, Wang B, Wang Z. A hybrid model for short-term pv output forecasting based on pca-gwo-grnn. J Mod Power Syst Clean Energy 2020;8(6):1268–75.
[27] VanDeventer W, Jamei E, Thirunavukkarasu GS, Seyedmahmoudian M, Soon TK, Horan B, Mekhilef S, Stojcevski A. Shortterm pv power forecasting using hybrid gasvm technique. Renew Energy 2019;140:367–79.
[28] 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. 74 822–74 834, 2019.
[29] Mellit A, Pavan AM, Lughi V. Deep learning neural networks for short-term photovoltaic power forecasting. Renew Energy 2021;172:276–88.
[30] Kazem HA, Yousif JH, Chaichan MT, Al-Waeli AH, Sopian K. Long-term power forecasting using frnn and pca models for calculating output parameters in solar photovoltaic generation. Heliyon 2022;8(1):e08803.
[31] A. Alkesaiberi, F. Harrou, and Y. Sun, "Efficient wind power prediction using machine learning methods: A comparative study," Energies, vol. 15, no. 7, 2022. [Online]. Available: https://www.mdpi.com/ 1996-1073/15/7/2327.
[32] T. Wang, Z. Wang, Z. Huang, and G. Xi. "Multi-domain physics-informed neural network for solving heat conduction and conjugate natural convection with discontinuity of temperature gradient on interface." *Sci. China Technol. Sci.*, Vol. 65, No.2442–2461, 2022.
[33] M. Edalatifar, M.B. Tavakoli, M. Ghalambaz, and F. Setoudeh. "Using deep learning to learn physics of conduction heat transfer." *J Therm Anal Calorim* Vol. 146, no. 1435–1452, 2021.
[34] M. Ghalambaz, M. Edalatifar, S. Moradi Maryamnegari, and M. Sheremet. "An intelligence parameter classification approach for energy storage and natural convection and heat transfer of nano-encapsulated phase change material: Deep neural networks." Neural Comput & Applic, Vol. 35, No. 19719–19727, 2023.
[35] Alizamir M, Shiri J, Fard AF, Kim S, Gorgij AD, Heddam S, et al. "Improving the accuracy of daily solar radiation prediction by climatic data using an efficient hybrid deep learning model: Long short-term memory (LSTM) network coupled with wavelet transform. Eng Appl Artif Intell 2023;123(Part A):106199.
[36] Singla P, Duhan M, Saroha S. An integrated framework of robust local mean decomposition and bidirectional long short-term memory to forecast solar irradiance. Int J Green Energy 2023;20(10):1073–85.
[37] Cuce E, Mert Cuce P, Karakas IH, Bali T. An accurate model for photovoltaic (PV) modules to determine electrical characteristics and thermodynamic performance parameters. Energy Conversion and Management 2017;146(2017):205–16.
[38] Cuce E, Bali T, Anil Sekucoglu S. Effects of passive cooling on performance of silicon photovoltaic cells. Int J Low-Carbon Technol 2011;6(4):299–308.
[39] S. B. Riffat, E. Cuce, "A review on hybrid photovoltaic/thermal collectors and systems, International Journal of Low-Carbon Technologies", Volume 6, Issue 3, September 2011, Pages 212–241, 2011.
[40] Cuce E, Cuce PM. Improving thermodynamic performance parameters of silicon photovoltaic cells via air cooling. Int J Ambient Energy 2014;35(4):193–9.