Renewable Energy 204 (2023) 39–58

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Near real-time wind speed forecast model with bidirectional LSTM networks

Lionel P. Joseph [a,*], Ravinesh C. Deo [a], Ramendra Prasad [b], Sancho Salcedo-Sanz [c,a,**], Nawin Raj [a], Jeffrey Soar [d]

[a] School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, QLD, 4300, Australia
[b] Department of Science, School of Science and Technology, The University of Fiji, Saweni, Lautoka, Fiji
[c] Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, 28805, Madrid, Spain
[d] School of Business, University of Southern Queensland, Australia

ARTICLE INFO

Keywords:
Wind speed forecasting
Intermittent renewable energy
Boruta feature selection
Bayesian optimization
Bidirectional long short-term memory

ABSTRACT

Wind is an important source of renewable energy, often used to provide clean electricity to remote areas. For optimal extraction of this energy source, there is a need for an accurate and robust wind speed forecasting. The intermittent nature of wind makes this goal quite challenging. This research proposes a novel hybrid bidirectional LSTM (BiLSTM) model for near real-time wind speed forecasting. The hybrid model is developed using wind speed and selected climate indices from a group of neighbouring reference stations as predictors to forecast wind speed of a target station. A 3-stage feature selection is applied on the predictors to robustly extract highly significant input features. Stage 1 employs partial auto-correlation and cross-correlation, stage 2 uses the RReliefF filter algorithm, and Boruta-RF wrapper method is implemented in the final stage to improve the BiLSTM model with an efficient Bayesian optimization used for hyperparameter tuning. The proposed model has been benchmarked with comparative models including standalone and hybrid LSTM, RNN, MLP and RF. The proposed hybrid BiLSTM algorithm is found to be superior in wind speed prediction for all tested sites with $\approx 76.6 - 84.8\%$ of errors being $\leq |0.5|$ ms^{-1}. The hybrid BiLSTM model also registered the lowest Relative Root Mean Square Error $(9.6 - 23.8\%)$ and Mean Absolute Percentage Error $(8.8 - 21.5\%)$ among all the tested algorithms. This research ascertains that the proposed model can accurately predict wind speed and capacitate wind energy availability to be regularly monitored at a near real-time level.

1. Introduction

An increased dependence on fossil-derived non-renewable energy sources has prompted countries to set climate mitigation targets, focused on the transition into renewable alternatives (RE) [1,2]. In 2015, 196 countries ratified the Paris Agreement commitments in support of Sustainable Development Goal (SDG) target 7 [3]. Similarly, Fiji, a Pacific Small Island Developing State (PSIDS) has set an ambitious target of 100% RE by 2036 [4]. Around 64% of this target has been met mainly through hydro- and bio-energy sources [5]. Fiji will need to add more RE capacities to attain 100% RE status, as highlighted in the updated Nationally Determined Contributions (NDC) roadmap [4]. As a vital component of greener resource, wind energy can significantly help Fiji in meeting its RE and NDC targets. In fact, Fiji has a good wind regime, so more energy could be harnessed [6,7]. However, the status of wind-based energy in Fijian electricity grid is currently negligible, with just one wind farm that has not yet realized its full potential [8]. Therefore, to meet the 100% RE target, it would require additional wind power to have a diverse mix of RE resources. Wind speed (U, hereafter) has a significant role in wind power generation. Wind power is cubically proportional to U, and a sudden twofold change in U can affect the wind power output by eight times [9]. The available U needs to be greater than the wind turbine's cut-in speed and less than its cut-out speed for a successful wind farm operation. During cyclonic events, the U often exceeds the cut-out speed. The turbines used in Fiji (i.e. Vergnet GEV 275 kW model) can be lowered during cyclonic period to avoid damages, but pre-planning is needed. However, this is not an option in Fiji since there are no forecasting tools in place. Hence, turbines get damaged resulting in monetary losses from repair and maintenance work. This leads to irregular wind power causing disturbance in the power quality and system stability [10]. Due to the stochastic nature of U, the energy generated is intermittent [11].

* Corresponding author.
** Corresponding author at: Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, 28805, Madrid, Spain.
E-mail addresses: lionel.joseph@usq.edu.au (L.P. Joseph), ravinesh.deo@usq.edu.au (R.C. Deo), ramendrap@unifiji.ac.fj (R. Prasad), sancho.salcedo@uah.es (S. Salcedo-Sanz), nawin.raj@usq.edu.au (N. Raj), jeffrey.soar@usq.edu.au (J. Soar).

https://doi.org/10.1016/j.renene.2022.12.123
Received 11 May 2022; Received in revised form 25 November 2022; Accepted 30 December 2022
Available online 4 January 2023

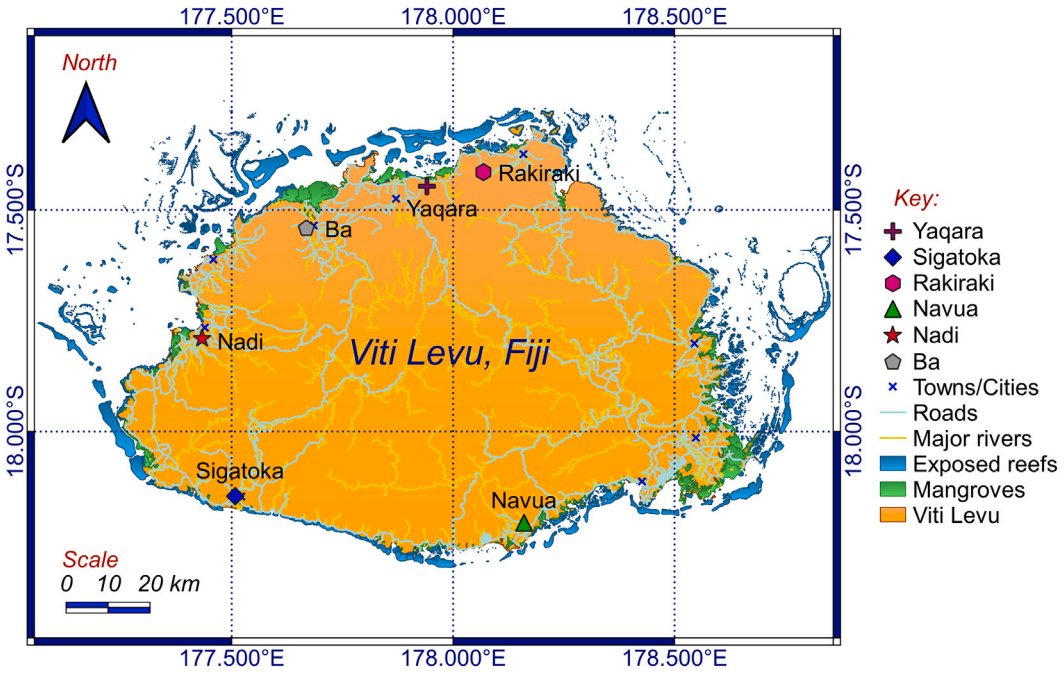


Fig. 1. Map of Viti Levu, Fiji Islands showing the six selected study sites and their geographical locations.

Therefore, an accurate near real-time wind speed predictive model is a prerequisite to ensure a safe and stable wind energy conversion.

Wind speed prediction can be achieved via the physical-based and artificial intelligence (AI) models. Physical-based models (e.g. numerical weather prediction — NWP) are used to forecast long-term U using physical data such as terrain, roughness, obstacle, atmospheric pressure and ambient temperature [12]. However, these models have a high computational demand and require detailed descriptions of the diverse weather variables, which are not always available [13]. Physical models perform poorly while forecasting near real-time U. Conversely, AI models (i.e. Machine Learning — ML and Deep Learning — DL) are useful for near real-time U forecasting. AI models are data-driven, and provide a strong nonlinear forecasting skill as they capture the future variations of near real-time U and extract relevant features from historical data [14]. The ML and DL models commonly applied in U forecasting include: artificial neural networks (ANN) [15], multilayer perceptron (MLP) [16], multiple linear regression (MLR) [17], support vector regression (SVR) [18], random forest (RF) [19], decision tree (DT) [20], k-nearest neighbour (KNN) [21], deep belief networks (DBN) [22], stacked autoencoders (SAE) [22] and recurrent neural networks (RNN) [23]. Among these, only ANN, MLR and RF have been applied to forecast U at selected Fijian sites [17,24]. These studies implemented standalone modelling that have innate drawbacks in terms of their generalization competence.

Unlike standalone models, hybrid methods offer better predictive performance. Alongside increased interest in hybrid methods, there has been an upsurge in the number of DL models like convolutional neural network (CNN) [25], gated recurrent unit (GRU) [26], long short-term memory (LSTM) [27] and bidirectional LSTM (BiLSTM) [28] used in the literature recently. DL can solve more complex problems relatively well provided sufficient datasets are available [29]. One of the commonly used DL models is RNN [30], which unfortunately face short-term memory problem due to exploding and vanishing of gradients. To solve this issue, two specialized versions of RNN introduced were GRU [31] and LSTM [32]. GRU is a less complex architecture with two gates (i.e. reset and update). LSTM is similar to GRU but has additional gates in its architecture (i.e. input, forget and output), which provides an added advantage in remembering longer sequences of data. Ma et al. [33] proposed an LSTM-based double decomposition hybrid tool for short-term U predictions. The double decomposition strategy helped in reducing the complexity and non-stationarity of each series, improving the LSTM's predictive performance. However, LSTM can only use the forward information to make predictions, but BiLSTM [34] considers the future sequence of information in reverse recursive pattern, allowing the model to learn the forward and backward details simultaneously.

The dual flow of information in BiLSTM facilitates efficient learning of long-term dependencies. Hence, BiLSTM networks have been applied to U prediction and related problems. To test this, Xiang et al. [35] hybridized the BiLSTM network with wavelet transform (WT) decomposition to forecast multi-step short-term U. WT-BiLSTM performed better in forecasting longer time horizon over shorter time scale as the average $MAPE$ at t_{L+1} was $(1.62 \pm 0.99)\%$ for WT-BiLSTM vs. $(1.20 \pm 0.80)\%$ for BiLSTM and the average $MAPE$ at t_{L+7} was $(16.03 \pm 12.08)\%$ for WT-BiLSTM vs. $(19.52 \pm 14.85)\%$ for BiLSTM. Neshat et al. [36] used evolutionary decomposition (ED) to split the original U data into sub-series, where BiLSTM model tuned with hybrid generalized normal distribution optimization (HGNDO) was used to forecast the sub-series. This model reported positive performance for 10-minute data with $MAPE$ of 7.53% vs. standalone BiLSTM (8.64%) and LSTM (8.66%). Jaseena & Kovoor [37] used empirical WT (EWT) to denoise the 10-minute U dataset into low and high-frequency sub-series. The results establish that the proposed model outperformed comparative models including standalone BiLSTM and LSTM, respectively. Review of these BiLSTM-related studies report positive results for short-term U forecasting. However, BiLSTM has twice the number of parameters as LSTM. Therefore, an efficient hyperparameter optimization tool is required for better performance.

The commonly used grid search (GS) and random search (RS) hyperparameter optimization tools are inefficient because every evaluation in their iterations is independent of prior evaluations, which increases time wastage in assessing poorly-performing regions of hyperparameter search space [38]. Bayesian optimization (BO) is proposed in [39,40] as an efficient framework for the global optimization of black-box functions. BO [41] is a sequential model-based technique that aims to locate the global optimum with the least number of trials balancing

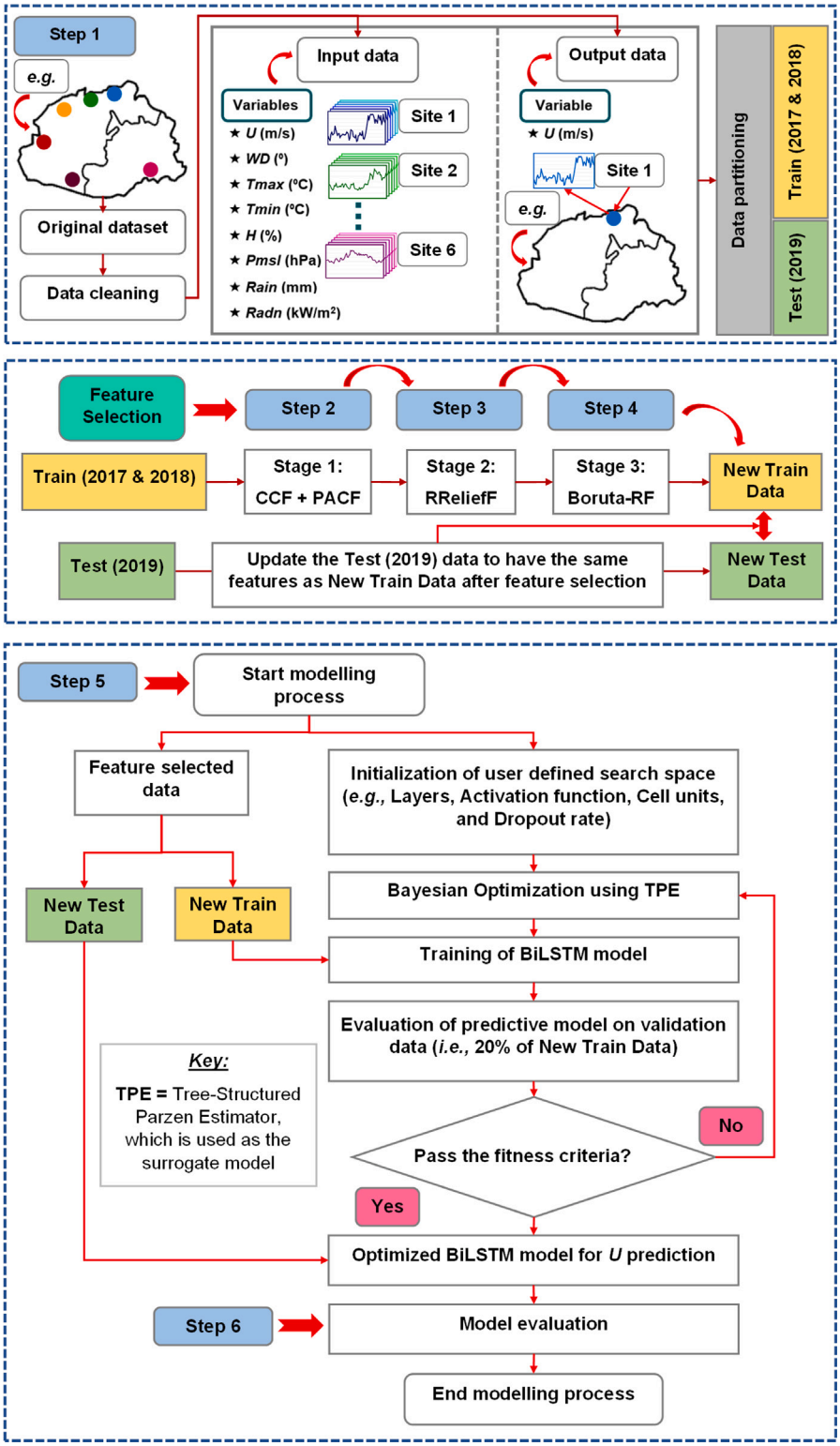


Fig. 2. Schematic of the proposed 3-phase hybrid model (i.e. FS-BO-BiLSTM) used for 10-minute ahead U (m s^{-1}) forecasting.

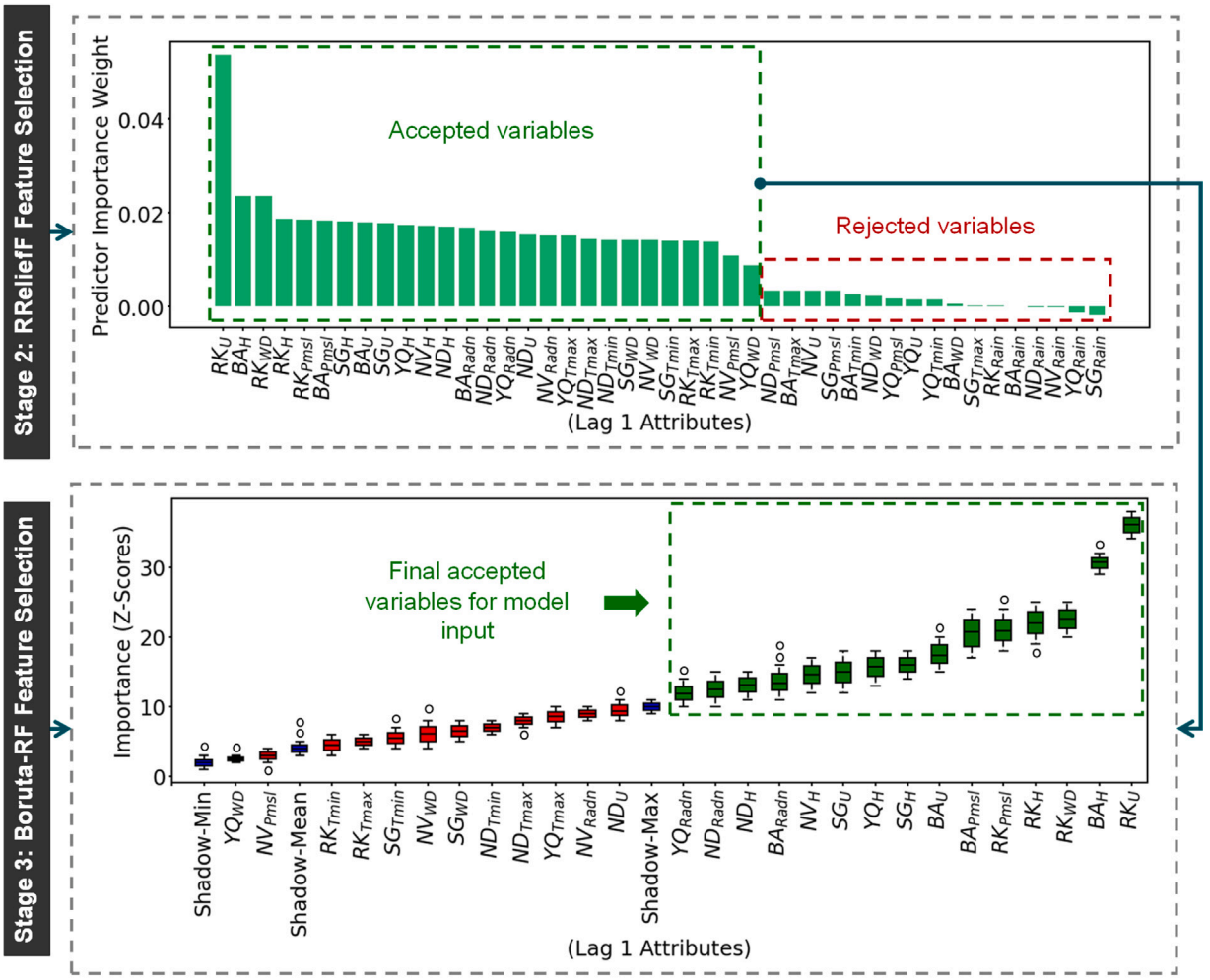


Fig. 3. Stage 2: RReliefF and **Stage 3:** Boruta-Random Forest hybridizer-based feature selection results (for **Site 1 – Rakiraki** as an example). The box plot shows the Z-scores obtained by the Boruta-RF algorithm in determining the best lagged 1 time series data for 10-minute ahead U (m s^{-1}) forecasting. "*Blue*" resembles the shadow inputs, "*green*" represents the Z-score distributions of the confirmed inputs with remarkably considerable importance and "*red*" correspond to the rejected predictor variables.

between exploration and exploitation. This helps in avoiding unnecessary trials needed to explore the search space. This optimizer is also advantageous when it comes to optimizing on-line systems. During on-line optimization of models; GS, RS and manual tuning are unable to tune various parameters simultaneously. However, BO can scale to a much larger number of parameters [42]. This is important for highly parametric models like BiLSTM, where there are frequent interactions between parameters that need optimization. BO is also preferred over evolutionary optimizers like genetic algorithm [43]. This is because the implementation of evolutionary optimizers requires the specification of various parameters like population size, number of generations, mutation rate and etc, which is difficult and relies mostly on trial and error. Various applications of BO have been explored to show its relevance. Few examples include optimization of: XGBoost and RF for accurate prediction of undrained shear strength [44], echo state network for short-term load forecasting [45] and CNN network for landslide susceptibility assessment [46]. BO has not been used to optimize BiLSTM for U forecasting till date. This research gap is narrowed in this study. Moreover, for BiLSTM architecture optimization, the type of predictors fed as model inputs matter considerably for achieving optimal results.

Predicting U is a challenging task, particularly in developing countries where the availability of quality data is a pressing issue [17]. In such areas, the available data from neighbouring reference stations can be used to predict the U of a target station. The studies that have used neighbouring reference station data as model inputs have shown positive results in terms of predictive performance. Bilgili et al. [47] estimated monthly U for eight stations in eastern Mediterranean region of Turkey using ANN. The study selected input data of neighbouring stations based on acceptable cross-correlation function (CCF) with target station data. Velázquez et al. [48] employed the hourly U and wind direction data of six sites in Canary Island (e.g. data for five sites = ANN inputs and U of sixth station = output). Currie et al. [49] evaluated the applicability of feed-forward neural networks for estimating hourly target station U using U and wind direction data from neighbouring reference sites along the south coast of Newfoundland, Canada. Deo et al. [50] devised an MLP integrated with the Firefly model optimizer for estimating monthly U of a target station using historical U of neighbouring reference stations in Iran. The authors suggested to assess the model performance by incorporating other atmospheric and land-surface data of neighbouring stations with application of an input feature selection procedure. Extensive training data were proposed as limited data can lead to significant biases during seasonal predictions. These research gaps are addressed in this study. The Use of extensive U and climate data of neighbouring stations leads to the presence of numerous predictors that require a robust feature selection (FS) strategy for dimensionality reduction.

For optimal predictive performance, a 3-stage FS strategy is proposed. For stage 1 FS, PACF and CCF is used to select the significant lag 1 variables. This method allows efficient selection of inputs [51],

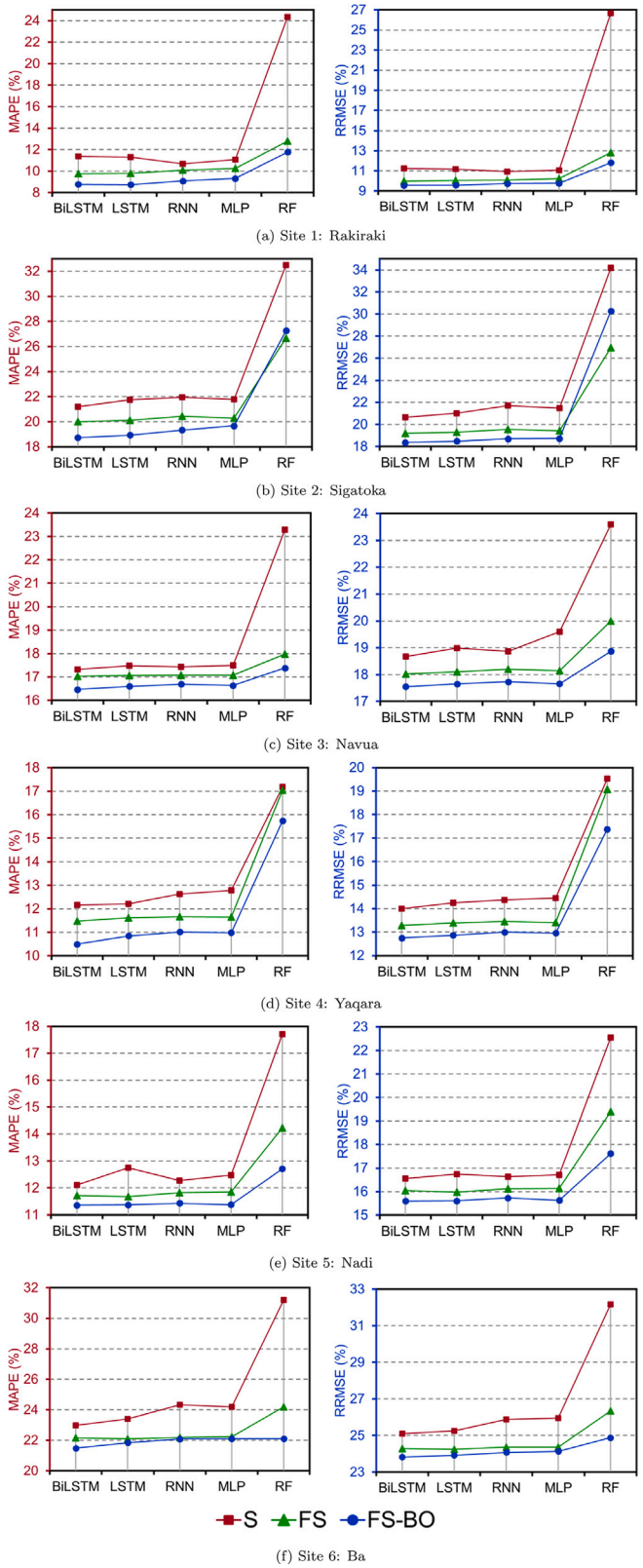


Fig. 4. Line plots of mean absolute percentage error (*MAPE* in %) and relative root mean square error (*RRMSE* in %) of the proposed vs. other comparative models for **(a) Site 1: Rakiraki, (b) Site 2: Sigatoka, (c) Site 3: Navua, (d) Site 4: Yaqara, (e) Site 5: Nadi** and **(f) Site 6: Ba** in the testing phase.

but variables with low correlation coefficient (as opposed by Bechrakis & Sparis [52]) are also selected. Thus, a hybrid filter and wrapper FS are used in stages 2 and 3, respectively. Wrapper-based Boruta-RF [53] is a highly recommended algorithm, which uses RF as the underlying tool. Boruta-RF has been utilized as an appropriate FS tool in modelling soil moisture [54], wave height [55] and streamflow [51]. Although wrapper methods can enhance the accuracy of models, they can be time-consuming. Filter methods can be used in conjunction with wrapper methods to remove redundant features with less computational power [56]. Hence, Boruta-RF is employed in the final stage once the majority inputs are efficiently removed by stage 2 filter method – Regression Relief-F (RReliefF), which is an efficient instance-based feature ranking algorithm for regression problems [57]. This 3-stage FS strategy has not been used to forecast U with the BiLSTM predictive model yet.

In this paper, a novel framework for near real-time U forecasting is proposed, based on the 3-stage FS, BO and BiLSTM model. The proposed approach is benchmarked against various standalone and hybrid ML and DL models in a real case study in Fiji. The proposed model outperforms all benchmarked models in terms of various statistical evaluation metrics. Hence, this tool can provide highly accurate U predictions that can be considered for near real-time control of wind turbines and load allocation planning. Thus, this can assist in designing a robust energy security platform in Fiji to help meet its NDC and RE targets. The remainder of the paper has been structured in the following way: Section 2 explains the theoretical details of the algorithms used. Section 3 outlines the materials and method. Section 4 presents the results and discussion. Finally, Section 5 concludes the paper.

2. Theoretical background

This section presents a description of the proposed BiLSTM network used for near real-time U forecasting. The optimization tool used for BiLSTM hyperparameter tuning and the 3-stage FS employed for dimensionality reduction is also outlined.

2.1. BiLSTM network

BiLSTM is an improved version of the LSTM model. LSTM [32] uses the concept of memory cells in the hidden layer(s) to manage long-term dependencies to mitigate the vanishing gradient problem of a simple RNN. The memory cells in the LSTM layer are updated at each time-step t in a multi-step process, and the cell states c_t and outputs h_t at time-step t in a forward pass are computed based on the following steps [58]:

i. The LSTM layer determines the information that should be disregarded in its previous cell states c_{t-1} using the forget gate (f_t). The current input data x_t, the outputs h_{t-1} of the memory cells at the previous time-step $(t-1)$, the bias terms b_f of the f_t, and weight matrices $W_{f,x}$ and $W_{f,h}$ are used to calculate the f_t value, given as:

$$f_t = \sigma\left(W_{f,x}x_t + W_{f,h}h_{t-1} + b_f\right) \tag{1}$$

ii. The LSTM layer determines the new information that needs to be kept in the cell states (c_t). The two computations to obtain the input gate (i_t) value at time-step t and the new candidate value ($\widetilde{c_t}$) are given as follows:

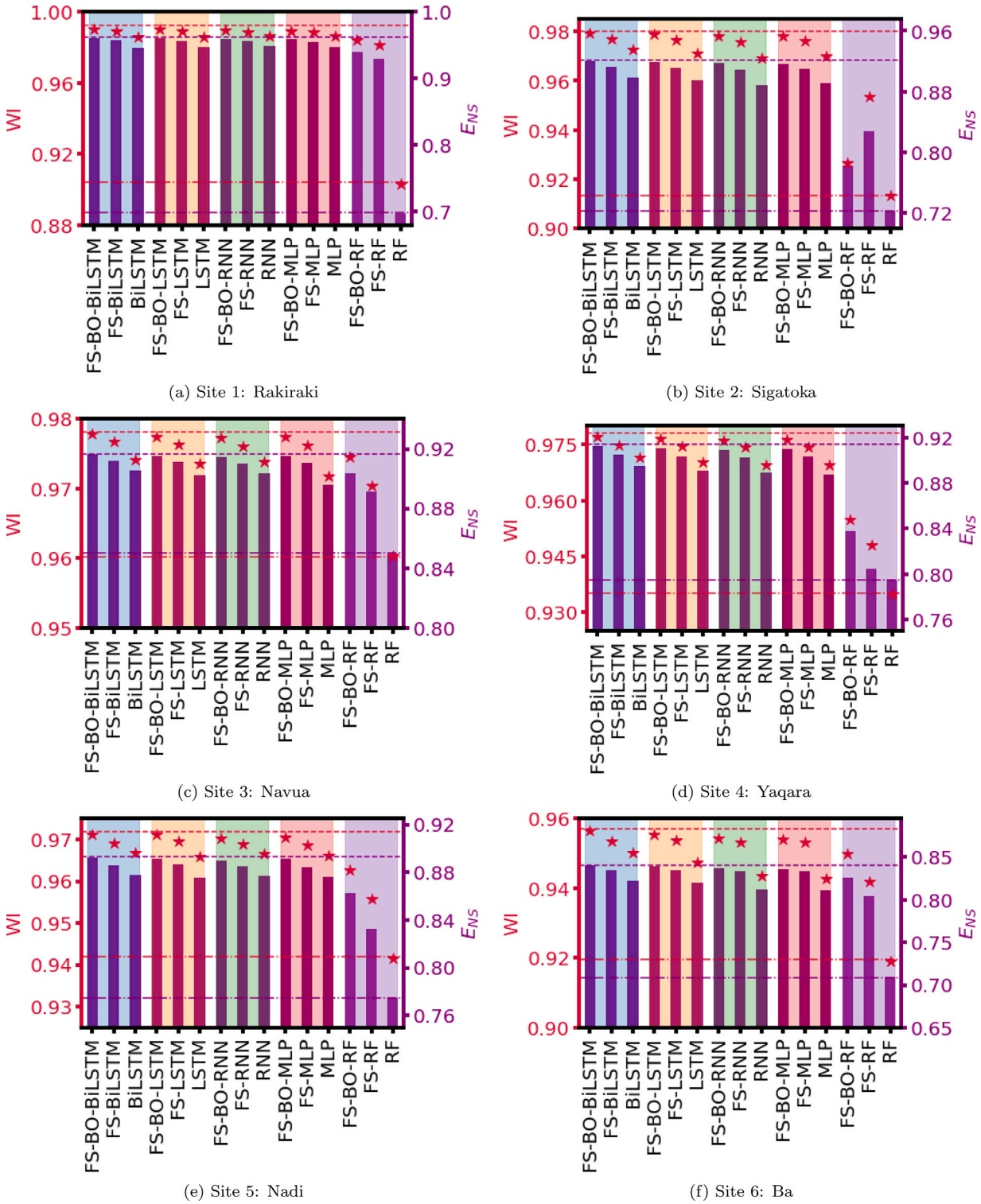
$$i_t = \sigma\left(W_{i,x}x_t + W_{i,h}h_{t-1} + b_i\right) \tag{2}$$

$$\widetilde{c_t} = \tanh\left(W_{\widetilde{c},x}x_t + W_{\widetilde{c},h}h_{t-1} + b_{\widetilde{c}}\right) \tag{3}$$

iii. The results obtained in the previous steps are used to compute new c_t as:

$$c_t = f_t \circ c_{t-1} + i_t \circ \widetilde{c_t} \tag{4}$$

Where $\circ$ depicts the Hadamard product.



(a) Site 1: Rakiraki

(b) Site 2: Sigatoka

(c) Site 3: Navua

(d) Site 4: Yaqara

(e) Site 5: Nadi

(f) Site 6: Ba

Fig. 5. Combined dot and line plots representing Willmott's Index (WI) and Nash–Sutcliffe coefficient (E_{NS}), respectively of the proposed vs. other comparative models in the testing phase for **(a) Site 1: Rakiraki, (b) Site 2: Sigatoka, (c) Site 3: Navua, (d) Site 4: Yaqara, (e) Site 5: Nadi** and **(f) Site 6: Ba**.

iv. The output h_t of the memory cells of the LSTM layer is obtained as follows:

$$o_t = \sigma\left(W_{o,x}x_t + W_{o,h}h_{t-1} + b_o\right) \tag{5}$$

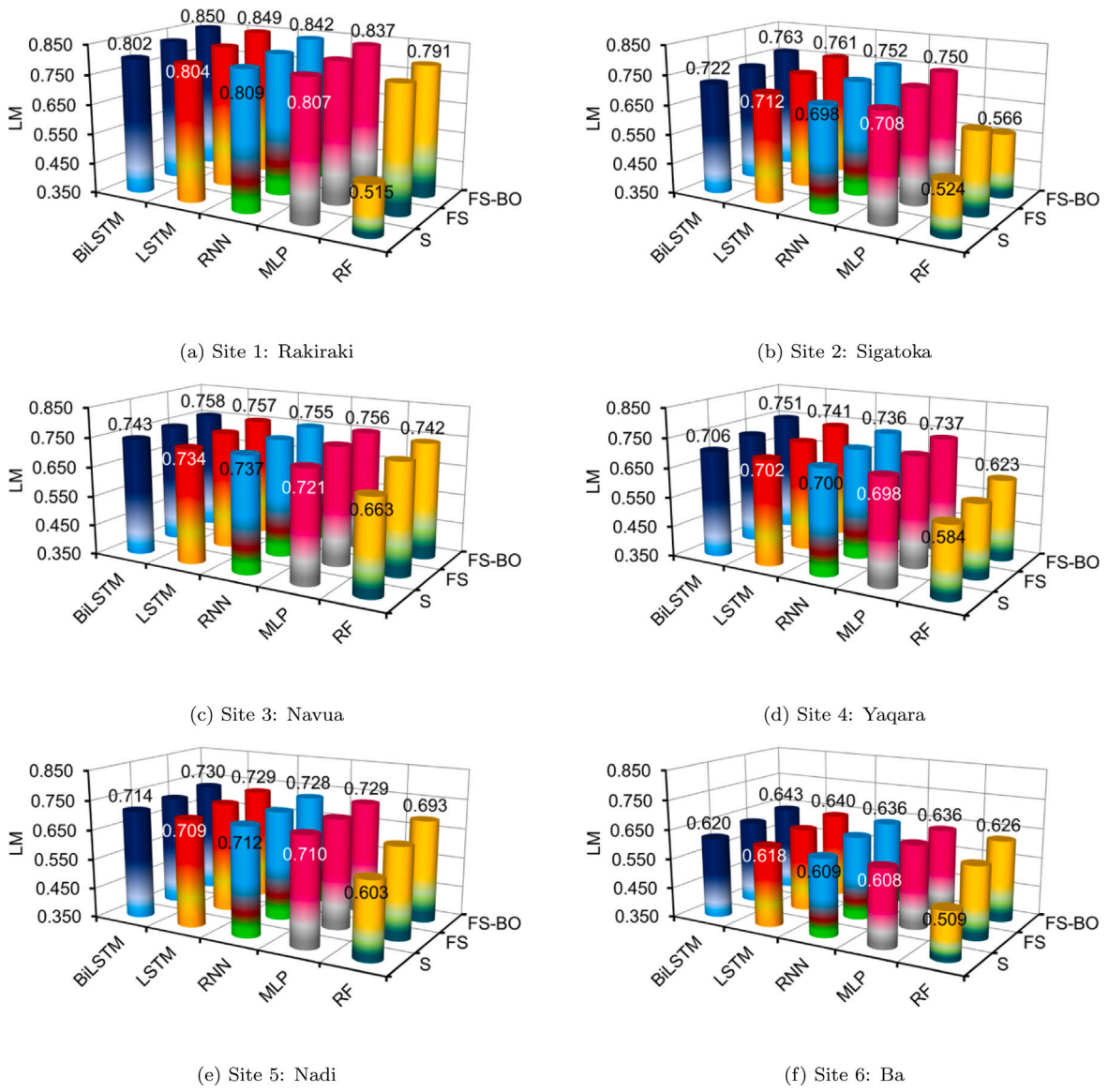
$$h_t = o_t \circ \tanh\left(c_t\right) \tag{6}$$

For BiLSTM, the architecture consists of both forward and backward LSTM layers, where the inputs from the forward and backward layers are handled simultaneously by the output layer as follows [59]:

$$\overrightarrow{h_t} = H\left(W_1 x_t + W_2\overrightarrow{h_{t-1}} + \overrightarrow{b}\right) \tag{7}$$

$$\overleftarrow{h_t} = H\left(W_3 x_t + W_5\overleftarrow{h_{t-1}} + \overleftarrow{b}\right) \tag{8}$$

$$y_t = W_4\overrightarrow{h_t} + W_6\overleftarrow{h_t} + by \tag{9}$$



(a) Site 1: Rakiraki

(b) Site 2: Sigatoka

(c) Site 3: Navua

(d) Site 4: Yaqara

(e) Site 5: Nadi

(f) Site 6: Ba

Fig. 6. Comparison of the Legates and McCabe's Index (*LM*) for the proposed 3-phase hybrid DL model against the other benchmarking models for **(a) Site 1: Rakiraki, (b) Site 2: Sigatoka, (c) Site 3: Navua, (d) Site 4: Yaqara, (e) Site 5: Nadi** and **(f) Site 6: Ba** in the testing phase.

Where $\overleftarrow{h_t}$, $\overrightarrow{h_t}$ and y_t are the vectors for backward propagation, forward propagation and an output layer, respectively; W_1, W_2, W_3, W_4, W_5 and W_6 are the corresponding weight coefficients; $\overleftarrow{b}$, $\overrightarrow{b}$ and b_y are the corresponding bias vectors.

The theoretical details of LSTM, RNN, MLP and RF methods used for model benchmarking are described elsewhere [30,32,60–62].

2.2. Bayesian optimization

Bayesian optimization (BO) [41] is based on Bayes' theorem, described as:

$$p(w|D) = \frac{p(D|w)\,p(w)}{p(D)} \tag{10}$$

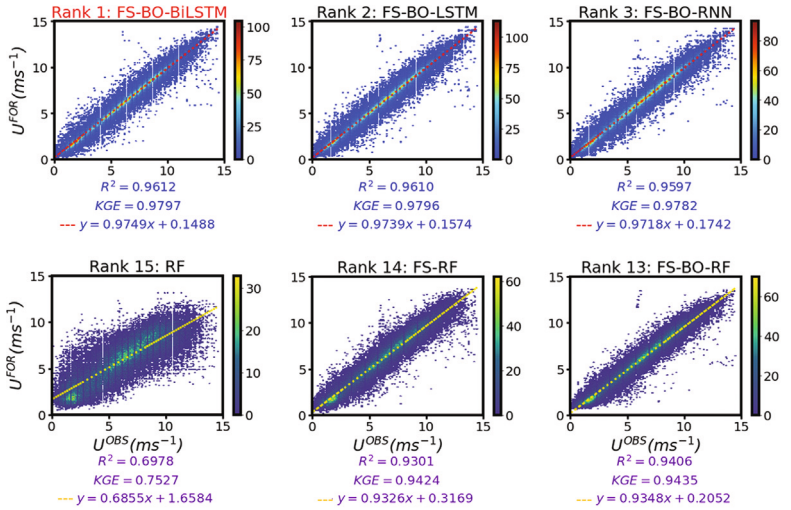
Where w indicates an unseen value, $p(w)$ refers to the preceding distribution, $p(D|w)$ indicates the probability and $p(w|D)$ denotes the posterior distribution.

BO employs two key aspects, while selecting the next hyperparameter configuration: a probabilistic surrogate model for modelling of the objective function and an acquisition function, which explores new areas in sample space and exploits areas that are already known for optimal results [63]. This makes BO more efficient over GS and RS.

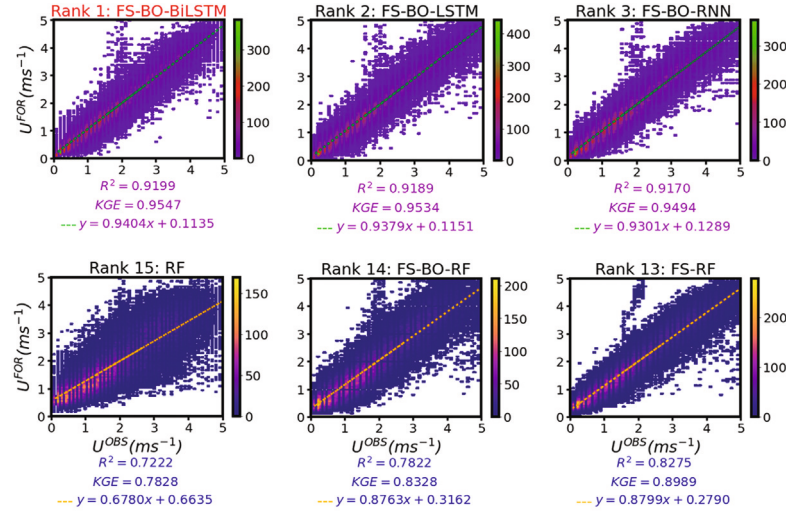
The basic steps of BO are given as [64]:

i. Build a probabilistic surrogate model of the objective function.
ii. Determine the ideal hyperparameter values on the surrogate model.
iii. Apply these selected hyperparameter values to the real objective function to assess them.
iv. Update the surrogate model with new results.
v. Repeat steps (ii–iv) until maximum number of iterations are evaluated.

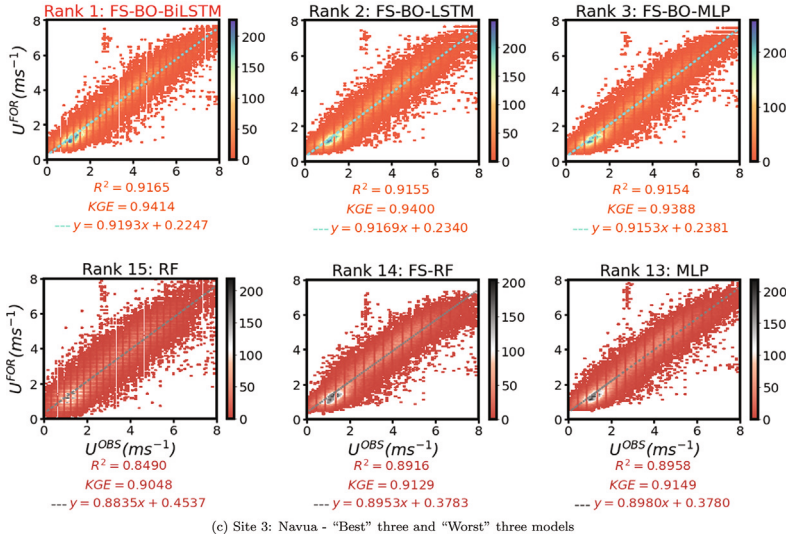
After each evaluation of the objective function, BO updates the surrogate model. Common surrogate models for BO include: Gaussian process (GP) [65], RF [66] and the tree-structured Parzen estimator (TPE) [67]. This study has employed TPE surrogate function to model the objective function, due to its lower time complexity [68].



(a) Site 1: Rakiraki - "Best" three and "Worst" three models

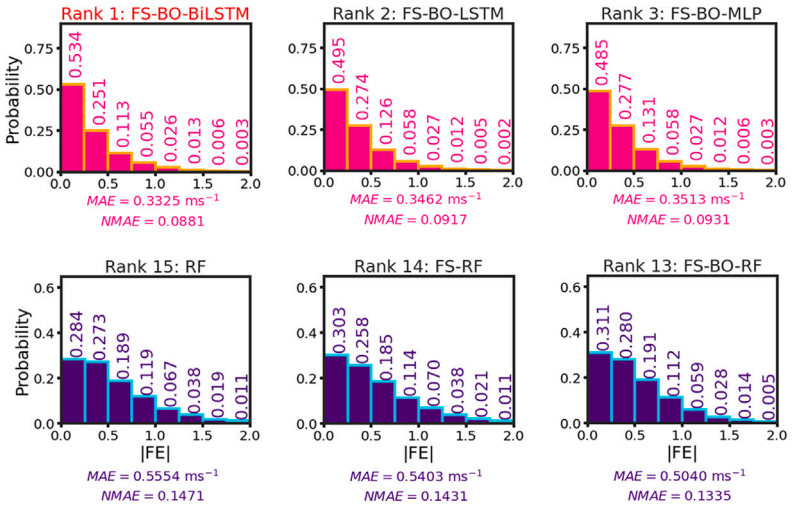


(b) Site 2: Sigatoka - "Best" three and "Worst" three models

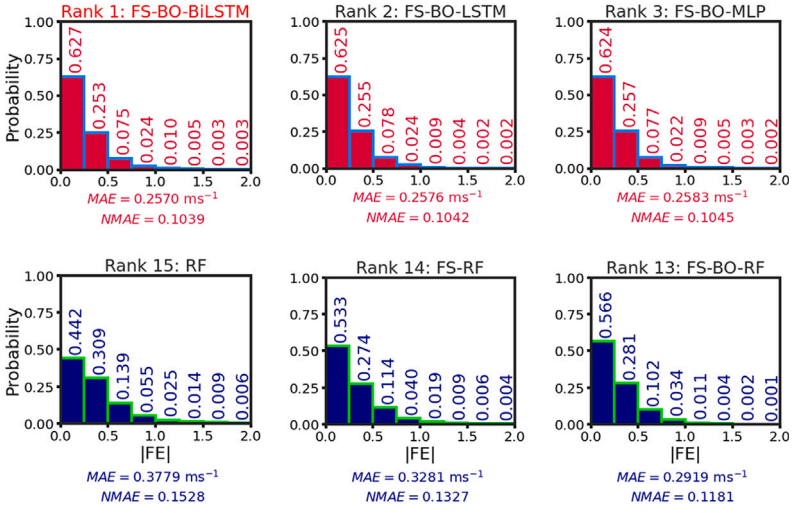


(c) Site 3: Navua - "Best" three and "Worst" three models

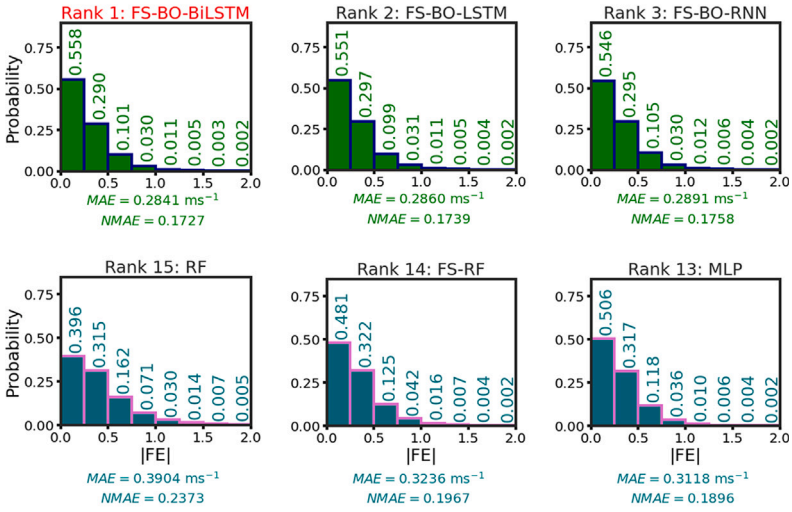
Fig. 7. Density scatter plots for forecasted U^{FOR} (m s^{-1}) vs. observed U^{OBS} (m s^{-1}) for **Site 1 – Rakiraki, Site 2 – Sigatoka** and **Site 3 – Navua** in the testing phase. Only selected best 3 and worst 3 ranked models are shown for each site. The vertical colour bars on the right indicate the frequency of samples within the binning area. For each panel, the coefficient of determination (R^2) and Kling–Gupta efficiency (KGE) are included, where highest values are considered optimal.



(a) Site 4: Yaqara - "Best" three and "Worst" three models



(b) Site 5: Nadi - "Best" three and "Worst" three models



(c) Site 6: Ba - "Best" three and "Worst" three models

Fig. 8. Histograms illustrating the probability of the absolute value of 10-minute forecasting error $|FE|$ for **Site 4 – Yaqara, Site 5 – Nadi** and **Site 6 – Ba**. Only selected best 3 and worst 3 ranked models in the testing phase are shown for each site.

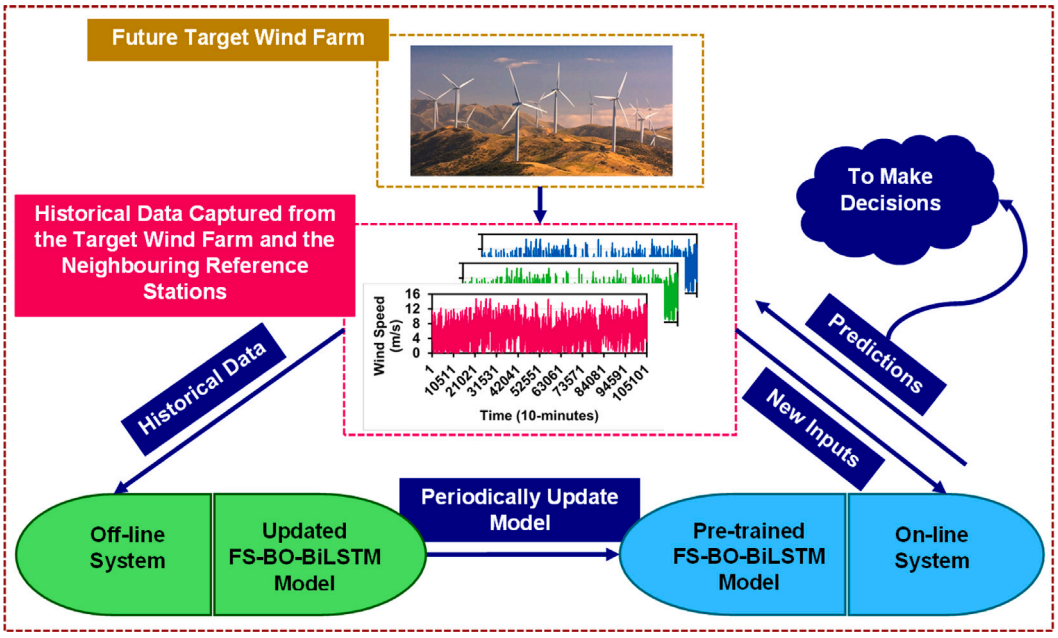


Fig. 9. Flowchart of the off-line and on-line proposed FS-BO-BiLSTM model to be used in real-life application.

2.3. Hybrid 3-stage feature selection approach

2.3.1. Stage 1: Correlated lagged input selection

In the first stage of input selection, the partial auto-correlation function (PACF) is employed to ascertain the best lag of antecedent U. The cross-correlation function (CCF) is used to retrieve the best lags of other climate variables. These are useful as the PACF quantifies the linear correlation between a time series (y_t) and a lagged form of itself $\left(y_{t_{L+k}}\right)$ with the removed linear dependence of $\left(y_{t_{L-1}}, y_{t_{L-2}}, \ldots, y_{t_{L-(k-1)}}\right)$, whereas the CCF quantifies the correlation of target output time series y with the lags of antecedent input variables $X = [X_1, X_2, \ldots, X_n]$. The variables with all lags within the 95% confidence band are considered insignificant and are disregarded. The most significant inputs obtained is then subject to stage 2 FS.

2.3.2. Stage 2: Regression Relief-F (RReliefF)

RReliefF algorithm (extended version of Relief and ReliefF [57,69]) is an instance-based filter FS method. Unlike the wrapper-based FS methods, RReliefF does not evaluate the features on a specific ML algorithm to find the optimal features. However, this efficient filter algorithm randomly selects some instances from the training data and then searches for the k-nearest neighbours from the same class. It ranks the features via allocation of weights [57] and is formulated as follows [70]:

Given a feature F with the weight W_F, where the weight is formulated based on Bayesian rule, given as such:

$$W_F = \frac{\left(p_{diffP|diffF}\right) p_{diffF}}{p_{diffP}} - \frac{\left(1 - p_{diffP|diffF}\right) p_{diffF}}{1 - p_{diffP}} \tag{11}$$

Where p is probability, $diffF$ in p_{diffF} signifies the different values of the feature F and $diffP$ in p_{diffP} denotes the different predictions. The probabilities p_{diffF} and p_{diffP} can be equated as:

$$p_{diffF} = p(diffF|NIs) \tag{12}$$

$$p_{diffP} = p(diffP|NIs) \tag{13}$$

Where NIs represent the nearest instances. The conditional probability $p_{diffP|diffF}$ in Eq. (11) is given as:

$$p_{diffP|diffF} = p(diffP|DFNIs) \tag{14}$$

Where $DFNIs$ denotes the $diffF$ and its NIs. Once the feature weight W_F is obtained using Eq. (11), the results of FS can be evaluated by feature weight ranking (e.g. a more important feature is given a greater weight). These weights (W_F) depend on the number of nearest neighbours k. Very small k values lead to unreliable estimates for noisy data, and larger k values hinder the stability of RReliefF in finding important features [57]. For reliable results, each RReliefF experiment is repeated 10 times and the weights are averaged as recommended in [71]. From the k search space of {5, 10, 15 and 20}, 10 is used in this study (also recommended in [72]). This is because $k = 10$ achieved stable and consistent results compared to other values in the search space. The selected features undergo a third round of FS using Boruta-RF.

2.3.3. Stage 3: Boruta-random forest (Boruta-RF)

Boruta is a wrapper-based FS method built around the RF algorithm [61]. The algorithm can be briefly explained as such [73]:

i. For a given set of predictor variables $(X_t \in R^n)$ and target variable $(y_t \in R)$, where n is the number of input features and $t = 1, 2, \ldots, T$ with T being the number of distinct samples; a randomly ordered duplicated variable, X'_t (i.e. shadow feature) is created for the respective input vector, X_t.
ii. The RF algorithm is utilized to predict the y_t using both X'_t and X_t.
iii. The variance importance measure (i.e. Mean Decrease Accuracy – MDA) for each X'_t and respective X_t is computed using [74]:

$$MDA = \frac{1}{m_{tree}} \sum_{m=1}^{m_{tree}} \frac{\sum_{t \in OOB} I\left(y_t = f\left(X_t\right)\right) - \sum_{t \in OOB} I\left(y_t = f\left(X_t^n\right)\right)}{|OOB|} \tag{15}$$

Where $I(\bullet)$ is the indicator function, Out-of-Bag, OOB is the prediction error of each of the training samples, $(y_t = f\left(X_t\right))$ and $(y_t = f\left(X_t^n\right))$ are predicted values before and after permutation, respectively. The m_{tree} parameter is the maximum number of trees in the forest of the RF algorithm. Higher number of trees give better performance by making the predictions more stable [53]. From the m_{tree} search space of {100, 200, 300 and 500}, 500 is used in this study due to its robust performance. This value is also recommended in studies [51,54].

Table 1
Geographic location and model input data description for the six potential wind farm sites in Viti Levu, Fiji. Note that the wind speed data are every 10-minutes over the period 01 January 2017 to 31 December 2019.

Site no.	Site name	Acronym	Latitude	Longitude	Elevation (m)	Expected data points	Data missing (%)	Training data split (%)	Testing data split (%)
1	Rakiraki	RK	17.34° S	178.22° E	8.1	157,680	3.34	66.67	33.33
2	Sigatoka	SG	18.14° S	177.50° E	6.7		7.95		
3	Navua	NV	18.22° S	178.17° E	6.2		1.84		
4	Yaqara	YQ	17.43° S	177.98° E	20.0		0.12		
5	Nadi	ND	17.76° S	177.44° E	20.7		0.67		
6	Ba	BA	17.54° S	177.67° E	10.0		17.82		

iv. The importance measure, $Z-\text{scores}$ can be calculated using MDA and the standard deviation of accuracy losses, SD as follows:

$$Z-\text{score} = \frac{MDA}{SD} \tag{16}$$

v. For input features with $Z-\text{scores} >$ maximum Z-score (MZSA) of shadow features are classed as "Confirmed", whereas input features having $Z-\text{scores} <$ MZSA are categorized as "Unimportant". To ensure that all features are categorized as either "Confirmed" or "Unimportant", and that no feature is left as "Tentative", a high iteration threshold (i.e. $maxRuns$) of 500 is selected in this study from the search space of {100, 200, 300 and 500}.

3. Materials and method

3.1. Study area and meteorological data

The case study area was Fiji, a PSIDS situated in the tropical southwest pacific between the latitudes: 15.5° S to 19.5°S and longitudes: 177°E to 179°W. The nation lies in the trade winds zone, with a predominant southeast wind direction, and has a tropical marine climate with austral summer (i.e. wet season: November–April) and austral winter (i.e. dry season: May–October) seasons. The U and other meteorological data (i.e. wind direction, minimum and maximum temperature, humidity, mean sea level pressure, solar radiation and rainfall) utilized for the six sites (Fig. 1) were provided by Fiji Meteorological Services (FMS) with 10-minute temporal resolution. The studied stations were selected based on the data availability, as the other sites (not included in this study) had over 20% of missing data. Latitude and longitude of the studied stations vary between 17.34°S to 18.22°S and 177.44°E to 178.22°E, respectively (Table 1). The shortest distance of ≈ 21.83 km was between the sites RK and YQ and the longest distance of ≈ 112.47 km was between the sites RK and SG. Table 2 shows the statistical characteristics of the data used. Relative to other sites, the lowest mean U was recorded at BA (i.e. 1.65 m s^{-1}), whereas the site with the windiest climate was RK (i.e. 5.86 m s^{-1}).

3.2. Design of the proposed U predictive model

The primary goal of this research was to devise FS-BO-BiLSTM near real-time U forecasting model. All the predictive models were developed using Google Colaboratory (Colab). Keras [75] and Tensorflow [76] libraries were used to develop the DL models. Sklearn library [77] was used for the RF model. Hyperopt [78] library was used for BO. The framework of the proposed model is illustrated in Fig. 2 and the overall process is as follows:

Step 1: The original dataset had few missing values (Table 1), which were backfilled with calendar averaged values [79]. During data cleaning, all extreme outliers were replaced with the median values of the respective time series. After data cleaning, the augmented Dickey–Fuller (ADF) test [80] was used to check the stationarity of the individual time series data used. Additionally, the Engle–Granger test [81] was conducted to test the cointegration of the target U variable and the predictors. For more details on these tests, readers are referred to Appendix B.1. After confirming the data stationarity and cointegration (Table B.1 and B.2), the primal task was to construct the data matrix. For U forecasting of site 1, the original input matrix was prepared to comprise the meteorological variables of site 1 plus U and meteorological variables of other 5 sites (Table 2). Data were partitioned into two parts, where 2017–2018 data were used for FS and model training and 2019 data were used for model testing (Table 1).

Step 2: PACF and CCF analysis was performed to select the significant lagged inputs. Figures B.1 and B.2 (in Appendix B.2) show the PACF and the CCF plots. Only 20 antecedent lags (i.e. past 200-minutes) were considered because much longer lags of U are unreliable in capturing the wind gust, and wakes in particular which are very short-lived and stochastic events. The cross-correlation coefficient (r_{cross}) of other climate indices were also seen to drop with higher antecedent lags of U (Figure B.2). The criteria for input combinations were determined by evaluating the r_{cross} of each predictor with U (for CCF) and correlation coefficient (r) of antecedent lags with U at lag 0 (for PACF). Features with all 20 lags lying within the 95% confidence interval band were insignificant. To keep the computational complexity of training DL models into consideration, only the best significant lag per variable was selected. Lag 1 (t_{L-1}) variable showed higher correlation compared to higher lags. Thus, all significant lag 1 variables were selected as inputs to forecast U at 2 steps ahead (t_{L+1}).

Step 3: The features selected in stage 1 FS, were fed to the RReliefF filter-based method. Based on its FS mechanism [82], the RReliefF algorithm (1) penalized the predictors that gave different values to neighbours with the same response values and (2) rewarded predictors that gave different values to neighbours with different response values. The penalty and rewards were weights obtained after each experimental run. Therefore, the algorithm used the intermediate weights of 10 experimental runs to compute the final weights of each predictor. Fig. 3 (stage 2) shows the weights assigned to individual predictors for RK. For all sites, RReliefF FS removed all rainfall predictors and good weights were registered for antecedent U, humidity, solar radiation and mean sea level pressure. Since many predictors were still present, a more robust wrapper-based FS was performed (i.e. stage 3).

Step 4: In the final stage, Boruta-RF computed the $Z-\text{scores}$ to exhibit the importance of the individual predictors, as shown in Fig. 3 (stage 3 for RK). All 12 variables within the shadow Min–Max ($Z-\text{score} = 2.7 - 10.1$) were rejected. The final 15 selected predictors for RK fell within the Z-score range of 12.12–37.04. This was the acceptable range since all the features within this range had Z-score values higher than the maximum Z-score of the shadow features (i.e. 10.1). The predictors confirmed through the stage 3 Boruta-RF method were selected as the final model inputs for U prediction. This method was highly reliable in removing all the irrelevant features as the variables were combined for training a RF regression model that was used to calculate the importance of each feature in form of a Z-score. Only the predictors with Z-score being higher than the maximum Z-score were selected as final inputs and the rest were disregarded. The removed features are summarized in Appendix B.2.

Table 2
Descriptive statistics of wind speed (U) (m s^{-1}) and other selected climate variables used for all six candidate sites.

Site acronym	Variable	Acronym (units)	Mean	Range	Standard deviation	Skewness	Kurtosis
RK	Target variable: Wind speed	U (m s^{-1})	5.86	14.70	2.94	0.08	−0.77
	Other predictor variables: Wind direction	*WD* (degrees)	155.81	358.00	58.23	1.61	2.43
	Max. Temp	*Tmax* (°C)	26.22	17.40	2.43	0.17	−0.29
	Min. Temp	*Tmin* (°C)	25.68	20.60	2.37	0.16	−0.20
	Humidity	*H* (%)	80.10	43.50	8.85	−0.15	−0.45
	Sea level pressure	*Pmsl* (hPa)	1010.98	22.70	3.55	−0.63	0.37
	Rainfall	*Rain* (mm)	0.04	17.50	0.35	19.64	544.49
	Solar radiation	*Radn* (kW m^{-2})	–	–	–	–	–
SG	Target variable	U (m s^{-1})	1.96	5.70	1.28	0.58	−0.46
	Other predictor variables	*WD* (degrees)	126.02	358.00	80.77	1.27	1.27
		Tmax (°C)	24.92	22.10	3.27	−0.01	−0.26
		Tmin (°C)	24.31	21.70	3.19	−0.05	−0.22
		H (%)	84.20	55.10	11.62	−0.49	−0.83
		Pmsl (hPa)	1011.89	21.50	3.89	−0.43	−0.25
		Rain (mm)	0.02	18.00	0.30	26.21	929.90
		Radn (kW m^{-2})	–	–	–	–	–
NV	Target variable	U (m s^{-1})	2.95	8.90	1.78	0.42	−0.74
	Other predictor variables	*WD* (degrees)	137.67	358.00	97.67	0.93	−0.22
		Tmax (°C)	25.17	18.90	2.81	0.09	−0.05
		Tmin (°C)	24.60	18.60	2.75	0.06	−0.03
		H (%)	83.20	48.50	10.21	−0.55	−0.70
		Pmsl (hPa)	1011.19	45.40	6.10	−1.46	3.16
		Rain (mm)	0.06	20.50	0.46	16.22	366.94
		Radn (kW m^{-2})	0.10	0.82	0.15	1.73	2.05
YQ	Target variable	U (m s^{-1})	3.95	9.10	1.79	0.29	−0.51
	Other predictor variables	*WD* (degrees)	152.13	358.00	61.69	0.97	3.33
		Tmax (°C)	26.85	21.70	2.98	0.31	−0.35
		Tmin (°C)	26.08	20.60	2.74	0.20	−0.26
		H (%)	76.17	56.40	11.93	−0.10	−0.48
		Pmsl (hPa)	1012.38	22.80	3.49	−0.49	0.17
		Rain (mm)	0.03	15.50	0.30	22.08	650.44
		Radn (kW m^{-2})	0.12	0.80	0.17	1.27	0.35
ND	Target variable	U (m s^{-1})	2.53	6.30	1.33	0.87	−0.06
	Other predictor variables	*WD* (degrees)	164.58	360.00	84.30	0.70	−0.55
		Tmax (°C)	25.90	22.60	3.24	−0.10	−0.44
		Tmin (°C)	25.55	22.30	3.10	−0.17	−0.33
		H (%)	81.83	54.80	11.57	−0.41	−0.67
		Pmsl (hPa)	1009.70	25.30	3.55	−0.75	0.66
		Rain (mm)	0.04	18.20	0.39	19.79	510.63
		Radn (kW m^{-2})	0.12	0.76	0.17	1.24	0.22
BA	Target variable	U (m s^{-1})	1.65	4.80	1.10	0.84	0.02
	Other predictor variables	*WD* (degrees)	165.05	358.00	94.24	0.37	−0.60
		Tmax (°C)	25.77	24.60	4.01	−0.12	−0.45
		Tmin (°C)	25.12	24.20	3.94	−0.14	−0.43
		H (%)	82.01	64.20	15.24	−0.57	−0.97
		Pmsl (hPa)	1010.65	26.90	3.62	−0.61	0.30
		Rain (mm)	0.03	18.50	0.37	20.75	577.39
		Radn (kW m^{-2})	0.11	0.77	0.16	1.27	0.41

Step 5: After FS using the training dataset, the data were arranged accordingly for the testing dataset. If RK site is considered, then using lag L, the input and output instances are created as follows:

$$RK_{U(input)} = \begin{bmatrix} RK_{U(t_{L-1})} & RK_{M(t_{L-1})} & \cdots & BA_{M(t_{L-1})} \\ RK_{U(t_L)} & RK_{M(t_L)} & \cdots & BA_{M(t_L)} \\ RK_{U(t_{L+1})} & RK_{M(t_{L+1})} & \cdots & BA_{M(t_{L+1})} \\ \vdots & \vdots & \vdots & \vdots \\ RK_{U(t_{N-2})} & RK_{M(t_{N-2})} & \cdots & BA_{M(t_{N-2})} \end{bmatrix} \quad (17)$$

M denotes other selected features like *WD*, *Tmax*, *Tmin*, *H*, *Pmsl*, *Rain* and *Radn*.

$$RK_{U(output)} = \begin{bmatrix} RK_{U(t_{L+1})} \\ RK_{U(t_{L+2})} \\ RK_{U(t_{L+3})} \\ \vdots \\ RK_{U(t_N)} \end{bmatrix} \quad (18)$$

The model input data (for training and testing) were normalized in the range (0 – 1) to minimize the overestimation of one variable to another [83] as X_n:

$$X_n = \frac{X_{actual} - X_{min}}{X_{max} - X_{min}} \quad (19)$$

Where X_{actual} is the respective input variables (e.g. RK_U, etc.), X_{min} and X_{max} are the minimum and maximum values of the inputs, respectively.

The training data (i.e.2017–2018) fed to the BiLSTM model was used for hyperparameter tuning, where 20% of the training data were used for validation. BO with TPE surrogate algorithm was used for hyperparameter optimization. The maximum number of evaluations was selected as 30 [84]. The hyperparameters tuned for the proposed model are furnished in Table 3. For visual reference, Figure B.3a (in Appendix B.3) reveals the selected hyperparameters for the proposed BiLSTM model for RK. Figures B.3b–B.3e present the hyperparameters

Table 3
Parameter search space assigned for Bayesian optimization (BO) used for developing the hybrid 3-phase models and default search space used in the development of hybrid 2-phase and standalone models.

Model type	Model notations	Param name	Param search space
3-Phase hybrid models	FS-BO-BiLSTM	Layers	Quantized uniform {1, 6, 1}
		Hidden units	Choice {64, 126, 256}
		Dropout rate	Uniform {0.1, 0.2}
		Activation function	Choice {ReLU, tanh}
		Batch size	1024
		Epochs	128
	FS-BO-LSTM	Layers	Quantized uniform {1, 6, 1}
		Hidden units	Choice {64, 126, 256}
		Dropout rate	Uniform {0.1, 0.2}
		Activation function	Choice {ReLU, tanh}
		Batch size	1024
		Epochs	128
	FS-BO-RNN	Layers	Quantized uniform {1, 6, 1}
		Hidden units	Choice {64, 126, 256}
		Dropout rate	Uniform {0.1, 0.2}
		Activation function	Choice {ReLU, tanh}
		Batch size	1024
		Epochs	128
	FS-BO-MLP	Layers	Quantized uniform {1, 6, 1}
		Hidden units	Choice {64, 126, 256}
		Dropout rate	Uniform {0.1, 0.2}
		Activation function	Choice {ReLU, tanh}
		Batch size	1024
		Epochs	128
	FS-BO-RF	Number of trees in the forest [n_estimators]	Choice {300, 400}
		Maximum features to consider for best split [max_features]	Choice {'auto', 'sqrt', 'log2', None}
		Minimum number of samples required to be at leaf node [min_samples_leaf]	Uniform {0, 0.5}
		Minimum samples to split internal node [min_samples_split]	Uniform {0, 1}
2-Phase Hybrid and Standalone Models	FS-BiLSTM & BiLSTM	Layers	1 hidden
		Hidden units	128
		Dropout rate	0.05
		Activation function	ReLU
		Batch size	1024
		Epochs	128
	FS-LSTM & LSTM	Layers	1 hidden
		Hidden units	128
		Dropout rate	0.05
		Activation function	ReLU
		Batch size	1024
		Epochs	128
	FS-RNN & RNN	Layers	1 hidden
		Hidden units	128
		Dropout rate	0.05
		Activation function	ReLU
		Batch size	1024
		Epochs	128
	FS-MLP & MLP	Layers	1 hidden
		Hidden units	128
		Dropout rate	0.05
		Activation function	ReLU
		Batch size	1024
		Epochs	128
	FS-RF & RF	n_estimators	100
		max_features	'auto'
		min_samples_leaf	1
		min_samples_split	2

selected for comparative models. The dense regions of the contour plots represent that out of the 30 evaluations of BO, the majority of samples were drawn to give the respective hyperparameter(s) on the *y*-axis. The horizontal line shows the selected hyperparameter and the vertical line shows the iteration at which it was selected. In addition, Adaptive Moment Estimation (Adam) optimizer was used to minimize the loss function during model training and validation.

Step 6: Finally, various performance metrics were used to evaluate the accuracy of the proposed method against alternative comparative models (i.e. standalone and 2-phase BiLSTM, LSTM, RNN, MLP, RF and 3-phase LSTM, RNN, MLP and RF).

3.3. Model performance criteria

For evaluation of the proposed FS-BO-BiLSTM model against others, two categories of statistical indicators were used: Class A (ideal value = 1) and Class B (ideal value = 0). Six Class A (Coefficient of Determination (R^2), Pearson's Correlation Coefficient (r), Willmott's Index (WI), Nash–Sutcliffe Efficiency (E_{NS}), Legates–McCabe's Index (LM) and Kling–Gupta Efficiency (KGE)) and six Class B (Mean Absolute Error (MAE; m s^{-1}), normalized MAE ($NMAE$), Root Mean Square Error ($RMSE$; m s^{-1}), normalized $RMSE$ ($NRMSE$), Relative $RMSE$ ($RRMSE$; %) and Mean Absolute Percentage Error ($MAPE$; %)) indicators were used. The Class A indicators were used to test the variance of predicted and actual U values, while the Class B error indicators were used to study model bias. Since both bias and variance are components of reducible error, the models were compared in terms of achieving low bias and variance. The rationale behind selection of these metrics and the mathematical notations are given in Appendix B.4.

3.3.1. Global performance indicator (GPI)

The global performance indicator (GPI) [85] was used to overcome the issue of comparing a large variety of models to conclude the overall performance. Mathematically, for ith model, the GPI is computed as:

$$GPI = \sum_{j=1}^{N} \alpha_j \left(\overline{y_j} - y_{ij}\right), \ (-\infty < GPI < +\infty) \tag{20}$$

Where a larger GPI indicates a more accurate model, N is the total number of statistical indicators used (i.e. 12 for this study), α_j is −1 for Class A indicators and +1 for Class B indicators, y_{ij} is the scaled value of indicator (j) for model (i) and $\overline{y_j}$ is the median of scaled values of indicator (j). The indicator y_{ij} was scaled in range (0 – 1) using:

$$y_{ij} = \frac{x_{ij} - \min\left(x_{ij}\right)}{\max\left(x_{ij}\right) - \min\left(x_{ij}\right)} \tag{21}$$

Where x_{ij} is the value of indicator (j) for model (i), $\min(x_{ij})$ and $\max(x_{ij})$ are the minimum and maximum values, respectively of indicator (j) for model (i). The $\overline{y_j}$ of all the models considered for GPI_j calculation of the Class A indicators (ideal value = 1) is always lower than the y_{ij} of the proposed model (i.e. 1). Since a higher positive GPI is optimal, an α_j value of −1 was used to make the calculation positive for Class A indicators. The $\overline{y_j}$ of the Class B indicators (ideal value = 0) is always higher than the y_{ij} of the proposed model (i.e. 0). Hence, an α_j value of +1 was used to keep the calculation unchanged for Class B indicators.

4. Results and discussion

In this section, empirical results of modelling experiments conducted are presented. The initial evaluation of the models are based on r and the non-normalized MAE and $RMSE$ measures (Table 4). The metric r is a non-dimensional and an absolute measure, which conjectures the strength and direction of linear association between the observed U^{OBS} and forecasted U^{FOR} values. The measures MAE and $RMSE$ are based on aggregation of residuals of U^{OBS} and U^{FOR} values. Table 4 shows that the proposed model obtained the highest r (0.917–0.981), and the lowest MAE (0.238–0.369) and $RMSE$ (0.343–0.581) for all sites. The r measure for three sites (SG, NV and ND) and $RMSE$ for one site (ND) showed comparable results between the objective model and FS-BO-LSTM. Conversely, the MAE at these sites were lower for the proposed model compared to FS-BO-LSTM. This points out the supremacy of BiLSTM, where two LSTMs are applied to the input data in both backward and forward directions. This dual flow of information facilitated efficient learning of long-term dependencies between the features and the target U. However, it should be noted that nearly all DL models perform exceptionally well with large datasets. The extensive dataset used in this study may have also allowed the LSTM models to effectively learn the important features in just a single flow of information for the three sites based on r and $RMSE$. Even so, r is scale and offset invariant and can even give larger values to mediocre models [86]. Also, the squaring of $RMSE$ term induces a bias towards high U values [87]. Thus, these two measures are unreliable at times, especially when comparing models with similar underlying structures. The absolute computation in MAE reduces the biases, making it reliable over r and $RMSE$ [87].

The U range of the six study sites varied (Table 2). This made MAE and $RMSE$ measures unreliable in assessing model performance across geographically disparate sites since both are expressed in absolute units [88]. For this reason, the normalized and the relative error measures were used to assess model bias. The normalized measures presented in Table 5 show that the proposed model registered the lowest $NMAE$ and $NRMSE$ at all sites. The $NMAE$ and $NRMSE$ measures allow comparison of model bias amongst different sites. For instance, the $NMAE$ of the proposed model for BA (0.172) and RK (0.060) revealed that the model bias was lower for the RK site. This further informs that RK site had features selected via the 3-stage FS that were more physically dependent to the forecasted U, making it a potentially good site for future wind farm commissioning. The relative measures $RRMSE$ and $MAPE$ are easier to interpret as percentage criteria are used to categorize the models as "Excellent" ($RRMSE/MAPE$ < 10%), "Good" (10% < $RRMSE/MAPE$ < 20%), "Fair" (20% < $RRMSE/MAPE$ < 30%) and "Poor" ($RRMSE/MAPE \geq 30\%$) [89]. Based on this, the proposed model was categorized as: "Excellent" for RK, "Good" for SG, NV, YQ and ND, and "Fair" for BA (Fig. 4).

As revealed in Fig. 4, the proposed FS-BO-BiLSTM model for all sites performed comparatively better over all the benchmarked models. When compared with the lowest performing standalone RF, the proposed model showed percentage decrease in $RRMSE|MAPE$ as follows: RK (−64.16%| − 64.09%), SG (−46.29%| − 42.39%), NV (−25.63%| − 29.30%), YQ (−34.70%| − 38.88%), ND (−30.80%| − 35.86%) and BA (−25.91%| − 31.14%). This clarifies the potential merits of the objective model, where FS and BO helped improve the performance of BiLSTM. The standalone BiLSTM also outperformed its standalone counterparts due to its ability to aptly capture the stochastic variation of U data via its bidirectional processing feature [90]. Conversely, the RF model performed poorly. This is because when RF is tasked with predicting the values not seen in the training dataset, it will predict an average of the values seen previously [61]. The range of training and testing datasets for the RK site were 13.23 and 14.70 m s^{-1}, respectively. There were 1,360 data points in the testing dataset with values > 13.23 m s^{-1}. For these values, the RF model estimated the U as ≈ 5.74 m s^{-1} (i.e. average U of training data). Also, RF performs poorly with sparse features (e.g. *Radn* and *Rain*). These features altered the performance of RF and its ability to make accurate predictions [61]. Based on these drawbacks of RF, the percentage error yielded was very high compared to the proposed model.

The E_{NS} [91] is a dimensionless and scaled variant of MSE. It is a commonly used criterion for model evaluation, but it exaggerates the larger values of outliers while overlooking the smaller ones [92]. WI [93,94] was introduced to address this issue of E_{NS} by considering

Table 4
Evaluation of proposed (i.e. FS-BO-BiLSTM) vs. all other comparative models in the testing phase, using the r = Pearson's correlation coefficient, *MAE* (m s^{-1}) = mean absolute error and *RMSE* (m s^{-1}) = root mean square error.

Model type	Model notations	Sites evaluated during testing phase																	
		RK			SG			NV			YQ			ND			BA		
		r	*MAE* (m s^{-1})	*RMSE* (m s^{-1})	*r*	*MAE* (m s^{-1})	*RMSE* (m s^{-1})	*r*	*MAE* (m s^{-1})	*RMSE* (m s^{-1})	*r*	*MAE* (m s^{-1})	*RMSE* (m s^{-1})	*r*	*MAE* (m s^{-1})	*RMSE* (m s^{-1})	*r*	*MAE* (m s^{-1})	*RMSE* (m s^{-1})
3-Phase hybrid models	FS-BO-BiLSTM	0.981	0.369	0.581	0.959	0.238	0.343	0.957	0.351	0.493	0.956	0.333	0.482	0.945	0.257	0.386	0.917	0.284	0.392
	FS-BO-LSTM	0.980	0.372	0.582	0.959	0.240	0.345	0.957	0.353	0.496	0.955	0.346	0.486	0.945	0.258	0.386	0.916	0.286	0.394
	FS-BO-RNN	0.980	0.390	0.591	0.958	0.251	0.349	0.956	0.355	0.498	0.954	0.352	0.491	0.943	0.259	0.389	0.915	0.289	0.396
	FS-BO-MLP	0.980	0.402	0.594	0.958	0.252	0.350	0.957	0.354	0.496	0.954	0.351	0.489	0.944	0.258	0.387	0.915	0.289	0.397
	FS-BO-RF	0.970	0.515	0.719	0.897	0.436	0.565	0.951	0.374	0.530	0.915	0.504	0.656	0.929	0.292	0.436	0.909	0.297	0.409
2-Phase hybrid models	FS-BiLSTM	0.979	0.422	0.607	0.955	0.259	0.358	0.955	0.363	0.506	0.952	0.369	0.502	0.941	0.267	0.397	0.914	0.292	0.399
	FS-LSTM	0.978	0.424	0.611	0.955	0.263	0.360	0.955	0.364	0.509	0.951	0.373	0.506	0.942	0.265	0.395	0.914	0.292	0.399
	FS-RNN	0.978	0.435	0.614	0.954	0.267	0.365	0.954	0.365	0.511	0.950	0.376	0.508	0.940	0.269	0.399	0.913	0.293	0.401
	FS-MLP	0.978	0.444	0.622	0.954	0.265	0.363	0.954	0.365	0.510	0.951	0.374	0.506	0.940	0.270	0.399	0.913	0.294	0.401
	FS-RF	0.965	0.563	0.779	0.912	0.375	0.503	0.945	0.400	0.562	0.901	0.540	0.720	0.915	0.328	0.480	0.898	0.324	0.433
Standalone Models	BiLSTM	0.974	0.488	0.684	0.948	0.280	0.385	0.952	0.373	0.525	0.946	0.393	0.529	0.937	0.278	0.410	0.908	0.304	0.413
	LSTM	0.974	0.482	0.679	0.947	0.290	0.392	0.952	0.386	0.533	0.944	0.398	0.538	0.936	0.279	0.414	0.907	0.304	0.416
	RNN	0.975	0.470	0.664	0.944	0.304	0.405	0.952	0.382	0.530	0.944	0.401	0.542	0.937	0.278	0.411	0.906	0.311	0.426
	MLP	0.974	0.475	0.673	0.946	0.293	0.401	0.951	0.404	0.551	0.943	0.403	0.546	0.936	0.278	0.413	0.905	0.312	0.427
	RF	0.840	1.196	1.620	0.850	0.479	0.638	0.924	0.488	0.663	0.895	0.555	0.737	0.889	0.378	0.558	0.850	0.390	0.529

Table 5
Evaluation of proposed (i.e. FS-BO-BiLSTM) vs. all other comparative models in the testing phase, using the *NMAE* = normalized mean absolute error and *NRMSE* = normalized root mean square error.

Model type	Model notations	Sites evaluated during testing phase											
		RK		SG		NV		YQ		ND		BA	
		NMAE	*NRMSE*	*NMAE*	*NRMSE*	*NMAE*	*NRMSE*	*NMAE*	*NRMSE*	*NMAE*	*NRMSE*	*NMAE*	*NRMSE*
3-Phase hybrid models	FS-BO-BiLSTM	0.060	0.095	0.127	0.183	0.124	0.175	0.088	0.127	0.103	0.156	0.172	0.238
	FS-BO-LSTM	0.061	0.096	0.129	0.185	0.126	0.177	0.092	0.129	0.104	0.156	0.174	0.239
	FS-BO-RNN	0.064	0.097	0.134	0.187	0.126	0.177	0.093	0.130	0.105	0.157	0.176	0.241
	FS-BO-MLP	0.066	0.098	0.135	0.187	0.126	0.177	0.093	0.130	0.105	0.156	0.176	0.241
	FS-BO-RF	0.085	0.118	0.234	0.303	0.133	0.189	0.134	0.174	0.118	0.176	0.181	0.249
2-Phase hybrid models	FS-BiLSTM	0.069	0.099	0.139	0.191	0.129	0.180	0.097	0.132	0.108	0.161	0.178	0.243
	FS-LSTM	0.070	0.101	0.141	0.193	0.130	0.181	0.099	0.134	0.107	0.160	0.178	0.242
	FS-RNN	0.072	0.101	0.143	0.196	0.130	0.182	0.100	0.135	0.109	0.161	0.178	0.244
	FS-MLP	0.073	0.102	0.142	0.194	0.130	0.182	0.099	0.134	0.109	0.162	0.179	0.244
	FS-RF	0.093	0.128	0.201	0.269	0.143	0.200	0.143	0.191	0.133	0.194	0.197	0.264
Standalone Models	BiLSTM	0.080	0.113	0.150	0.207	0.133	0.187	0.104	0.140	0.112	0.166	0.185	0.251
	LSTM	0.079	0.112	0.155	0.210	0.137	0.190	0.105	0.143	0.113	0.167	0.185	0.253
	RNN	0.077	0.109	0.163	0.217	0.136	0.189	0.106	0.144	0.113	0.166	0.189	0.259
	MLP	0.078	0.111	0.157	0.215	0.144	0.196	0.107	0.145	0.113	0.167	0.190	0.260
	RF	0.197	0.267	0.257	0.342	0.174	0.236	0.147	0.195	0.153	0.226	0.237	0.322

the MSE ratio instead of the differences. The results of E_{NS} and WI are presented in Fig. 5. It is evident that the Bayesian optimized BiLSTM models integrated with the 3-stage FS demonstrated a dramatic improvement, where both the E_{NS} and WI were higher than the standalone models. The proposed vs. standalone BiLSTM performance in terms of $E_{NS}|WI$ were as follows: RK (0.961 vs. 0.946|0.991 vs. 0.986), SG (0.920 vs. 0.899|0.979 vs. 0.973), NV (0.917 vs. 0.905|0.978 vs. 0.974), YQ (0.912 vs. 0.895|0.977 vs. 0.972), ND (0.892 vs. 0.878|0.971 vs. 0.967) and BA (0.840 vs. 0.823|0.957 vs. 0.950). This validates that the hybridization based on BO and 3-stage FS enhanced the performance of the standalone BiLSTM for all sites. However, WI and E_{NS} are both oversensitive to peak residual values and registers relatively high values due to the squared values of residual terms [92,93].

LM [92] is not overestimated as it considers absolute values and assigns appropriate weights to errors and discrepancies, being preferred over WI and E_{NS}. The LM results (Fig. 6) concertedly revealed the supremacy in the performance of the proposed FS-BO-BiLSTM model at all six sites over other models. The proposed model at all the sites ensued $LM > 0.643$, with the maximum magnitude recorded at RK ($LM = 0.85$). Considering all six sites, the average LM index of the proposed model showed an improvement of: 4.42%, 5.05%, 5.39%, 5.71% and 32.30% over the standalone BiLSTM, LSTM, RNN, MLP and RF models, respectively.

To further explore the suitability of FS-BO-BiLSTM in U forecasting, diagnostic plots were used. Fig. 7 shows the density scatter plots of the U^{OBS} and U^{FOR} for sites 1 (RK), 2 (SG) and 3 (NV). For efficient comparison, only the "Best" and "Worst" three ranked models were plotted. In each panel, an R^2 measure is used to evaluate the goodness-of-fit between the U^{OBS} and U^{FOR}. The plots (Fig. 7) also incorporated the KGE measure. A perfectly fitted model is ought to have an ideal value of +1 for R^2 and KGE. The magnitudes of R^2 registered for the optimal FS-BO-BiLSTM models were close to unity, which were 0.9612 for RK, 0.9199 for SG and 0.9165 for NV. The KGE values were also in line with the R^2 values, registering 0.9797 for RK, 0.9547 for SG and 0.9414 for NV. These two measures clearly outlined that the proposed model achieved lower variance between the U^{OBS} and U^{FOR} compared to FS-BO-LSTM.

Analysis of the spread of forecasting errors (FE) was also implemented to evaluate the competence of the proposed hybrid FS-BO-BiLSTM model. Fig. 8 shows plots of histogram for sites 4 (YQ), 5 (ND) and 6 (BA) revealing the probability distribution of $|FE|$ computed in error brackets of 0.25 step-sizes. Only the "Best" and "Worst" three ranked models were plotted for better comparative evaluation. The "Best" three ranked models for all three sites registered very small spreads in forecasting error that were closer to 0. A closer examination of the probability of $|FE|$ for all models further revealed the strength

Table 6
Global Performance Index (*GPI*) values to rank models. The models are ranked as: (*a*) in relation to the phases with ranks 1–3 (i.e. standalone, 2-phase and 3-phase), (*b*) in relation to all 5 models per phase with ranks 1–5 and (*c*) in relation to all the 15 models generated per site (ranks 1–15). Colours "*green*" represents higher ranks, while "*red*" denotes lower performing models.

Model Type	Model Notations	Sites Evaluated for Model Ranking During Testing Phase																							
		RK				SG				NV				YQ				ND				BA			
		GPI	Rank			*GPI*	Rank			*GPI*	Rank			*GPI*	Rank			*GPI*	Rank			*GPI*	Rank		
			a	*b*	*c*		*a*	*b*	*c*		*a*	*b*	*c*		*a*	*b*	*c*		*a*	*b*	*c*		*a*	*b*	*c*
3-Phase Hybrid Models	FS-BO-BiLSTM	2.841	1	1	1	0.739	1	1	1	0.547	1	1	1	1.379	1	1	1	0.744	1	1	1	-1.884	1	1	1
	FS-BO-LSTM	2.820	1	2	2	0.685	1	2	2	0.502	1	2	2	1.254	1	2	2	0.731	1	2	2	-1.965	1	2	2
	FS-BO-RNN	2.707	1	3	3	0.534	1	3	3	0.464	1	4	4	1.171	1	4	4	0.676	1	4	4	-2.058	1	3	3
	FS-BO-MLP	2.644	1	4	4	0.494	1	4	4	0.489	1	3	3	1.195	1	3	3	0.713	1	3	3	-2.082	1	4	4
	FS-BO-RF	1.581	1	5	13	-5.079	2	5	14	0.077	1	5	9	-1.504	1	5	13	-0.236	1	5	13	-2.492	1	5	9
2-Phase Hybrid Models	FS-BiLSTM	2.509	2	1	5	0.311	2	1	5	0.358	2	1	5	0.955	2	1	5	0.482	2	2	6	-2.183	2	2	6
	FS-LSTM	2.481	2	2	6	0.245	2	2	6	0.312	2	2	6	0.894	2	2	6	0.525	2	1	5	-2.162	2	1	5
	FS-RNN	2.412	2	3	7	0.137	2	4	8	0.285	2	4	8	0.858	2	4	8	0.435	2	3	7	-2.208	2	3	7
	FS-MLP	2.346	2	4	8	0.184	2	3	7	0.299	2	3	7	0.884	2	3	7	0.421	2	4	8	-2.215	2	4	8
	FS-RF	1.082	2	5	14	-2.892	1	5	13	-0.484	2	5	14	-2.307	2	5	14	-1.119	2	5	14	-3.360	2	5	14
Standalone Models	BiLSTM	1.782	3	4	12	-0.293	3	1	9	0.044	3	1	10	0.522	3	1	9	0.282	3	1	9	-2.568	3	1	10
	LSTM	1.828	3	3	11	-0.540	3	2	10	-0.114	3	3	12	0.391	3	2	10	0.176	3	4	12	-2.772	3	2	11
	RNN	1.990	3	1	9	-0.854	3	4	12	-0.060	3	2	11	0.300	3	3	11	0.258	3	2	10	-3.147	3	3	12
	MLP	1.919	3	2	10	-0.680	3	3	11	-0.381	3	4	13	0.244	3	4	12	0.235	3	3	11	-3.200	3	4	13
	RF	-7.480	3	5	15	-6.699	3	5	15	-2.064	3	5	15	-3.147	3	5	15	-2.778	3	5	15	-6.040	3	5	15

Table 7
Seasonal evaluation of proposed (i.e. FS-BO-BiLSTM) vs. all other comparative models in the testing phase, using the *RRMSE* (%) – relative root mean square error.

Model type	Model notations	RRMSE (%) of all Sites evaluated during testing phase											
		RK		SG		NV		YQ		ND		BA	
		Seasons		Seasons		Seasons		Seasons		Seasons		Seasons	
		Dry	Wet	Dry	Wet	Dry	Wet	Dry	Wet	Dry	Wet	Dry	Wet
3-Phase hybrid models	FS-BO-BiLSTM	7.988	11.616	17.364	19.384	17.125	17.982	11.151	14.789	14.452	16.610	22.538	25.232
	FS-BO-LSTM	8.016	11.709	17.462	19.502	17.207	18.088	11.221	14.950	14.519	16.660	22.665	25.291
	FS-BO-RNN	8.173	11.781	17.684	19.713	17.283	18.207	11.403	15.046	14.585	16.840	22.707	25.559
	FS-BO-MLP	8.202	11.928	17.726	19.760	17.221	18.137	11.396	14.965	14.582	16.775	22.830	25.568
	FS-BO-RF	10.085	14.155	28.764	31.817	18.380	19.380	15.394	19.924	16.318	18.888	23.373	26.516
2-Phase hybrid models	FS-BiLSTM	8.389	12.122	18.135	20.262	17.592	18.471	11.650	15.373	14.886	17.104	22.914	25.751
	FS-LSTM	8.399	12.247	18.280	20.352	17.650	18.550	11.743	15.426	14.845	17.080	22.910	25.688
	FS-RNN	8.512	12.289	18.410	20.713	17.826	18.569	11.798	15.584	14.941	17.301	23.010	25.760
	FS-MLP	8.574	12.454	18.377	20.513	17.729	18.562	11.771	15.521	14.973	17.311	23.092	25.822
	FS-RF	10.813	15.521	25.845	28.070	19.683	20.246	17.235	21.418	17.673	21.050	24.998	27.819
Standalone models	BiLSTM	9.356	13.785	19.442	21.883	18.331	19.000	12.261	16.214	15.273	17.818	23.571	26.734
	LSTM	9.248	13.720	19.764	22.273	18.620	19.334	12.417	16.566	15.490	17.975	23.918	26.767
	RNN	9.059	13.422	20.528	22.964	18.502	19.197	12.529	16.689	15.402	17.844	24.476	27.380
	MLP	9.168	13.625	19.979	22.937	19.105	20.107	12.666	16.718	15.420	17.955	24.639	27.404
	RF	23.981	30.255	33.438	34.986	22.514	24.876	17.087	22.639	19.940	24.986	30.368	34.108

Note: Fiji has two seasons — Dry between May to October and Wet between November to April.

and suitability of the proposed model, which registered the greatest percentage of $|FE|$ of 53.4%, 62.7% and 55.8% in the first bin ($0 \leq FE \leq 0.25$) for YQ, ND and BA, respectively. Thus, the forecasted error distributions also demonstrate that the integration of 3-stage FS, BO and BiLSTM model presents a distinct advantage for 10-minute U forecasting.

Although a wide range of evaluation metrics and diagnostic plots were used for model comparison, it is a strenuous task to rank a large number of models reliably based on this. Hence, a more robust global performance indicator (GPI) was used for this purpose. Table 6 presents the GPI and the ranks of models. The 3-phase models had superior performance (Rank 1 $\approx$ 96.7%) over the 2-phase (Rank 2 $\approx$ 96.7%) and standalone models (Rank 3 $\approx$ 100%). Overall, the proposed FS-BO-BiLSTM model had the greatest GPI and the best predictive performance as Rank 1/15 model for all six sites. This is because the 3-stage FS and BO were successful in selecting the best features and hyperparameters, respectively that boosted the forecasting accuracy of BiLSTM over all other models.

4.1. Application of the proposed framework

Based on the promising results obtained, the proposed framework is useful in designing a robust energy security platform for a developing nation like Fiji to help meet its NDC and RE targets. Fiji has an excellent wind regime with high potential for wind power production [7]. However, strategic policy and motivation is required from investors to promote wind energy, which is somewhat hindered due to the relatively low energy output from the present 10 MW Butoni wind farm [9]. This is evidenced from the data [5] that shows that wind has contributed to only 0.12% to Fiji's electricity mix in 2020, and this was due to only 28 out of 37 functional turbines and a tropical cyclone occurring during the wet season.

The wet season is when wind energy is severely affected due to more wind variations because of cyclonic events. To support this claim, the $RRMSE$ (%) results of seasonal forecasts presented in Table 7 reveals higher $RRMSE$ for the wet season amongst all sites. This is because of the highly stochastic variations in U during the wet season. Therefore, the failure of Butoni in Fiji, while the same wind turbine model has proven to be successful in Vanuatu and New Caledonia [9] indicating that Fiji's issue is more to do with site selection and inability to address the intermittency issues. The site selection issue can be solved by conducting a feasibility study before the implementation of wind energy project(s). The complex intermittency issue can be overcome by the proposed near real-time forecasting tool developed in this study.

In real applications, the proposed framework depicted as Fig. 9 can be used, where one system is used to separately tune and train the off-line proposed model and the other one is used to make on-line

predictions. The on-line system can use the pre-trained FS-BO-BiLSTM model to make real-time predictions using new input data. Simultaneously, the off-line system can update the proposed model by tuning and training the model using additional new data. The pre-trained model will be replaced with the updated model every once in a while, to ensure that accurate and reliable predictions are being made.

This accurate forecasting tool can provide reliable predictions to help the grid operators make carefully planned decisions. For instance, it can inform in advance whether the future U will be below the cut-in U or above the cut-out U of the turbine. This information will help prevent the frequent start-up or shut-down of wind energy converting turbines. Hence, this will reduce the wear and tear of the gearbox, reduce the failure rate of the turbines and increase its operational lifespan [95]. Moreover, the predictions from the proposed model can also help prevent the sudden increase or decrease in the capacity factor of the wind turbines due to variations in U. This can further enhance the stability, reliability, and the security of the electric power systems and avoid unwarranted power brownouts [96].

5. Conclusions

To endorse novel modelling and energy simulation technologies, that ideally monitor wind energy extraction and furnish a secured energy distribution platform, this study aimed to assess the robustness of a 3-phase hybrid model (FS-BO-BiLSTM) to forecast 10-minute wind speed (U) at six target sites in Fiji. This study appends merit to previous studies that have used neighbouring reference station data as model inputs, by using: additional meteorological variables, a 3-stage feature selection (FS) strategy to encapsulate the best predictors, and Bayesian optimization (BO) for efficient hyperparameter tuning of the BiLSTM model. The findings explicitly outline the superior predictive performance of the proposed model attaining the highest r (0.917 – 0.981), WI (0.957 – 0.991) and LM (0.643 – 0.850); and the lowest $RRMSE$ (9.556 – 23.832%) and $MAPE$ (8.750 – 21.480%) for all six sites. The proposed model also registers the largest proportion of predicted errors ($\approx$ 76.6 – 84.8%) in the smallest range $\leq |0.5|\,\mathrm{ms}^{-1}$ amongst all tested sites. Extending the scope of the proposed hybrid approach, the following work can validate the FS-BO-BiLSTM model in other emerging areas of interest, such as forecasting solar radiation, tidal and wave energy, energy demand, electricity price and etc. This study introduces a useful near real-time forecasting tool for wind energy providers to make the energy generation and distribution decisions effectively, in terms of smart grids and economic integration of wind energy, and energy management systems in Fiji Islands and related countries.

However, the proposed model has a few limitations, and the efforts to overcome these challenges can be considered as potential future research directions. These are as follows:

i. The proposed model has implemented a single-step U forecasting approach. This gives limited information regarding the future U. Hence, the model needs to be tested at different time scales prior to being implemented in energy management systems.
ii. The single-step outputs are expressed as point forecasts. In future, interval and probabilistic forecasting approaches should be considered. This is because these methods can provide an estimate of the possible future range of U and the uncertainty in the forecasts.
iii. Model hybridization via data decomposition strategies is not tested. It is suggested that the performance of the proposed model can be further improved via techniques like improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN), robust local mean decomposition (RLMD), stationary wavelet transform (SWT) and etc.
iv. Better alternatives to the 3-stage FS adopted in this study can be tested in future by integrating the proposed model with CNN and autoencoder methods to facilitate effective feature extraction.
v. The FS-BO-BiLSTM model designed in this study is a black-box model. Hence, it does not explain the influence of climate features on U forecasts. Due to which, the physical dynamics between the climate indices used and the U is ignored. To resolve this, the proposed model needs to be interpreted via the eXplainable Artificial Intelligence (XAI) techniques in future to understand the underlying mechanism of the proposed black-box model.

Table A.1
List of acronyms.

Acronym	Full name
AI	Artificial Intelligence
ANN	Artificial Neural Network
BiLSTM	Bidirectional LSTM
BO	Bayesian Optimization
Boruta-RF	Boruta-Random Forest Hybridizer Algorithm
CCF	Cross-Correlation Function
CNN	Convolutional Neural Network
CPU	Central Processing Unit
DL	Deep Learning
DT	Decision Tree
EFL	Energy Fiji Limited
E_{NS}	Nash–Sutcliffe Efficiency
FS	Feature Selection
GPI	Global Performance Index
GPU	Graphics Processing Unit
GRU	Gated Recurrent Unit
GS	Grid Search
KGE	Kling–Gupta Efficiency
LM	Legates–McCabe's Index
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
$MAPE$	Mean Absolute Percentage Error
ML	Machine Learning
MLP	Multilayer Perceptron
NDC	Nationally Determined Contribution
$NMAE$	Normalized MAE
$NRMSE$	Normalized $RMSE$
PACF	Partial Auto-Correlation Function
PDF	Probability Distribution Function
PM	Persistence Model
PSIDS	Pacific Small Island Developing States
r	Pearson's Correlation Coefficient
R^2	Coefficient of Determination
r_{cross}	Cross-Correlation Coefficient
RE	Renewable Energy
ReLU	Rectified Linear Unit
RF	Random Forest
$RMSE$	Root Mean Square Error
RNN	Recurrent Neural Network
RReliefF	Regression Relief-F
$RRMSE$	Relative $RMSE$
RS	Random Search
SDG	Sustainable Development Goal
U	Wind Speed
U^{FOR}	Forecasted U
U^{OBS}	Observed U
WI	Willmott's Index

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sancho Salcedo-Sanz reports financial support was provided by Universidad de Alcalá. Sancho Salcedo Sanz reports financial support was provided by Spanish Ministry of Science and Innovation (MICINN).

Acknowledgements

The authors would like to acknowledge the Fiji Meteorological Services (FMS) for providing data that facilitated this study. This study was supported by USQ International PhD Stipend and International PhD Tuition Fee Scholarships awarded to the first author by the University of Southern Queensland (USQ), managed by Graduate Research School (GRS). This research was partially supported by project PID2020-115454GB-C21 of Spanish Ministry of Science and Innovation (MICINN), to continually build research synergies between Professors Ravinesh Deo (USQ, Australia) and Sancho Salcedo-Sanz (UAH, Spain). The authors also wish to thank the anonymous reviewers and the Editor-in-Chief for the useful comments that improved the quality of this paper.

Appendix A. Acronyms list

See Table A.1.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.renene.2022.12.123.

References

[1] F. Bilgili, E. Koçak, Ü. Bulut, S. Kuşkaya, Can biomass energy be an efficient policy tool for sustainable development? Renew. Sustain. Energy Rev. 71 (2017) 830–845, http://dx.doi.org/10.1016/j.rser.2016.12.109.

[2] H. Lucas, S. Fifita, I. Talab, C. Marschel, L. Cabeza, Critical challenges and capacity building needs for renewable energy deployment in Pacific small Island developing states (Pacific SIDS), Renew. Energy 107 (2017) 42–52, http://dx.doi.org/10.1016/j.renene.2017.01.029.

[3] C. Streck, P. Keenlyside, M. von Unger, The Paris agreement: A new beginning, J. Eur. Environ.Plan. Law 13 (2016) 3–29, http://dx.doi.org/10.1163/18760104-01301002.

[4] Ministry of Economy of Fiji Islands, Fiji's Updated Nationally Determined Contribution, Tech. rep, Ministry of Economy, Suva, Republic of Fiji, 2020, pp. 1–20.

[5] EFL, 2020 Annual Report, Tech. rep, Energy Fiji Limited, Suva, Republic of Fiji, 2020, pp. 1–96.

[6] A.S.K. Dalabeeh, Techno-economic analysis of wind power generation for selected locations in Jordan, Renew. Energy 101 (2017) 1369–1378, http://dx.doi.org/10.1016/j.renene.2016.10.003.

[7] K. Sharma, M.R. Ahmed, Wind energy resource assessment for the Fiji Islands: Kadavu Island and Suva Peninsula, Renew. Energy 89 (2016) 168–180, http://dx.doi.org/10.1016/j.renene.2015.12.014.

[8] E. Michalena, V. Kouloumpis, J.M. Hills, Challenges for Pacific small Island developing states in achieving their nationally determined contributions (NDC), Energy Policy 114 (2018) 508–518, http://dx.doi.org/10.1016/j.enpol.2017.12.022.

[9] L. Joseph, R. Prasad, Viability of commercial on-shore wind farm sites in Viti Levu, Fiji, in: A. Singh (Ed.), Translating the Paris Agreement Into Action in the Pacific, Springer, Switzerland, 2020, pp. 151–176, http://dx.doi.org/10.1007/978-3-030-30211-5_7.

[10] J. Chen, G.-Q. Zeng, W. Zhou, W. Du, K.-D. Lu, Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization, Energy Convers. Manage. 165 (2018) 681–695, http://dx.doi.org/10.1016/j.enconman.2018.03.098.

[11] Y. Gao, S. Ma, T. Wang, T. Wang, Y. Gong, F. Peng, A. Tsunekawa, Assessing the wind energy potential of China in considering its variability/intermittency, Energy Convers. Manage. 226 (2020) 1–13, http://dx.doi.org/10.1016/j.enconman.2020.113580.

[12] H. Bouzgou, N. Benoudjit, Multiple architecture system for wind speed prediction, Appl. Energy 88 (7) (2011) 2463–2471.

[13] C. Wu, J. Wang, X. Chen, P. Du, W. Yang, A novel hybrid system based on multi-objective optimization for wind speed forecasting, Renew. Energy 146 (2020) 149–165.

[14] F. Famoso, S. Brusca, D. D'Urso, A. Galvagno, F. Chiacchio, A novel hybrid model for the estimation of energy conversion in a wind farm combining wake effects and stochastic dependability, Appl. Energy 280 (2020) 1–16, http://dx.doi.org/10.1016/j.apenergy.2020.115967.

[15] S. Salcedo-Sanz, Á.M. Pérez-Bellido, E.G. Ortiz-García, A. Portilla-Figueras, L. Prieto, F. Correoso, Accurate short-term wind speed prediction by exploiting diversity in input data using banks of artificial neural networks, Neurocomputing 72 (4–6) (2009) 1336–1341, http://dx.doi.org/10.1016/j.neucom.2008.09.010.

[16] S. Salcedo-Sanz, Á.M. Pérez-Bellido, E.G. Ortiz-García, A. Portilla-Figueras, L. Prieto, D. Paredes, Hybridizing the fifth generation mesoscale model with artificial neural networks for short-term wind speed prediction, Renew. Energy 34 (6) (2009) 1451–1457.

[17] R. Prasad, L. Joseph, R.C. Deo, Modeling and forecasting renewable energy resources for sustainable power generation: Basic concepts and predictive model results, in: A. Singh (Ed.), Translating the Paris Agreement Into Action in the Pacific, sixty eight ed., Springer, Switzerland, 2020, pp. 59–79, http://dx.doi.org/10.1007/978-3-030-30211-5_3.

[18] S. Salcedo-Sanz, E.G. Ortiz-Garcı, Á.M. Pérez-Bellido, A. Portilla-Figueras, L. Prieto, et al., Short term wind speed prediction based on evolutionary support vector regression algorithms, Expert Syst. Appl. 38 (4) (2011) 4052–4057.

[19] Q. Han, F. Meng, T. Hu, F. Chu, Non-parametric hybrid models for wind speed forecasting, Energy Convers. Manage. 148 (2017) 554–568, http://dx.doi.org/10.1016/j.enconman.2017.06.021.

[20] A. Troncoso, S. Salcedo-Sanz, C. Casanova-Mateo, J. Riquelme, L. Prieto, Local models-based regression trees for very short-term wind speed prediction, Renew. Energy 81 (2015) 589–598.

[21] A. Kusiak, Z. Zhang, Short-horizon prediction of wind power: A data-driven approach, IEEE Trans. Energy Convers. 25 (4) (2010) 1112–1122, http://dx.doi.org/10.1109/TEC.2010.2043436.

[22] A.T. Sergio, T.B. Ludermir, Deep learning for wind speed forecasting in north-eastern region of Brazil, in: 2015 Brazilian Conference on Intelligent Systems, BRACIS, IEEE, Natal, Brazil, 2015, pp. 322–327, http://dx.doi.org/10.1109/BRACIS.2015.40.

[23] Q. Cao, B.T. Ewing, M.A. Thompson, Forecasting wind speed with recurrent neural networks, European J. Oper. Res. 221 (2012) 148–154, http://dx.doi.org/10.1016/j.ejor.2012.02.042.

[24] A. Kumar, A.B.M.S. Ali, Prospects of wind energy production in the western Fiji — An empirical study using machine learning forecasting algorithms, in: 2017 Australasian Universities Power Engineering Conference, AUPEC, IEEE, Melbourne, VIC, Australia, 2018, pp. 1–5, http://dx.doi.org/10.1109/AUPEC.2017.8282443.

[25] H. Acikgoz, U. Budak, D. Korkmaz, C. Yildiz, WSFNet: An efficient wind speed forecasting model using channel attention-based densely connected convolutional neural network, Energy 233 (2021) 1–16, http://dx.doi.org/10.1016/j.energy.2021.121121.

[26] C. Li, G. Tang, X. Xue, A. Saeed, X. Hu, Short-term wind speed interval prediction based on ensemble GRU model, in: IEEE Transactions on Sustainable Energy, IEEE, 2020, pp. 1370–1380, http://dx.doi.org/10.1109/TSTE.2019.2926147.

[27] X. Liao, Z. Liu, W. Deng, Short-term wind speed multistep combined forecasting model based on two-stage decomposition and LSTM, Wind Energy (2021) 1–22, http://dx.doi.org/10.1002/we.2613.

[28] T. Liang, Q. Zhao, Q. Lv, H. Sun, A novel wind speed prediction strategy based on Bi-LSTM, MOOFADA and transfer learning for centralized control centers, Energy 230 (2021) 1–16, http://dx.doi.org/10.1016/j.energy.2021.120904.

[29] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359, http://dx.doi.org/10.1109/TKDE.2009.191.

[30] J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. 79 (8) (1982) 2554–2558, http://dx.doi.org/10.1073/pnas.79.8.2554.

[31] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, 2014, pp. 1–15, ArXiv Preprint arXiv:arXiv:1406.1078.

[32] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, http://dx.doi.org/10.1162/neco.1997.9.8.1735.

[33] Z. Ma, H. Chen, J. Wang, X. Yang, R. Yan, J. Jia, W. Xu, Application of hybrid model based on double decomposition, error correction and deep learning in short-term wind speed prediction, Energy Convers. Manage. 205 (2020) 1–17, http://dx.doi.org/10.1016/j.enconman.2019.112345.

[34] A. Graves, J. Schmidhuber, Framewise phoneme classification with bidirectional LSTM and other neural network architectures, Neural Netw. 18 (5–6) (2005) 602–610, http://dx.doi.org/10.1016/j.neunet.2005.06.042.

[35] J. Xiang, Z. Qiu, Q. Hao, H. Cao, Multi-time scale wind speed prediction based on WT-bi-LSTM, in: MATEC Web of Conferences, Vol. 309, 2020, pp. 1–12, http://dx.doi.org/10.1051/matecconf/202030905011.

[36] M. Neshat, M.M. Nezhad, E. Abbasnejad, S. Mirjalili, L.B. Tjernberg, D.A. Garcia, B. Alexander, M. Wagner, A deep learning-based evolutionary model for short-term wind speed forecasting: A case study of the lillgrund offshore wind farm, Energy Convers. Manage. 236 (2021) 1–25, http://dx.doi.org/10.1016/j.enconman.2021.114002.

[37] K. Jaseena, B.C. Kovoor, Decomposition-based hybrid wind speed forecasting model using deep bidirectional LSTM networks, Energy Convers. Manage. 234 (2021) 1–26, http://dx.doi.org/10.1016/j.enconman.2021.113944.

[38] M.-A. Zöller, M.F. Huber, Benchmark and survey of automated machine learning frameworks, J. Artificial Intelligence Res. 70 (2021) 409–472, http://dx.doi.org/10.1613/jair.1.11854.

[39] L. Cornejo-Bueno, E.C. Garrido-Merchán, D. Hernández-Lobato, S. Salcedo-Sanz, Bayesian optimization of a hybrid system for robust ocean wave features prediction, Neurocomputing 275 (2018) 818–828.

[40] M.S. Alam, N. Sultana, S.Z. Hossain, Bayesian optimization algorithm based support vector regression analysis for estimation of shear capacity of FRP reinforced concrete members, Appl. Soft Comput. 105 (2021) 1–11, http://dx.doi.org/10.1016/j.asoc.2021.107281.

[41] J. Močkus, On Bayesian methods for seeking the extremum, in: G.I. Marchuk (Ed.), Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974, in: Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 1975, pp. 400–404, http://dx.doi.org/10.1007/3-540-07165-2_55.

[42] A. Verma, Z. Dai, B.K.H. Low, Bayesian optimization under stochastic delayed feedback, in: International Conference on Machine Learning, PMLR, 2022, pp. 22145–22167.

[43] A.D. Bethke, Genetic Algorithms As Function Optimizers, University of Michigan, 1980.

[44] W. Zhang, C. Wu, H. Zhong, Y. Li, L. Wang, Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization, Geosci. Front. 12 (2021) 469–477, http://dx.doi.org/10.1016/j.gsf.2020.03.007.

[45] G.T. Ribeiro, J.G. Sauer, N. Fraccanabbia, V.C. Mariani, L.d.S. Coelho, Bayesian optimized echo state network applied to short-term load forecasting, Energies 13 (9) (2020) 1–19, http://dx.doi.org/10.3390/en13092390.

[46] M.I. Sameen, B. Pradhan, S. Lee, Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment, CATENA 186 (2020) 1–13, http://dx.doi.org/10.1016/j.catena.2019.104249.

[47] M. Bilgili, B. Sahin, A. Yasar, Application of artificial neural networks for the wind speed prediction of target station using reference stations data, Renew. Energy 32 (14) (2007) 2350–2360, http://dx.doi.org/10.1016/j.renene.2006.12.001.

[48] S. Velázquez, J.A. Carta, J. Matías, Comparison between ANNs and linear MCP algorithms in the long-term estimation of the cost per kW h produced by a wind turbine at a candidate site: A case study in the Canary Islands, Appl. Energy 88 (11) (2011) 3869–3881, http://dx.doi.org/10.1016/j.apenergy.2011.05.007.

[49] J.J. Currie, P.J. Goulet, A.W. Ratsimandresey, Wind conditions in a Fjordlike bay and predictions of wind speed using neighboring stations employing neural network models, J. Appl. Meteorol. Climatol. 53 (6) (2014) 1525–1537, http://dx.doi.org/10.1175/JAMC-D-12-0339.1.

[50] R.C. Deo, M.A. Ghorbani, S. Samadianfard, T. Maraseni, M. Bilgili, M. Biazar, Multi-layer perceptron hybrid model integrated with the firefly optimizer algorithm for windspeed prediction of target site using a limited set of neighboring reference station data, Renew. Energy 116 (Part A) (2018) 309–323, http://dx.doi.org/10.1016/j.renene.2017.09.078.

[51] A.M. Ahmed, R.C. Deo, Q. Feng, A. Ghahramani, N. Raj, Z. Yin, L. Yang, Deep learning hybrid model with Boruta-random forest optimiser algorithm for streamflow forecasting with climate mode indices, rainfall, and periodicity, J. Hydrol. 599 (2021) 1–23, http://dx.doi.org/10.1016/j.jhydrol.2021.126350.

[52] D. Bechrakis, P. Sparis, Correlation of wind speed between neighboring measuring stations, IEEE Trans. Energy Convers. 19 (2) (2004) 400–406, http://dx.doi.org/10.1109/TEC.2004.827040.

[53] M.B. Kursa, A. Jankowski, W.R. Rudnicki, Boruta – A system for feature selection, Fund. Inform. 101 (4) (2010) 271–285, http://dx.doi.org/10.3233/FI-2010-288.

[54] A.M. Ahmed, R.C. Deo, A. Ghahramani, N. Raj, Q. Feng, Z. Yin, L. Yang, LSTM integrated with Boruta-random forest optimiser for soil moisture estimation under RCP4.5 and RCP8.5 global warming scenarios, Stoch. Environ. Res. Risk Assess. 35 (2021) 1851–1881, http://dx.doi.org/10.1007/s00477-021-01969-3.

[55] N. Raj, J. Brown, An EEMD-BiLSTM algorithm integrated with Boruta random forest optimiser for significant wave height forecasting along coastal areas of Queensland, Australia, Remote Sens. 13 (8) (2021) 1–20, http://dx.doi.org/10.3390/rs13081456.

[56] R.J. Urbanowicz, R.S. Olson, P. Schmitt, M. Meeker, J.H. Moore, Benchmarking relief-based feature selection methods for bioinformatics data mining, J. Biomed. Inform. 85 (2018) 168–188, http://dx.doi.org/10.1016/j.jbi.2018.07.015.

[57] M. Robnik-Šikonja, I. Kononenko, Theoretical and empirical analysis of relieff and RReliefF, Mach. Learn. 53 (2003) 23–69, http://dx.doi.org/10.1023/A:1025667309714.

[58] T. Fischer, C. Krauss, Deep learning with long short-term memory networks for financial market predictions, European J. Oper. Res. 270 (2) (2018) 654–669, http://dx.doi.org/10.1016/j.ejor.2017.11.054.

[59] Ö. Yildirim, A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification, Comput. Biol. Med. 96 (2018) 189–202, http://dx.doi.org/10.1016/j.compbiomed.2018.03.016.

[60] D.E. Rumelhart, J.L. McClelland, Learning internal representations by error propagation, in: Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations, MIT Press, 1987, pp. 318–362.

[61] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, http://dx.doi.org/10.1023/A:1010933404324.

[62] F. Rosenblatt, The perceptron: A probabilistic model for information storage and organization in the brain, Psychol. Rev. 65 (6) (1958) 386–408, http://dx.doi.org/10.1037/h0042519.

[63] M. Injadat, F. Salo, A.B. Nassif, A. Essex, A. Shami, Bayesian optimization with machine learning algorithms towards anomaly detection, in: 2018 IEEE Global Communications Conference, GLOBECOM, IEEE, Abu Dhabi, United Arab Emirates, 2018, pp. 1–6, http://dx.doi.org/10.1109/GLOCOM.2018.8647714.

[64] J. Snoek, H. Larochelle, R.P. Adams, Practical Bayesian optimization of machine learning algorithms, in: F. Pereira, C. Burges, L. Bottou, K. Weinberger (Eds.), Advances in Neural Information Processing Systems, Twentyfifith ed., NeurIPS Proceedings, 2012, pp. 1–9, URL https://proceedings.neurips.cc/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf.

[65] M. Seeger, Gaussian processes for machine learning, Int. J. Neural Syst. 14 (2) (2004) 69–106, http://dx.doi.org/10.1142/S0129065704001899.

[66] F. Hutter, H.H. Hoos, K. Leyton-Brown, Sequential model-based optimization for general algorithm configuration, in: C.A.C. Coello (Ed.), Learning and Intelligent Optimization, Six thousand six hundred eight-two ed., LION 2011, in: Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2011, pp. 507–523, http://dx.doi.org/10.1007/978-3-642-25566-3_40.

[67] J. Bergstra, R. Bardenet, Y. Bengio, B. Kégl, Algorithms for hyper-parameter optimization, in: J. Shawe-Taylor, R.S. Zemel, P.L. Bartlett, F. Pereira, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems, Twenty four ed., Curran Associates Inc., 2011, pp. 2546–2554, URL https://proceedings.neurips.cc/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf.

[68] R. Elshawi, M. Maher, S. Sakr, Automated machine learning: State-of-the-art and open challenges, 2019, pp. 1–23, ArXiv Preprint arXiv:1906.02287.

[69] I. Koprinska, M. Rana, V.G. Agelidis, Correlation and instance based feature selection for electricity load forecasting, Knowl.-Based Syst. 82 (2015) 29–40, http://dx.doi.org/10.1016/j.knosys.2015.02.017.

[70] M. Robnik-Sikonja, I. Kononenko, An adaptation of relief for attribute estimation in regression, in: ICML '97: Proceedings of the Fourteenth International Conference on Machine Learning, 1997, pp. 296–304.

[71] M.A. Ferraciolli, F.F. Bocca, L.H.A. Rodrigues, Neglecting spatial autocorrelation causes underestimation of the error of sugarcane yield models, Comput. Electron. Agric. 161 (2019) 233–240, http://dx.doi.org/10.1016/j.compag.2018.09.003.

[72] A. Rafati, M. Joorabian, E. Mashhour, An efficient hour-ahead electrical load forecasting method based on innovative features, Energy 201 (2020) 1–13, http://dx.doi.org/10.1016/j.energy.2020.117511.

[73] M.B. Kursa, W.R. Rudnicki, Feature selection with the Boruta package, J. Stat. Softw. 36 (11) (2010) 1–13, http://dx.doi.org/10.18637/jss.v036.i11.

[74] C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, A. Zeileis, Conditional variable importance for random forests, BMC Bioinformatics 9 (307) (2008) 1–11, http://dx.doi.org/10.1186/1471-2105-9-307.

[75] F. Chollet, et al., Keras, GitHub, 2015, http://github.com/fchollet/keras.

[76] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, et al., TensorFlow: A system for large-scale machine learning, 2016, pp. 1–18, ArXiv Preprint arXiv:1605.08695.

[77] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830, http://dx.doi.org/10.5555/1953048.2078195.

[78] J. Bergstra, D. Yamins, D. Cox, Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures, in: S. Dasgupta, D. McAllester (Eds.), Proceedings of the 30th International Conference on Machine Learning, Twenty eight ed., PMLR, Atlanta, Georgia, USA, 2013, pp. 115–123, URL https://proceedings.mlr.press/v28/bergstra13.html.

[79] R.C. Deo, X. Wen, X. Wen, F. Qi, A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset, Appl. Energy 168 (2016) 568–593, http://dx.doi.org/10.1016/j.apenergy.2016.01.130.

[80] D.A. Dickey, W.A. Fuller, Distribution of the estimators for autoregressive time series with a unit root, J. Amer. Statist. Assoc. 74 (366a) (1979) 427–431.

[81] R.F. Engle, C.W. Granger, Co-integration and error correction: Representation, estimation, and testing, Econometrica (1987) 251–276.

[82] S.P. Kumar, D. Lopez, Feature selection used for wind speed forecasting with data driven approaches, J. Eng. Sci. Technol. Rev. 8 (5) (2015) 124–127.

[83] S. Ghimire, R.C. Deo, N. Raj, J. Mi, Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms, Appl. Energy 253 (2019) 1–20, http://dx.doi.org/10.1016/j.apenergy.2019.113541.

[84] T.T. Joy, S. Rana, S. Gupta, S. Venkatesh, Batch Bayesian optimization using multi-scale search, Knowl.-Based Syst. 187 (2020) 1–11, http://dx.doi.org/10.1016/j.knosys.2019.06.026.

[85] O. Behar, A. Khellaf, K. Mohammedi, Comparison of solar radiation models and their validation under Algerian climate – the case of direct irradiance, Energy Convers. Manage. 98 (2015) 236–251, http://dx.doi.org/10.1016/j.enconman.2015.03.067.

[86] R. Prasad, R.C. Deo, Y. Li, T. Maraseni, Ensemble committee-based data intelligent approach for generating soil moisture forecasts with multivariate hydro-meteorological predictors, Soil Tillage Res. 181 (2018) 63–81, http://dx.doi.org/10.1016/j.still.2018.03.021.

[87] T. Chai, R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)?–arguments against avoiding RMSE in the literature, Geosci. Model Dev. 7 (3) (2014) 1247–1250.
[88] J. Hora, P. Campos, A review of performance criteria to validate simulation models, Expert Syst. 32 (2015) 578–595, http://dx.doi.org/10.1111/exsy.12111.
[89] M.-F. Li, X.-P. Tang, W. Wu, H.-B. Liu, General models for estimating daily global solar radiation for different solar radiation zones in mainland China, Energy Convers. Manage. 70 (2013) 139–148, http://dx.doi.org/10.1016/j.enconman.2013.03.004.
[90] A. Saeed, C. Li, M. Danish, S. Rubaiee, G. Tang, Z. Gan, A. Ahmed, Hybrid bidirectional LSTM model for short-term wind speed interval prediction, IEEE Access 8 (2020) 182283–182294, http://dx.doi.org/10.1109/ACCESS.2020.3027977.
[91] J. Nash, J. Sutcliffe, River flow forecasting through conceptual models part I — A discussion of principles, J. Hydrol. 10 (3) (1970) 282–290, http://dx.doi.org/10.1016/0022-1694(70)90255-6.
[92] D.R. Legates, G.J. McCabe., Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation, Water Resour. Res. 35 (1999) 233–241, http://dx.doi.org/10.1029/1998WR900018.
[93] C.J. Willmott, On the validation of models, Phys. Geogr. 2 (2) (1981) 184–194, http://dx.doi.org/10.1080/02723646.1981.10642213.
[94] C.J. Willmott, On the evaluation of model performance in physical geography, in: G.L. Gaile, C.J. Willmott (Eds.), Spatial Statistics and Models, forty ed., Springer, Dordrecht, 1984, pp. 443–460, http://dx.doi.org/10.1007/978-94-017-3048-8_23.
[95] U. Bhardwaj, A. Teixeira, C.G. Soares, Reliability prediction of an offshore wind turbine gearbox, Renew. Energy 141 (2019) 693–706, http://dx.doi.org/10.1016/j.renene.2019.03.136.
[96] A.M. Foley, P.G. Leahy, A. Marvuglia, E.J. McKeogh, Current methods and advances in forecasting of wind power generation, Renew. Energy 37 (2012) 1–8, http://dx.doi.org/10.1016/j.renene.2011.05.033.