Draft version March 8, 2023
Typeset using LaTeX default style in AASTeX631

Renewable energy prediction using hybrid LSTM-BiLSTM-GRU model v6.3.1*

Deepak Kanneganti 20380423[1]

[1] *Curtin University*
Bentley, Perth, WA, Australia,6152

ABSTRACT

Renewable energy sources are widely adopted as they are much safer to generate energy with less atmospheric harm. Solar Power prediction depends on sunlight to gather solar power effectively, and one of the significant disadvantages is weather dependency. The solar system's efficiency drops due to weather changes, like a cloudy and rainy day, and these weather changes have a noticeable effect on the energy system. To overcome the problem of weather uncertainty, we present a novel Deep-Learning based on a hybrid multivariate time series model LSTM-BiLSTM-GRU mode. A continuous learning framework with a model decay modeller is introduced to retraining to prevent weather drift and ensure that models in production provide beneficial results. Real-time data is used in this paper to evaluate the efficiency of the proposed model. The hybrid model proposed in a study performs better compared to the individual algorithms designed for solar power prediction. The model achieved a score of 0.945 r2 scores.

Keywords: solar power prediction; Long short-term memory (LSTM), Bidirectional Long short-term memory (BiLSTM), Gated recurrent units(GRU), model decay modeller

1. INTRODUCTION

According to the recent statistic published by BDO Global [1], by the end of 2024, renewable energy will hold 33 per cent of the world's electricity. Solar energy is a fast-growing source that has reached around 60 per cent, according to the International Energy Agency 2019 report[1]. Compared to other energy sources like fossil fuel and nuclear power plants, the operations are intermittent in solar production and only depend on the weather conditions Barque et al. (2018). Grid operators use solar power forecasting to manage the electric grid and balance energy generation effectively. This process is unavoidable because high variability in power predictions can lead to fire accidents and damage electronic devices.

Hence, the solar forecasting model plays a crucial role in alarming the grid operations to make different operations like planning electricity distribution and meeting market demands. This paper proposes a solar forecasting model for a season head prediction. A wide range of literature reviews focusing on solar power prediction has categorized power prediction as physical, statistical and artificial Intelligence based techniques Yang et al. (2021). Simple and complex are the two categories in physical power prediction. According to the research by Hammer et al. (1999), a simple model uses weather and satellite observations to forecast power. The study conducted Cheng et al. (2021) supports that applying a simple model combined with HIRLAM mesoscale weather pattern has generated poor results on 21 power stations in the Jutland peninsula. The physical methods base model can achieve better accuracy under reasonable conditions. The study shows Perez et al. (2002) that these characteristics will get highly complex at stages.

The statistical method uses a large volume of data related to power output and other unknown features to explain a relationship between power and unknown features. Based on the count of unknown features, it is further classified to use a liner, multiple and nonlinear regression. The literature adopts Li et al. (2011) multi liner relationship between solar radiation and ambient temperature and establishes a linear relationship between these features.The Support vector machine method is adopted in the literature Li & Li (2008)to predict solar farm power due to the minimization of risk and generalization ability. Further, SVM is considered the optimal solution to deal with quadratic optimization

* Released on March, 1st, 2021

problems. An improved least squares version supports vector machine (LSSVM) referred to in Zhu & Tian (2011). The study conducted Di Piazza et al. (2016) is an instance of nonlinear regression models. The researcher proposed NARX and NARMAX models with features like solar radiation, daytime and temperatures as input variables. The literature Pedro & Coimbra (2012), Bouzerdoum et al. (2013) proposed using ARIMA and SARIMA models. ARIMA mode is a non-stationary method applied to predict the power plant.

Ye et al. (2022)

The latest studies show that forecasting power based on artificial intelligence algorithms has recently become very popular. Nowadays, Artificial intelligence-based design models are back propagation (BP) neural network prediction models Kaushika et al. (2014),Artificial Neural Networks (ANN) Praynlin & Jensona (2017),Regularized-ELM (R-ELM) Cao & Lin (2008) and deep neural network models like Long Short Term Memory [LSTM] Gensler et al. (2016), Bidirectional Long Short Term Memory [Bi-LSTM] Peng et al. (2021), and Gated recurrent units (GRU) Wojtkiewicz et al. (2019).

Even though these models perform well to a specific limit, solar forecasting models are weather-dependent. Our primary objective is to design a hybrid multivariate model that can predict solar energy with minimal error. For this goal, a critic will evaluate the model performance of all the individual models and takes the values with minimal error. The second goal is to overcome the data drift and model delay. Our idea is to design a model decay monitor to identify the model decay performance after every quarter and help the algorithms to gain new knowledge by retraining the model. The novelty of our continuous retraining learning approach is that using new data is contradictory to a stable dataset. This novel approach helps the model extract new insights from the weather data. This paper will describe the dataset, methodology, hybrid models structure, continuous retraining framework and results.

2. CONTRIBUTION

Thus, this paper aims to develop a fine-tuned BERT text classification model that works for the cross-domain and develop local agnostic Explainable models.

1. List of contributions

   - Designed a hybrid multivariate LSTM-BiLSTM-GRU model that performs better than the individual models
   - Designed a continuous learning frame work with a model decay monitor that improves the life time of the model

3. LITERATURE REVIEW

The research on solar power prediction is necessary to maintain electricity distribution and meet market demands. Due to the following request, much research has been proposed in the broader literature to solve the challenges with solar power prediction.

The modelling means of solar forecasting techniques are mainly three types physical, statistical and artificial intelligence-based methods Ye et al. (2022). These methods have their limitations in forecasting power generation. Physical modelling uses numerical weather prediction (NWP), classified as local and wide area predictions. Models NAM Mathiesen & Kleissl (2011), MM5 Fernandez-Jimenez et al. (2012), and WRF Lima et al. (2016) are developed and applied in the PV power prediction of local area models. The wide-area methods are covered in Mathiesen & Kleissl (2011); the author used GFS and ECMWF. These models are more accurate in forecasting under cloudly and cloudless situations.One drawback of the physical model technique is determining the output character and obtaining accurate predictions.

statistical modelling techniques do not require geographical data but mainly use a considerable volume of historical data to identify the insights using statistics. The authors in Li et al. (2011) implement linear regression and support vector machine Li & Li (2008), advance SVM in Zhu & Tian (2011), ARIMA and SARIMA models in Pedro & Coimbra (2012) , Bouzerdoum et al. (2013). The complex part of dealing with the statistical model is estimating the regression and statistical equations for high accuracy.

Nowadays, Artificial intelligence-based modelling for power predictions has become popular due to its self-adoption and learning abilities. The author in Praynlin & Jensona (2017) implemented an Artificial neural network using solar radiation data. The author of this study Praynlin & Jensona (2017) implements the Back Propagation algorithm and Radial basis function network. Extreme learning machine(ELM) and Wavelet Neural Networks (WNN) are other AI models implemented by the authors in the study Suyono et al. (2018) and Cao & Lin (2008)

Another possible way to categorize different approaches is to sort them by their length forecasting horizon. The area of short-term forecasting represents less than three months discussed in Giebel et al. (2011), Costa et al. (2008), mid-term forecasting is less than a year Mirasgedis et al. (2006), and long-term forecasting techniques are for more than two years forcing Hong et al. (2013).

4. DATA DESCRIPTION

This paper uses ten years of data solar grid and weather data from (January 2013 to December 2022) at Curtin University Green energy electricity park located in Bentley, Western Australia. Weather parameters like temperate(C), humidity (), solar radiation(w/m2), wind speed(m/s), and wind direction are the independent features in this analysis, along with the module temperature from the solar grid. The solar panel change direction throughout the day based on the sun's moment—a poly-crystal solar panel-generated data power is the target variable for this analysis.The nominal power ranges between 50W and 1100W.Table 1 contains detailed information on the list of features and their units used in this study.

All the datasets are available monthly, with all the weather and solar power features for every minute. The data is categorized into three different volumes according to the model requirement. The multivariate model LSTM, BiLSTM, GRU and hybrid mode LSTM-BiLSTM-GRU are trained with seven years of consecutive data and evaluated in the coming seasons to evaluate their performance. The continuous retraining framework uses four years of historical data to train the hybrid model, and the model's performance is further evaluated quarterly for retraining purposes.

S.No	Features	units	range
1	Power(P) TS1	Watt/m	15-1300w
2	Module Temperature	C/m	1-63
3	Barometer data	Hg	1000-1033
4	Outside Humidity	g.m3	15-98
5	Outside Temperature	C/m	7-34
6	Solar Radiation	W/m2	0-1388
7	Wind Speed	m/s	0-74
8	Wind Direction	degrees	0-255
9	Date	dd/mm/yyyy	01-01-2013-31-12-2022
10	Time	hh:mm	06:00 - 18:35

Table 1. List of different domains datasets used in the experiments

5. DATA PROCESSING AND MODELING

5.1. *Data Processing*

Data reprocessing is the initial stage to ensure reliable information gets training the model. In this stage, a massive volume of historical data is extracted from the power station in text file format. These data files have gone through two stages of data cleaning discussed in the further analysis.

5.1.1. *Data Cleaning and feature extraction*

The data cleaning is the initial stage, where we identify the percentage of missing values, outliers, and duplicates and remove them as they are ineffectual for the analysis. Generated solar power gets stored in the inverter after random intervals. During this period, the software and sensor value are offline, resulting high amount of outliers in the data. These contextual outliers are identified and removed to increase the quality of the data. The weather data may not change every minute. Due to this, it contains servile duplicated values identified and removed from the data. All the sensor data points are at different ranges as mentioned in table 1, data is normalized between 0 and 1 using the min-max normalization.

The historical weather data contain several features, and feature selection algorithms are applied to identify the factors contributing to the power prediction. They ensure better performance of the LSTM-BiLSTM-GRU model; features with high correlation coefficients to the target power prediction are considered for model training.

5.1.2. *Time series generator*

The primary challenge with time series data is to design a structure for the data with input and output components. The deep learning models discussed in this study require a multivariate tie serial sable. The image gives a visual refraction of the structure of the time series generator. The timestamp (t) contains the time steps with input and target. For instance, the time steps t0 to t7 are the inputs, and t8 is the target. The tables on the left side of the images show the movement of the input and target values at each time step. In this study, the one-time step contains 12 h of data as input and predicts the following hour's prediction. Each time step includes the following parameters like batch size 32, sample rate 1 and 12 features are used. This data structure divided the dataset with this study helps the model learn the change in the weather patterns for better prediction. The model's performance gets evaluated by comparing the output values of the times step t0 with the labels.

S1	S2	S3	S4	S5	S6	S7	Target
1	2	3	4	5	6	7	8
2	3	4	5	6	7	8	9
3	4	5	6	7	8	9	10
4	5	6	7	8	9	10	11
5	6	7	8	9	10	11	12
6	7	8	9	10	11	12	13

Table 2 Time series and target

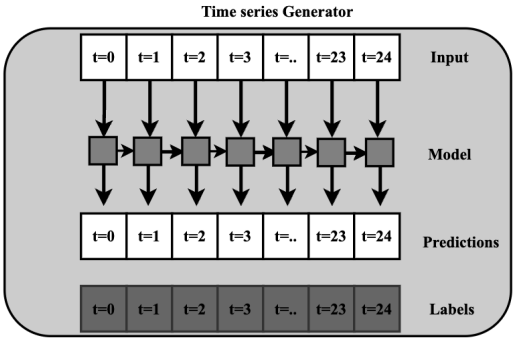

Figure 1. Time series generator

5.2. *LSTM-BILSTM-GRU Model*

5.2.1. *Long short-term memory (LSTM) and Bidirectional-Long short-term memory (LSTM)*

LSTM neural network stands for Long short-term memory. LSTM model was introduced in 1997 by LSTM Hochreiter & Schmidhuber (1997). This neural network aims to solve the vanishing and exploding gradient descent problem in recurrent neural networks (RNNs).RNN works for sequence data tasks like stock price prediction Zhu (2020) and machine translation Cho et al. (2014); the major drawback is that the increase in long sequence can affect the learning. LSTM has overcome this issue due to its unique studies different from ordinary neural networks. The concept of gates inside the model structure helps to process sentiment analysis Ombabi et al. (2020), clustering analysis Liu et al. (2020), processing Xiang et al. (2020) and time series data Sagheer & Kotb (2019).

The structure of the LSTM mode is in the left side of **figure 3**.The LSTM model has four major input parts: The input layer, the input gate's control signal, the forgetting gate's control signal, and the output gate's control signal. The LSTM model takes the input vector (h(t-1),x(t)) at time step t, and the output number will be 0 or 1 at each cell state (Ct-1). The Sigma layer is the decision-making layer, also called a forget gate layer.

$$f_t = \sigma(W_f.[h_t - 1, x_t] + b_f) \tag{1}$$

In time series data, the cell stage might include a set of patterns to predict the following states, and later for a new set of patterns, previous information is the forgotten. The input gate i(t) layer decides to update the information. Moreover, the tanh layer creates a new vector for new cell state C'. An updated state is created using equation (2) and (3). The old sell C(t-1) is update with Ct.The older state ft is multiplied by the previous cells state C(t-1) and added (2)*(3) resulting the new candidate value equation (4).

$$i_t = \sigma(W_i.[h_t - 1, x_t] + b_i) \tag{2}$$

$$\hat{C}_t = tanh(W_c.[h_t - 1, x_t] + b_c) \tag{3}$$

$$C_i = (f_t * C_t - 1) + (i_t * \hat{C})) \tag{4}$$

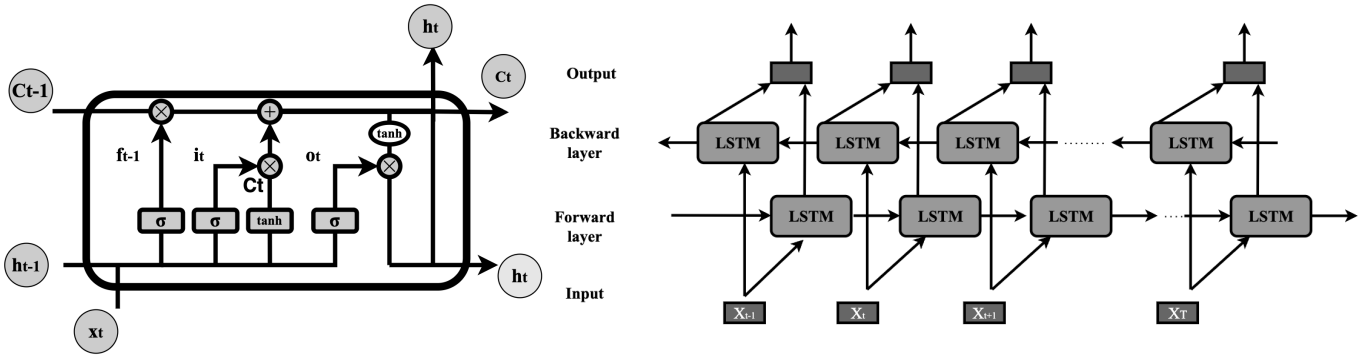

Figure 2. LSTM and Bi-LSTM model

Finally, the output will depend on the cell state of the filtered version. In the initial stages, the sigmoid layer outputs the cell state and then goes through the tanh layer, resulting in an output ranging from -1 to 1.The sigmoid gate generates the output of a required part. Finally the last step ,output value of ht represent in equation 5.

$$o_t = \sigma(W_o.[h_t - 1, x_t] + b_o), h_t = o_t * tanh(C_t) \tag{5}$$

Bi-direction LSTM mode is an extended form of the LSTM model; it uses two hidden layers to process time series data in forward and backward directions. The right side diagram of **figure 3** shows the block diagram of the Bidirectional LSTM model. Schuster and Paliwal introduced the bi-LSTM model in 1997 Schuster & Paliwal (1997). Studies Baldi et al. (1999) show that applying LSTM twice for forward and backward propagation has improved the performance of long-term dependence and the model's learning and accuracy.The literature study shows BiLSTM is for Natural language processing tasks like sentiment analysis Sharfuddin et al. (2018), name entity recognition Dai et al. (2019), clustering Fan et al. (2018) ,and text classification tasks Liu & Guo (2019). BiLSTM expression is showed in formula (6)and (7) where sigma is the activation function, Ht is the hidden layer input, and the output is generated by updating the forward propagation $\overrightarrow{ht}$ backward propagation $\overleftarrow{ht}$.

$$\overrightarrow{h_t} = \sigma(W_t.[h_t - 1, x_t] + b_t) \tag{6}$$

$$\overleftarrow{h_t} = \sigma(W_t.[h_t - 1, x_t] + b_t) \tag{7}$$

$$H_t = \sigma(W_f + W_b) + b_y \tag{8}$$

5.2.2. *Gated recurrent units(GRU))*

Gated recurrent units model is a recurrent neural network introduced by Kyunghyun cho Cho et al. (2014)in the year 2014. GRU models are similar to LSTM models with a forget gate and additional parameters.GRU models are similar to LSTM models with a forget gate and additional parameters. Researcher studies show that the GRU model can provide high performance on time series predictions Zhang et al. (2017), speech emotions recognition Zhu et al. (2020), polyphonic music modelling Liu & Yang (2019), and natural language processing tasks like text classification Zulqarnain et al. (2019).

The standing RNN has drawbacks with vanishing gradient problem, GRU uses to update and reset the gate, two vectors decide to pass the information as output.**Fig 4** shows the structure of the GRU model. The standing RNN has drawbacks with the vanishing gradient problem, which GRU uses to update and reset the gate; two vectors decide to pass the information as output.Equation (9) and (10) represents the output and reset gate. The input x(t) and h(t-1) gets multiplied by their weight W(t) and U(t); these both results gets summed, and the output goes through the sigmoid function, which gives a result ranging from 0 to 1.The reset gate decides the past information needs to rest. The reset gate decides the past information needs to rest. The formal is the same as s output, h(t-1) and x(t) multiply them with their weights, sum these weights, and apply the sigmoid function.

$$Z_t = \sigma(W_t * x_t + U_t * h_t{-}_1) \tag{9}$$

$$r_t = \sigma(W_r * x_t + U_r * h_t{-}_1) \tag{10}$$

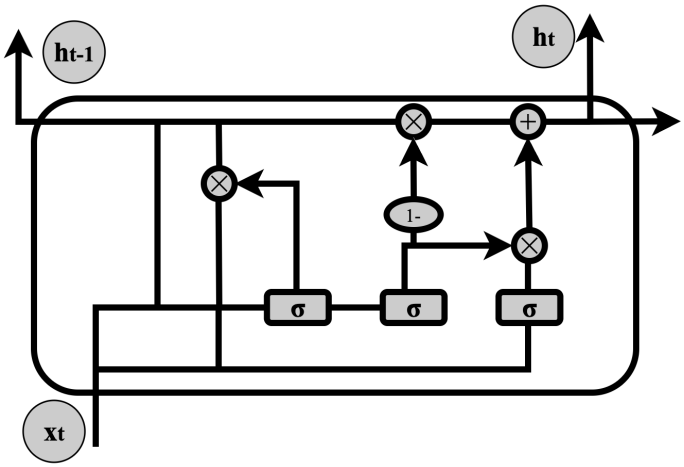

Figure 3. Gated recurrent units(GRU)

Equations 11 and 12 represent the current state's content memory unit and final memory, and memory content units title the related information from the past. The initial step is multiplying the input x(t) and weight along with h(t-1) and weight Ut. The Hadamard product is the sum of the x(t), h(t-1), and weights. The final step is to apply the nonlinear activation function tanh. For the final output equation (12), the network calculates the h(t) vector containing the current unit and pass to the network. The update gate calls the current memory h' (t) and previous state h' (t-1). An element-wise multiplication is applied on updated gates z(t) and h' (t-1), and for (1-Zt) and h' (t). The final out of the model is the sum of the two products, mentioned in equation 12

$$\hat{h} = tanh(W x_t + r_t * U_{ht}{-}_1) \tag{11}$$

$$h_t = (Z_t * h_t{-}_1 + (1 - Z_t)) * \hat{h}_t \tag{12}$$

5.2.3. *Architecture of LSTM-BILSTM-GRU*

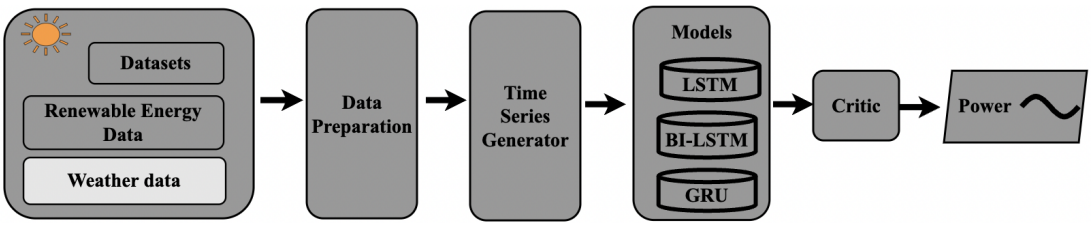

Figure 4. LSTM-BILSTM-GRU Model

The Hybrid model LSTM-BILSTM-GRU model combines three multivariate time series LSTM Hochreiter & Schmidhuber (1997), BILSTM Schuster & Paliwal (1997) and GRU Cho et al. (2014) deep neural network models used to predict solar power predictions. All these models use the wealth and solar features discussed in **section 4**. **Figure 4** illustrates the structure of the proposed hybrid model.

This study uses TensorFlow to design the deep neural network to forecast solar predictions. The DNN model constructed in those papers contains three DNN recurrent layers and three fully connected layers. Dropout layers are added after each DNN model to overcome the over-fitting problem. The study by Gal & Ghahramani (2016) suggest a dropout rate of 0.3 for medium-size models, making the neurons discharge and readjust for each training

Parameters	**LSTM**	**BiLSTM**	**GRU**	**LSTM-BiLSTM-GRU**
Layer	(512,256,128-64)	(512,256,128-64))	(512,256,128-64)	(512,256,128-64)
Activation function	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU
aplha	0.5	0.5	0.5	0.5
learning rate	0.001	0.001	0.001	0.001
Drop out	0.3	0.3	0.3	0.3
Alpha	0.5	0.5	0.5	0.5
Loss function	MSE	MSE	MSE	MSE
Training epochs	25	25	25	40

Table 2. Evaluation index result table of LSTM-BiLSTM-GRU model

sample. The table contains the parameters used to design the models discussed in this study. Layer, learning rate, dropout, alpha, loss function and episodes are the parameter assigned manually. The DNN models' architecture has four hidden layers with a neuron size of 64,128,256 and 512. Out of various comparisons, the model structure has no sign of over-fitting and validation error is minimal. Mean square error(MSE) is the loss function for these models. Adam optimizes Kingma & Ba (2014)the algorithm with a learning rate of 0.001 by updating the network weights. Leaky ReLU is the activation function used to avoid the dying rule problem with a constant alpha value of 0.5.Epochs are single training iterations for all batches in the forward and backward propagation. The models are trained for 25 Epochs, and a hybrid model with 40 epochs.

Finally, a critic evaluates these model's values by calculating the difference between the expected value. The critic considers the predictions with minimal errors. The main idea is to calculate the prediction difference of each model with the actual values. N represents the samples, yi original values,$\hat{y}$ represents the model's prediction,k indicates the models as the critic evaluated the k=1,2,3 three models and finds the optimal value to predict. The difference between the yi, and $\hat{y}$ is indicated in the formula (13). Optimal values generated by the critic are considered the final predictions. The performance of this model is evaluated and discussed in the following sections visually.

$$Critic = \sum_{i=1}^{N} (yi - \hat{yi}, k), i = 1, 2, 3, 4, 5, 6... \tag{13}$$

5.3. *Retraining Framework*

The figure describes the overview of the proposed new novel approach to retraining the model. The objective is to improve the accuracy of solar predictions using a continuous retraining framework after a specific period. The model decay monitor evaluates the performance of the model after every quarter. The data from 2013-16 got trained, and the model decay monitor estimates the version after every quarter and displays the statistics. The BiLSTM model performance is better when compared with the other deep learning models used in this study. The main objective is to develop a retraining model framework to overcome the data drifting problem. The BiLSTM model for this framework contains four hidden layers with a neuron size of 64,128,256 and 512 and Leakey Re-LU activation function with an alpha value of 0.5. The model updated the network weights with the Adam optimizer, which uses a learning rate of 0f 0.001. The model got trained on 24 epochs.

The model decay monitor's role is to evaluate the model's promise after the specified intervals. Model decaying is a significant challenge due to a change in the character of the input data. The model trained on weather data is vulnerable to the model decaying problem due to the unexpected change in the climate. A model decay monitor is introduced in this study. This monitor's objective is to perform the statistical and visual analysis of the model performance to make further decisions on the model retraining. The statistical analysis contains the results of the evaluation metrics to understand the change in the model performance. The visual study shows a graphical representation of the model decaying after each interval. Considering the outcomes of the two analyses, the decision to retrain will be further by the model decay monitor.Deciding the volume of the data that gets retrained is an issue, and retraining the model again with an entirely new dataset requires more computing resources. This study proposes a new approach to retraining the model considering the following factors. A continuous decaying in the model performance will leverage to retrain the model with the last quarter. This act will help the model to exclude the old data and include the knowledge from the new data.

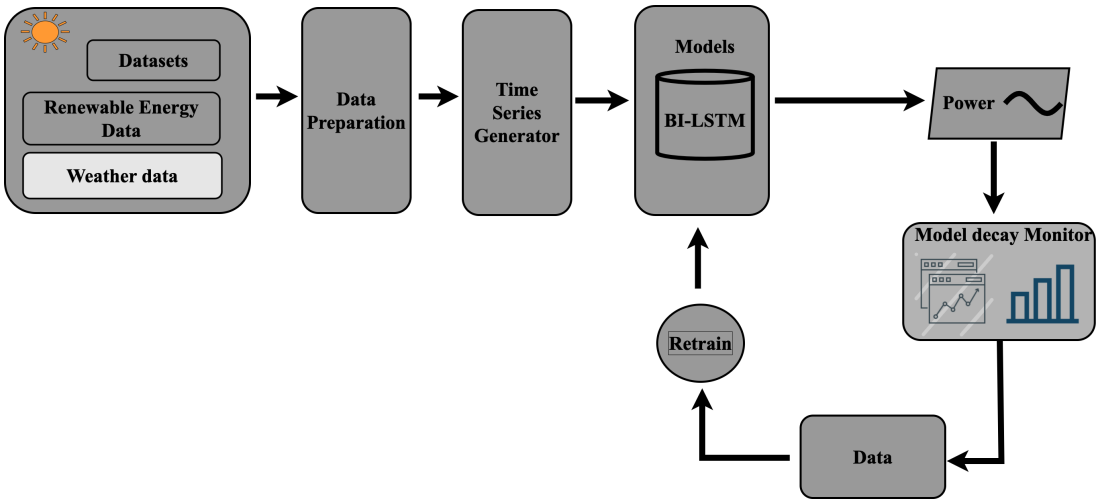

Figure 5. Continuous learning LSTM-BiLSTM-GRU

6. EVALUATION METRICS

summarizes the quantitative comparison results of all the models using the Mean Absolute Percentage Error (MAPE),R2 Score, Mean Square Error (RMSE), Mean Absolute Error (MAE), and which can be expressed according to **Eqs. 14, 15,16,17 and 18**

$$MAPE = \frac{100}{N}\sum_{t=1}^{N}\frac{|Et - At|}{At} \tag{14}$$

r2 score is the coefficient of determination, a regression score function. The value ranges from[0,1], the higher the value. The R2 score formula is given in the equation 14.

$$R2 = 1 - \frac{\sum_{i=1}^{N}(yi - \hat{y})^2}{\sum_{i=1}^{N}(yi - \overline{y})} \tag{15}$$

Mean Absolute error calculates the absolute of the observation of the dataset to the model prediction. The average of the observation is calculated to consider the opposing error. Equation 15 describes the formula of the MAE.

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|yi - \hat{y}| \tag{16}$$

The final evaluation metric is Mean squared error (MSE); it calculates the error by assessing the mean square difference between the actual and predicted values. The lesser the MSE value, the better the model performs, expressed in equation 16.

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(yi - \hat{y})^2 \tag{17}$$

7. RESULTS AND DISCUSSION

7.1. *LSTM-BiLSTM-GRU*

The results of individual multivariate LSTM, BiLSTM, GRU and hybrid model outcomes are in the **table 2**. The regression metrics discussed in previous section 6 are the evaluation factors for the model's performance. In this study, all these models are processed parallelly with the same time series data to generate the results. The experimental results show that the individual models cannot activate ideal prediction. After detailed examinations of the individual model results, a hybrid model with a critic that chooses the optimal value is proposed. The result shows that the mixed model LSTM-BiLSTM-GRU proposed by this study is the best solution for solar power prediction. The following

graphs give a visual understanding of models that forecast the power prediction for the next quarter for every minute. In this experiment, the models were trained on the data from January 2013 to August 2017 and predicted for October, November and December 2017 for every month. **Table 2** has the list of experiment results of all the models. This paper makes the minute solar forecasting from [01/10/2017- to 31/12/2017]. The model's performance for each month is displaced separately in the table.

Time period	**Model**	**R2 score**	**MSE**	**MAE**
5 minutes time step	LSTM	0.92	0.008	0.05
	Bi-LSTM	0.921	0.0086	0.046
	GRU	0.87	0.95	0.087
	LSTM-BiLSTM-GRU	0.937	0.006	0.051
Monthly	**Model**	**R2 score**	**MSE**	**MAE**
October	LSTM	0.906	0.009	0.056
	Bi-LSTM	0.91	0.008	0.049
	GRU	0.856	0.014	0.086
	LSTM-BiLSTM-GRU	0.9347	0.006	0.047
November	LSTM	0.908	0.0098	0.051
	Bi-LSTM	0.906	0.009	0.050
	GRU	0.850	0.016	0.093
	LSTM-BiLSTM-GRU	0.927	0.0078	0.0541
December	LSTM	0.929	0.007	0.043
	Bi-LSTM	0.930	0.007	0.039
	GRU	0.878	0.012	0.0839
	LSTM-BiLSTM-GRU	0.940	0.006	0.051

Table 3. Evaluation index result table of LSTM-BiLSTM-GRU model

The first four rows of the table display the evaluation metrics values of the LSTM model, BiLSTM model, GRU model and LSTM-BiLSTM-GRU for the next two days with a time step of 5 minutes. Among them, the hybrid model has the highest value of the R2 score, which determines the model's over or under-fitting. This value ranges from [0,1]; in this case, the hybrid model has the highest R2 score, near 0.94. The LSTM AND BiLSTM model R2 score value is the same, which is 0.92. Mean absolute error and Mean Square Error reflects the actual situation of the predicted value error. These metrics give a better understanding of the standard of the model. MAE AND MSE of the hybrid mode are significantly less than individual model predictions, which are 0.05 and 0.006. Out of all the models, the GRU performance is less.

The following rows in the table show the results of the models for the next quarter. The October, November and December data set is tested to evaluate the model performance and consistency in the prediction. In October, the hybrid model's performance remained better than the individual forecasts with an R2 value of 0.93, MSE 0.006 and MAE 0.05. The R2 score value of LSTM and BiLSTM is 0.92. The results of November and December show clear evidence of consistency in the model performance. The hybrid model results are better than the individual prediction.

The plots in **figure 6** visually represent all the models with the actual values—the solar forecasting for the following two days with a time step of five minutes. The pots give a visual understating that the forecasts of the LSTM and BiLSTM are almost the same. The GRU model performance is when compared with these two models individually. One of the drawbacks of the GRU model, from the left side graph**figure 7** , is that the model performance is not good in predicting solar power more fabulous than 1000 kW.

7.2. *Continuous learning LSTM-BiLSTM-GRU*

The plots in **figure 8** visually represent the model performance for the following quarters. In this retraining framework, the model gets trained on three years of data from January 2013 to December 2016. The model decay monitor evaluates the model's performance after every quarter, and results show that the version of the mode is decaying over a while. According to the researchers, the weather data drifts after a period due to climate changes. The evaluation metrics like Mean Absolute Percentage Error (MAPE), R2 Score, Root Mean Square Error (RMSE)

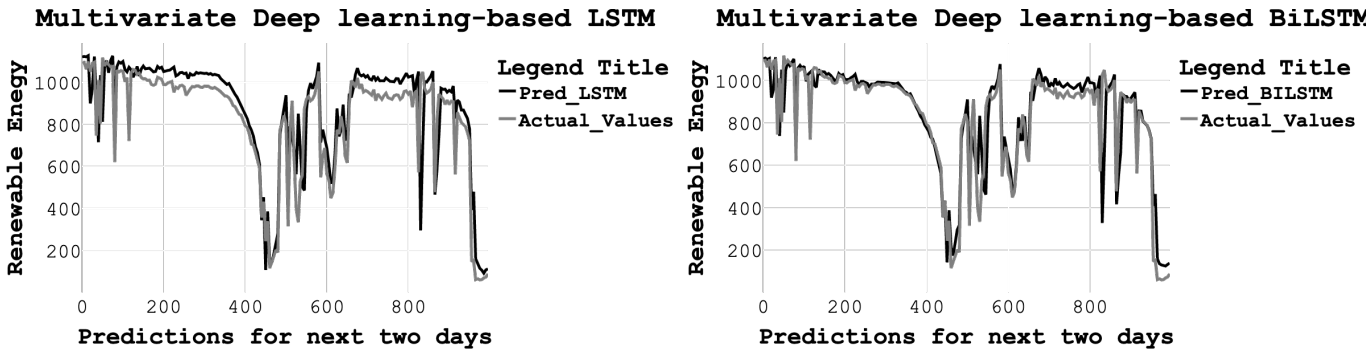

Figure 6. LSTM and BiLSTM model

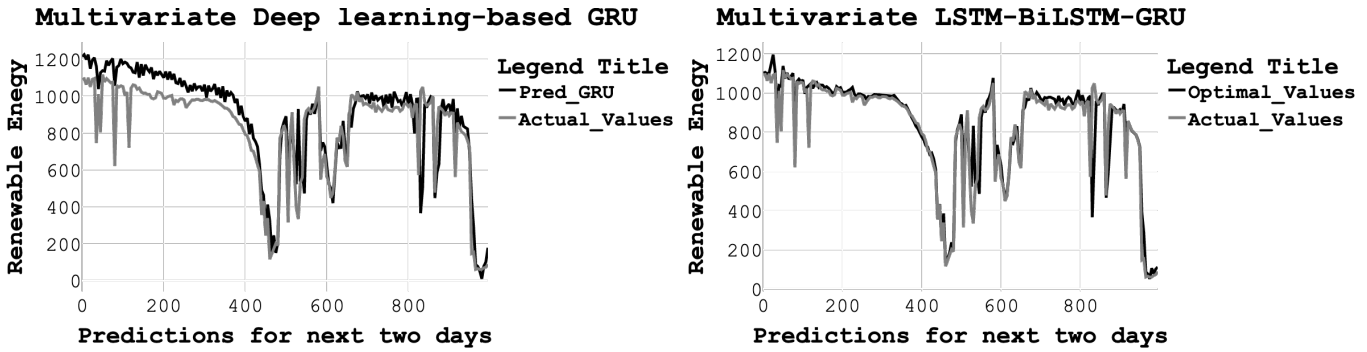

Figure 7. LSTM and BiLSTM model

and Mean Absolute Error (MAE) evaluated the model performance for the following quarters. Table 4 contains the results of the model at different time intervals. The second rows show the model's performance in training from 2013 to 2016—the BiLSTM model R2 score value is 0.88, RMSE 129.04 and MAE 92.42. MAPE is a forecasting evaluation metric, the sum of the individual absolute errors divided by each period separately. The MAPE score of the training model is 23.13. Considering this as the optimal parameter, the model decay monitor evaluates the model performance after every quarter. The following three rows contain the model results for the subsequent quarters. The results show that quarter 1 and 2 has shown a closer performance in the model.

Time period	**Model**	**R2 score**	**RMSE**	**MAE**	**MAPE**
Trained 2013-16	Bi-LSTM	0.883	129.01	92.42	23.54
Period	**Model**	**R2 score**	**RMSE**	**MAE**	**MAPE**
Quarter 1	Bi-LSTM	0.8897	123.3	91.85	19.8
Quarter 2	Bi-LSTM	0.877	116.45	77.34	18.18
Quarter 3	Bi-LSTM	0.8747	131.549	86.95	22.11
Retraining	Bi-LSTM	0.9486	70.9436	29.5691	5.71

Table 4. Evaluation index result table of retraining framework

The performance of the model got decayed in quarter three. The RMSE and MAE for quarter three are 131.54 and 86.95, higher than the previous period. The Rsquare score reached 0.87; add the MAPE is 22.11. Comparing the model's volume and performer with training, it is understandable that its performance is decaying after some time. Weather change and data quality are the primary reasons behind the performance deteriorating.

8. CONCLUSIONS

This study proposes the LSTM-BILSTM-GRU model to predict better solar predictions for every minute for the following season. Data pre-processing is the initial stage which includes the following aspects: data cleaning, data

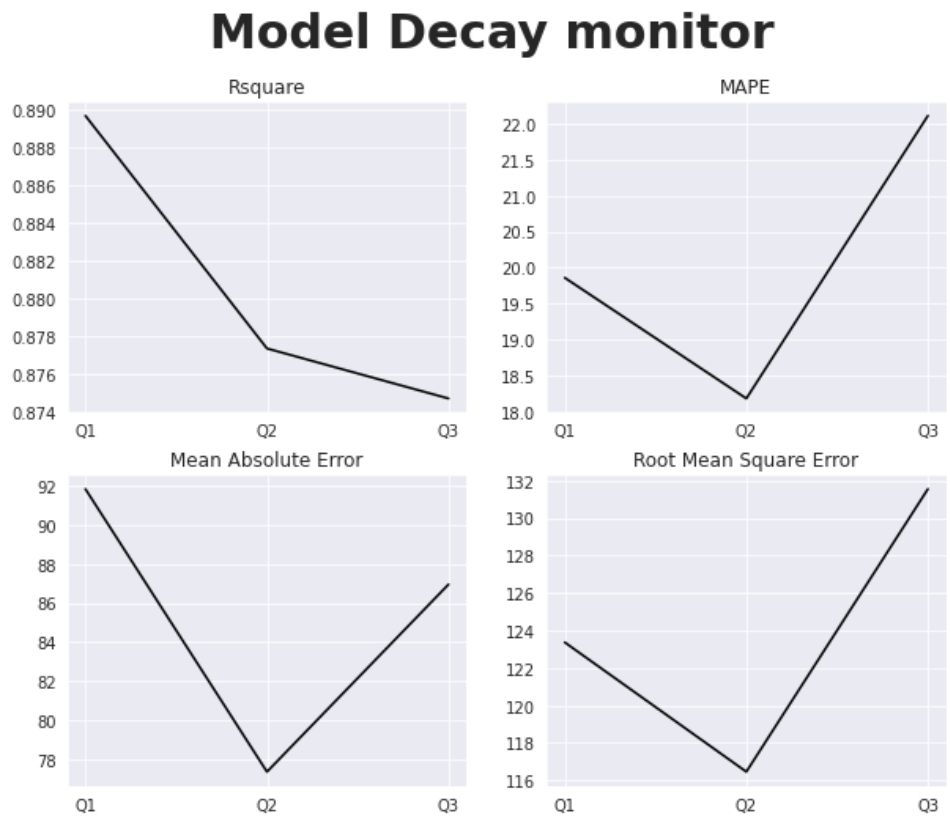

Figure 8. Model decay monitor results

mining feature engineering, feature extraction and feature scaling. Later, multivariate time series models like the LSTM, Bi-LSTM, and GRU models carried out the solar forecasting for the same dataset. Then, a critic evaluated the three models' predictions and decided on the prediction values with minimal error. Finally, regression evaluation metrics like R2 score, MSE and MAE assess the model performance. The model's performance at different time intervals is compared visually using time series plots. The hybrid model proposed in this study enables us to reach the following conclusions.

1. List of outcomes

    - Experiment results of the models show that the accuracy of the hybrid model is better than the individual models. Due to the mechanism of the critic, the final prediction combines all three models' best values, resulting in a better performance outcome.
    - The comparative analysis of the models shows the following results. The R2 score of the Hybrid model is higher than that of the individual LSTM model, Bi-LSTM model and GRU model for seasonal and monthly predictions. The Mean Square Error (MSE) and Mean Absolute error(MAE) values are less for the Hybrid model than for the individual models. These results conclude that the proposed model is more effective when compared with the individual models for solar forecasting.

.

9. ACKNOWLEDGMENTS

This research is part of the Green Electric Energy Park(GEEP) funded by Curtin University, WA. The authors would also like to thank Pawsey super-computing for providing computing resources. In addition, the author would like to thank Dr Sajib Mistry and Dr Sumedha Raja karuna for their helpful insights and for providing us with relevant code and data related to their projects.

APPENDIX

REFERENCES

Baldi, P., Brunak, S., Frasconi, P., Soda, G., & Pollastri, G. 1999, Bioinformatics, 15, 937

Barque, M., Martin, S., Vianin, J. E. N., Genoud, D., & Wannier, D. 2018, in 2018 international workshop on big data and information security (IWBIS), IEEE, 43–48

Bouzerdoum, M., Mellit, A., & Pavan, A. M. 2013, Solar energy, 98, 226

Cao, J., & Lin, X. 2008, Energy Conversion and Management, 49, 1396

Cheng, L., Zang, H., Wei, Z., et al. 2021, IEEE Transactions on Sustainable Energy, 13, 629

Cho, K., Van Merriënboer, B., Gulcehre, C., et al. 2014, arXiv preprint arXiv:1406.1078

Costa, A., Crespo, A., Navarro, J., et al. 2008, Renewable and Sustainable Energy Reviews, 12, 1725

Dai, Z., Wang, X., Ni, P., et al. 2019, 1

Di Piazza, A., Di Piazza, M. C., & Vitale, G. 2016, Renewable Energy and Environmental Sustainability, 1, 39

Fan, Y., Gongshen, L., Kui, M., & Zhaoying, S. 2018, IEEE Access, 6, 57460

Fernandez-Jimenez, L. A., Muñoz-Jimenez, A., Falces, A., et al. 2012, Renewable Energy, 44, 311

Gal, Y., & Ghahramani, Z. 2016, Advances in neural information processing systems, 29

Gensler, A., Henze, J., Sick, B., & Raabe, N. 2016, in 2016 IEEE international conference on systems, man, and cybernetics (SMC), IEEE, 002858–002865

Giebel, G., Draxl, C., Brownsword, R., Kariniotakis, G., & Denhard, M. 2011

Hammer, A., Heinemann, D., Lorenz, E., & Lückehe, B. 1999, Solar Energy, 67, 139

Hochreiter, S., & Schmidhuber, J. 1997, Neural computation, 9, 1735

Hong, T., Wilson, J., & Xie, J. 2013, IEEE Transactions on Smart Grid, 5, 456

Kaushika, N., Tomar, R., & Kaushik, S. 2014, Solar Energy, 103, 327

Kingma, D. P., & Ba, J. 2014, arXiv preprint arXiv:1412.6980

Li, G., Liao, H., & Li, J. 2011, J. Yunnan Norm. Univ, 31, 33

Li, R., & Li, G. 2008, Electric power, 41, 74

Lima, F. J., Martins, F. R., Pereira, E. B., Lorenz, E., & Heinemann, D. 2016, Renewable Energy, 87, 807

Liu, G., & Guo, J. 2019, Neurocomputing, 337, 325

Liu, J., Gong, M., Guo, W., et al. 2020, International Journal of Computational Intelligence Systems, 13, 717

Liu, J.-Y., & Yang, Y.-H. 2019, arXiv preprint arXiv:1906.01203

Mathiesen, P., & Kleissl, J. 2011, Solar Energy, 85, 967

Mirasgedis, S., Sarafidis, Y., Georgopoulou, E., et al. 2006, Energy, 31, 208

Ombabi, A. H., Ouarda, W., & Alimi, A. M. 2020, Social Network Analysis and Mining, 10, 1

Pedro, H. T., & Coimbra, C. F. 2012, Solar Energy, 86, 2017

Peng, T., Zhang, C., Zhou, J., & Nazir, M. S. 2021, Energy, 221, 119887

Perez, R., Ineichen, P., Moore, K., et al. 2002, Solar Energy, 73, 307

Praynlin, E., & Jensona, J. I. 2017, in 2017 Innovations in Power and Advanced Computing Technologies (i-PACT), IEEE, 1–7

Sagheer, A., & Kotb, M. 2019, Neurocomputing, 323, 203

Schuster, M., & Paliwal, K. K. 1997, IEEE transactions on Signal Processing, 45, 2673

Sharfuddin, A. A., Tihami, M. N., & Islam, M. S. 2018, 1

Suyono, H., Santoso, H., Hasanah, R. N., Wibawa, U., & Musirin, I. 2018, Indonesian Journal of Electrical Engineering and Computer Science, 12, 691

Wojtkiewicz, J., Hosseini, M., Gottumukkala, R., & Chambers, T. L. 2019, Energies, 12, 4055

Xiang, S., Qin, Y., Zhu, C., Wang, Y., & Chen, H. 2020, ISA transactions, 106, 343

Yang, B., Zhong, L., Wang, J., et al. 2021, Journal of Cleaner Production, 283, 124628

Ye, H., Yang, B., Han, Y., & Chen, N. 2022, Frontiers in Energy Research, 702

Zhang, X., Shen, F., Zhao, J., & Yang, G. 2017, in Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14–18, 2017, Proceedings, Part V 24, Springer, 523–532

Zhu, Y. 2020, 1650, 032103

Zhu, Y., & Tian, J. 2011, Power system technology, 35, 54

Zhu, Z., Dai, W., Hu, Y., & Li, J. 2020, Pattern Recognition Letters, 140, 358

Zulqarnain, M., Ghazali, R., Ghouse, M. G., & Mushtaq, M. F. 2019, JOIV: International Journal on Informatics Visualization, 3, 377