IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 January 2024, accepted 16 April 2024, date of publication 24 April 2024, date of current version 1 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3392899

RESEARCH ARTICLE

Multi-Step-Ahead Wind Speed Forecast System: Hybrid Multivariate Decomposition and Feature Selection-Based Gated Additive Tree Ensemble Model

LIONEL P. JOSEPH[1,2], RAVINESH C. DEO[1,2], (Senior Member, IEEE), DAVID CASILLAS-PÈREZ[3], RAMENDRA PRASAD[4], NAWIN RAJ[1], AND SANCHO SALCEDO-SANZ[5]

[1]School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, QLD 4300, Australia
[2]Centre for Applied Climate Sciences, University of Southern Queensland, Toowoomba, QLD 4350, Australia
[3]Department of Signal Processing and Communications, Universidad Rey Juan Carlos, Fuenlabrada, 28942 Madrid, Spain
[4]Department of Science, School of Science and Technology, The University of Fiji, Lautoka, Fiji
[5]Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, 28805 Madrid, Spain

Corresponding authors: Ravinesh C. Deo (ravinesh.deo@usq.edu.au) and Lionel P. Joseph (joseph.lionel73@gmail.com)

The work of Lionel P. Joseph was supported by the University of Southern Queensland (UniSQ) International Ph.D. Stipend and International Ph.D. Tuition Fee Scholarships Awarded through the Graduate Research School (GRS), UniSQ, under Grant 2020. The work of Ravinesh C. Deo and Sancho Salcedo-Sanz was supported in part by Spanish Ministry of Science and Innovation (MICINN) under Project PID2020-115454GB-C21.

ABSTRACT Wind, being a clean and sustainable resource, boasts environmental advantages. However, its electricity generation faces challenges due to unpredictable variations in wind speed (*WS*). Accurate predictions of these variations would allow mixed grids to adjust their energy mix in real-time, ensuring overall stability. For this purpose, the paper develops a new hybrid gated additive tree ensemble (H-GATE) model for accurate multi-step-ahead *WS* predictions. First, the multivariate empirical mode decomposition (MEMD) simultaneously demarcates the multivariate data into intrinsic mode functions (IMFs) and residuals. These components represent underlying trends, periodicity, and stochastic patterns in *WS* variations. The IMF and residual components are pooled in respective sets, and an opposition-based whale optimization algorithm (OBWOA) is applied for dimensionality reduction. The selected features are used by GATE tuned with Bayesian optimization (BO) to forecast the individual IMF and residual components. The outputs are summed to obtain the final multi-step-ahead *WS* forecasts. The proposed H-GATE is benchmarked against standalone (S-GATE, S-CLSTM, and S-ABR) and hybrid (H-CLSTM and H-ABR) models. Based on all statistical metrics and diagnostic plots, H-GATE outperforms all comparative models at all forecast horizons, accumulating the lowest mean absolute percentage error (*MAPE*) of 6.13 - 9.93% (at t_{L+1}), 8.67 - 14.07% (at t_{L+2}), and 11.60 - 18.37% (at t_{L+3}) across all three sites. This novel multi-step-ahead *WS* forecasting strategy can significantly benefit grid operators by helping anticipate fluctuations in wind power generation. This can assist in optimizing energy dispatch schedules, reducing reliance on backup power sources, and enhancing overall grid stability. Practical implementation of this method can help meet the rising energy demands through renewable wind energy.

INDEX TERMS Wind speed forecasting, gated additive tree ensemble, multivariate empirical mode decomposition, opposition-based whale optimization algorithm, Bayesian optimization.

I. INTRODUCTION

Rapid economic development has increased human demand for energy, which comes mostly from fossil fuels. As a

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

result of the use of conventional fossil fuels, greenhouse gases are released into the atmosphere, contributing to global warming. Due to their non-renewable nature, fossil fuels will eventually deplete. To overcome the environmental burden and energy crisis, Sustainable Development Goal 7 (SDG 7) was implemented in 2015 to increase global energy generation using clean renewable sources by 2030 [1]. Wind energy is one of the most promising renewable energy (RE) sources, which has received much attention globally due to its benefits of being abundant, eco-friendly, and cost-effective [2]. According to the Global Wind Energy Council (GWEC) report [3], the latest installed capacity of wind power in 2022 was 77.6 GW, which brings the cumulative global wind capacity to 906 GW, a $\approx 9\%$ increase since 2021. Further addition of wind power capacities can bring considerable economic benefits and reduce the adverse environmental effects caused by fossil fuels [4]. However, maximizing the utilization of wind power is challenging considering that the wind energy conversion systems pose serious risks to the power system stability due to the intermittent nature of wind speed (*WS*) [5]. For maximum wind energy generation and effective wind power management, accurate and stable *WS* forecasting is essential.

Generally, *WS* forecasting is performed at different time scales, including very short-term ($<$ 30 minutes), short-term (30 minutes - 6 hours), medium-term (6 hours - 1 day), and long-term ($>$ 1 day) [6]. This study tests the application of short-term *WS* forecasting as this range is crucial for turbine control, operational safety, and economic load dispatch [7]. Short-term *WS* forecasting helps promote better system stability, security, and wind power quality [8]. To achieve accurate *WS* forecasts, researchers have presented many strategies grouped into physical, statistical, and artificial intelligence (AI)-based methods, as summarized in Table 1.

Physical methods operate by simulating the underlying meteorological and physical principles governing atmospheric conditions and are mainly based on numerical weather prediction (NWP) models [9]. NWP models generally require a large amount of data to perform complex computations and are preferred for long-term forecasting [9]. NWP models are often combined with other methods to improve the forecast accuracy. For instance, Table 1 shows that the NWP model is combined with Gaussian process regression (GPR) in [10] and Kalman filter (KF) in [11] to correct the *WS* forecast errors. Despite bias correction, both studies revealed notable discrepancies. This is common for NWP models as uncertainties in parameterizations and limitations in understanding certain atmospheric processes can easily amplify forecast errors. Furthermore, the inherent complexity of implementing physical models renders them unsuitable for short-term *WS* forecasting [12].

Contrary to physical methods, statistical models, including autoregressive (AR), autoregressive moving average (ARMA), and autoregressive integrated moving average (ARIMA), exhibit distinct advantages in the short-term *WS* forecasting domain [13]. These models operate efficiently with relatively simple computations and excel in capturing and predicting short-term trends and patterns in time series data. For instance, Table 1 shows that ARMA used in [14] and ARIMA employed in [15] registered a low percentage of forecast errors when trained with univariate data. While AR-based variants work exceptionally well with linear univariate data, their applicability is unsuitable for nonlinear data with multivariate features [16]. Hence, using multivariate data with nonlinearity and complex fluctuations made statistical models inapt for this study. Contrariwise, AI-based machine learning (ML) and deep learning (DL) are superior in mapping nonlinear data with better self-learning capability [17]. Hence, ML and DL models have gained widespread acceptance over physical and statistical approaches.

The commonly applied ML methods for short-term *WS* forecasting include artificial neural network (ANN), support vector regression (SVR), and tree-based regressors like decision tree (DTR), random forest (RFR), adaptive boosting (ABR), gradient boosting (GBR), and extreme gradient boosting (XGBR) [18]. The prevalence of these models is evident in numerous studies, some highlighted in Table 1. For instance, ANN excels in multi-step-ahead wind speed (*WS*) forecasting in [19], outperforming ARIMA across all forecast horizons. Another study [20] applies a single-step forecasting approach using SVR, yielding reliable outcomes for short-term *WS* forecasting. In the tree-based category, ABR is employed in [21] for seasonal *WS* forecasting, investigating the stochastic nature of wind in different seasons. Despite notable achievements in these studies, these models exhibit certain limitations. ANN often gets stuck in local minima during model training, failing to locate the global minimum of the loss function [22]. While SVR can effectively reach the global solution, its scalability for large datasets is limited due to its computational and memory requirements [23]. Tree-based models are sensitive to outliers, and when used with larger datasets, a single tree may develop a large number of nodes, which causes overfitting [18]. To overcome these limitations, DL techniques are preferred, which can effectively capture complex hidden patterns in larger datasets.

Recurrent neural network (RNN) is a widely applied DL model for temporal dependence modelling but is prone to the vanishing and exploding gradient issue [24], [25]. Gated recurrent unit (GRU) [26] and long short-term memory (LSTM) [27] emerge as improved RNN variants, particularly prevalent in *WS* forecasting (Table 1). GRU is computationally less intensive, having only two gates and a single hidden state. Given its simple gating structure, GRU outperformed LSTM, SVR, and ANN in [28] for short-term *WS* forecasting. However, LSTM dominates GRU in tasks that require modelling long-term dependencies by capturing complex patterns in sequential data using its two separate memory cells [18]. For instance, LSTM used for short-term *WS* forecasting in [29] outperformed GRU and

ML-based XGBR, extreme learning machine (ELM), SVR, and ANN models. Additionally, LSTM applied for medium-term *WS* forecasting in [30] outperformed a statistical-based nonlinear AR (NAR) model by registering an average of 36.16% decrease in mean absolute percentage error (*MAPE*). The superior predictive performance of LSTM is due to its additional gates and memory cells, which offer added benefits in remembering longer sequences in data. LSTM can be further enhanced through bidirectional processing, enabling the network to capture information from past and future time steps [31]. This is evident in [18], where bidirectional LSTM (BiLSTM) outperformed LSTM, RNN, ANN, and RFR for short-term *WS* forecasting. Moreover, a one-dimensional (1)-D) convolutional neural network (CNN) is another DL architecture that uses convolutional kernels for time series-based feature extraction [32]. CNN was fused with LSTM (CLSTM) in [33], where CNN extracted the spatial correlation of meteorological variables, and LSTM used the reconstructed input for multi-step-ahead wind power forecasting. The *MAPE* of CLSTM reduced by 8.55% and 23.49% compared to LSTM and CNN, respectively. Additionally, CNN is merged with the attention mechanism (AM) and BiLSTM in [34] for optimal short-term *WS* prediction. These studies show that the feature extraction capability of 1-D CNN has proven essential in improving the prediction accuracy of time series data. However, the convolutional filters are applied across the entire sequence simultaneously, which causes a loss of temporal context [35]. Therefore, acknowledging the limitations of the reviewed ML and DL models, this study uses a gated additive tree ensemble (GATE) model [36], which is a novel high-performance DL architecture for tabular data.

GATE is a parameter-efficient DL structure, which uses a gating mechanism inspired by GRU. It includes four main components [36]: gated feature learning units (GFLUs), differential nonlinear decision trees (DTs) as the key inductive bias, self-attention between the DT outputs, and additive ensemble of DT outputs to obtain the desired output. The hierarchical gating mechanism in GATE facilitates feature extraction at each decision step, allowing the model to selectively retain valuable information while discarding irrelevant details. Additionally, the adoption of stacked soft oblivious DTs, where the output of preceding DTs informs subsequent ones, has proven effective in enhancing prediction accuracy, as evidenced in [37]. Thus, integrating gating mechanisms and stacked DTs enhances the model's robustness and elevates its predictive capabilities, making it a suitable choice for *WS* forecasting. To experimentally explore its potential, Joseph and Raj [36] compared GATE against state-of-the-art tree-based GBR and tabular DL-based (i.e., NODE, Transformer, and TabNet) models using several public datasets. GATE outperformed both tree-based and DL-based models in terms of accuracy, was more efficient than the DL counterparts, and handled a large number of features effectively using its excellent feature representation learning unit. Unlike most DL models, GATE offers higher stability in achieving the least errors [36]. However, GATE has more parameters than the soft tree-based models and its performance is largely dependent on optimal hyperparameter selection. An efficient Bayesian optimization (BO) [38] is proposed for this task. BO is a probabilistic optimization method, which aims to locate the global optimum of an objective function while minimizing the number of function evaluations. It efficiently explores the hyperparameter search space, balancing 'exploration' to sequentially select new promising regions and 'exploitation'to narrow the search around the best-known solutions [38]. The efficiency of BO was explored in [18] to optimize BiLSTM for short-term *WS* forecasting, where extensive data (i.e., 105,120 data points) was used during model training and optimization. Despite the large dataset used, the optimization time of BO was significantly lower than the popular grid search and random search methods. Therefore, given its excellent optimization ability, BO is used to optimize GATE in this study.

Furthermore, *WS* data is erratic in nature; hence, utilizing raw data to develop predictive models would result in large errors [39], [40]. Data decomposition techniques are proposed to reduce noise from such complex data [17]. Various multiresolution analysis (MRA) tools have been tested to extract the embedded information within the non-stationary data. Wavelet transform (WT) is the most commonly used MRA tool for *WS* data decomposition, being combined with several models including ARIMA [15], SVR [20], and LSTM [41] for short-term *WS* forecasting. Although WT is a powerful MRA tool, its dependence on the pre-selection of the mother wavelet function is one of its major drawbacks [42]. Unlike WT, empirical mode decomposition (EMD) is an adaptive method, which does not rely upon a predefined basis function, making it well-suited for handling nonlinear signals [43]. EMD helps segregate high-frequency raw data into several intrinsic mode function (IMF) components and a residual component of different frequencies without any information loss. The effectiveness of EMD is evaluated in [44], where its integration with Elman neural network (ENN) consistently achieved the largest reduction in *MAPE* when compared with persistent (−53.76%), ANN (−45.58%), and standalone ENN (−29.82%).

The commonly used variants of EMD include ensemble EMD (EEMD) [45], complete EEMD with adaptive noise (CEEMDAN) [46], and improved CEEMDAN (ICEEMDAN) [47], which have been developed to alleviate the shortcomings of their respective predecessors. These four methods hybridized LSTM in [30], where the improvements from basic EMD to advanced ICEEMDAN were evident through a significant decrease in forecast errors. The resulting *MAPE* scores demonstrated this improvement, showcasing values of 17.76% for standalone LSTM, 11.04% for EMD-LSTM, 7.38% for EEMD-LSTM, 7.34% for CEEMDAN-LSTM, and 6.62% for ICEEMDAN-LSTM.

EMD-variants help preserve the physical configurations of the temporal input data. However, despite the progressive evolution of these four MRA tools, their application is limited to univariate data [48], [49]. These EMD-variants can only decompose a single time series at once, and the IMFs retrieved after processing multiple time series data cannot assure a consistent frequency separation [50]. Hence, these variants are not suitable for this study since multivariate data is used. Therefore, a modified multivariate EMD (MEMD) [51] is used for simultaneous data decomposition.

The variability of *WS* is dependent on many meteorological factors including wind direction, temperature, humidity, sea level pressure, solar radiation, rainfall, etc [18]. These variables need to be appropriately decomposed into sub-series with consistent frequency division. MEMD helps achieve this by resolving the mode alignment issue in the joint analysis of multi-oscillatory components within $n-$dimensional data [52].

The application of MEMD has shown remarkable results in forecasting solar radiation using RFR [42], [53], solar photovoltaic power using LSTM [54], evapotranspiration with LSTM [55], and building heat load using SVR [50]. Using MEMD in these studies preserved the cross-variable relationships in the decomposed components, allowing for a more accurate representation of the interdependencies between different variables employed. For more robust performance, hybrid singular spectrum analysis (SSA) and MEMD method are used in [34] for *WS* forecasting using CNN combined with AM and BiLSTM. In this study, SSA denoised the original multivariate data and MEMD decomposed the denoised series to obtain superior forecast accuracy. The use of MEMD in various domains affirms its effectiveness as a multichannel MRA tool. Yet, its application is not fully explored in short-term *WS* forecasting, especially with the GATE model. Thus, this study aims to narrow this research gap.

The multivariate data decomposed via MEMD generates numerous sub-series of IMFs and residual components. For each IMF of the target *WS* data, there are several corresponding IMFs of the predictor variables. To achieve optimal model performance, relevant predictors must be selected using a suitable feature selection (FS) method. The efficacy of a wrapper-based Boruta FS technique is highlighted in [31], successfully addressing challenges related to the curse of dimensionality and overfitting. Another study [18] integrated Boruta with a filter-based RReliefF method for FS, achieving notably accurate results in short-term *WS* forecasting. However, the study reported that integrating different FS techniques can be time-consuming, especially for large datasets. For ideal FS outcomes, meta-heuristics have become widely recommended for the following reasons [56]: ease of implementation, not needing any gradient information, and can bypass local optima. Meta-heuristics mainly mimic the physical and biological behaviour of nature [57].

Among these methods, the algorithms inspired by the intelligence of natural swarms are widely used. Some examples include particle swarm optimization (PSO) [58], grey wolf optimization (GWO) [59], firefly algorithm (FA) [60], crow search algorithm (CSA) [61], and whale optimization algorithm (WOA) [56]. Recently, WOA has been applied successfully in *WS* forecasting studies [62], [63] for hyperparameter selection. The WOA, which mimics the hunting behaviour of humpback whales, has many properties that allow it to locate the optimal solution for optimization problems. The implementation of WOA is easy and it has minimal algorithm-specific parameters to tune, which are adjusted along the iterations [64]. However, during optimization, some search agents (i.e., whales) update their location based on the position of the best solution. In real, the best solution might be in the opposite direction. This causes poor exploration ability and a tendency to get trapped in local optima solutions [65]. To resolve this, opposition-based WOA (OBWOA) is proposed [66], which explores the search space in two directions at the same time to obtain a global optimum solution. Leveraging its advanced exploration capabilities within the search space, OBWOA is implemented to formulate accurate mechanisms for estimating solar cell parameters in [65], yielding robust outcomes. Consequently, in this study, OBWOA is employed to harness similar advantages for optimizing the performance of FS for accurate *WS* forecasting.

The contributions and novelty of this paper are summarized as follows:

1) A new gated additive tree ensemble (GATE) architecture is employed to forecast short-term hourly *WS* using multivariate ground-based meteorological inputs. Within its architectural framework, GATE boasts exceptional gated feature representation learning units, demonstrating a superior ability for feature extraction compared to the popularly used CNN.
2) Most data decomposition strategies in the literature focus on univariate data segregation. When applied to multivariate data, these methods do not guarantee consistent frequency division. Consequently, a multivariate empirical mode decomposition (MEMD) tool is applied to concurrently decompose multichannel data into sub-series with uniform intrinsic mode function (IMF) and residual components to structure a hybrid GATE (i.e., H-GATE) model.
3) The decomposition of multivariate data yields numerous predictors, some of which are redundant, noisy, and poorly correlated to the target variable. Using such predictors lowers the predictive accuracy and increases the computational complexity. Hence, an opposition-based whale optimization algorithm (OBWOA) with improved exploration ability than WOA and most meta-heuristic algorithms is used for robust dimensionality reduction.
4) Unlike most DL models, GATE exhibits a parameter-efficient structure. However, to maximize its predictive

performance, the model hyperparameters undergo tuning via Bayesian optimization (BO). In contrast to the conventional grid and random search methods, BO excels by adapting its sequential search based on past evaluations, leading to more efficient optimization of the GATE model.

5) A multi-step-ahead forecasting approach is employed to predict *WS* at one-step-ahead (1 hour), two-step-ahead (2 hours), and three-step-ahead (3 hours) forecast horizons. The results derived from this method offer advanced insights into future wind energy production, facilitating proactive grid management.

The remainder of this paper is organized as follows: Section II outlines the theoretical overview of the algorithms. Section III describes the data and methodology. Section IV discusses the results. Section V presents the conclusions. Finally, Section VI addresses the study limitations and proposes future recommendations.

II. THEORETICAL OVERVIEW

The proposed H-GATE forecast model was built using the MEMD, OBWOA, and GATE algorithms, as described in this section. The predictive models used for benchmarking purposes: ABR [67] and CLSTM [33] are well-known methods and hence, are not explained in detail here.

A. MULTIVARIATE EMPIRICAL MODE DECOMPOSITION (MEMD)

The EMD [43] technique is commonly used for non-stationary data decomposition due to its ease of application. However, it suffers from the problem of mode mixing and is not suitable for multivariate data [68]. The IMFs generated via EMD for multivariate data decomposition, rarely correspond to the same frequency for different variables and often results in different number of IMFs for different variables. To overcome the inherent drawbacks of EMD, MEMD [51] is proposed for multivariate data decomposition. The MEMD method is an advanced version of EMD, which is designed to deal with multichannel signals and overcomes the mode mixing issue by utilizing white Gaussian noises [69]. MEMD effectively concurrently decomposes multivariate data into respective IMFs and a residue function. The first decomposed component (i.e., IMF-1) has the longest wavelength, and the frequency decreases with higher-order IMFs. The final component of MEMD computation is the residue function, which extracts the trend information. The steps of decomposition via MEMD are described as follows [51]:

- An appropriate pointset is located to sample on a $(n-1)$ hypersphere using the Hammersley sequences for $n-$dimensional time series:

$$x(t) = \{x_1(t), x_2(t), \ldots, x_n(t)\}, \quad t = 1, 2, \ldots, T \quad (1)$$

- The angles $\theta^k = \{\theta_1^k, \theta_2^k, \ldots, \theta_{n-1}^k\}, k = 1, 2, \ldots, K$ of the normalized Hammersley sequences are computed, where K represents the total number of projections.
- A set of projections $\{q^{\theta_k}(t)\}_{k=1}^K$ is determined for the multivariate signal $x(t)$ on k^{th} direction using the direction vectors $v^{\theta_k} = \{v_1^k, v_2^k, \ldots, v_n^k\}$.
- The instantaneous moment $\{t_i^{\theta_k}\}_{k=1}^K$ of the extreme value of the projection set is determined, where i represents the location of the extreme dot $i \in [1, T]$.
- The coordinates of the extreme dots $[t_i^{\theta_k}, x(t_i^{\theta_k})]$ are interpolated to obtain multiple envelope curves $\{e^{\theta_k}(t)\}_{k=1}^K$. The mean of multiple envelope curves $m(t)$ is obtained using:

$$m(t) = \frac{1}{K}\sum_{k=1}^{K} e^{\theta_k}(t) \quad (2)$$

- The final output is computed as:

$$h(t) = x(t) - m(t) \quad (3)$$

where output $h(t)$ is the multivariate IMFs if the stopping condition is met [70]. If the stopping criterion is not met, the sifting process is continued by repeating the above steps.

B. OPPOSITION-BASED WHALE OPTIMIZATION ALGORITHM (OBWOA)

The standard WOA mimics the hunting practice of humpback whales [56]. In WOA, each whale is considered as a potential candidate solution for a given optimization problem. For searching the prey location and hunting, the whale updates its position via two mechanisms: 1) encircling prey and 2) spiral movement. The 'exploration' of the search space is performed during optimization when the whales look for prey and the 'exploitation' happens when the whales attack the prey. However, a limitation of WOA is that it can get stuck in the local optimal point; hence, the convergence during FS becomes time-consuming [65]. This issue arises because some whales update their position according to the position of the best whale (i.e., solution). While in real, the optimal solution might be in the opposite direction to the current solution. To overcome this, the proposed OBWOA is designed to locate the global optimal solution in both forward and opposite directions [66]. The proposed OBWOA consists of two phases (Figure 1).

In phase one, a random population of whales (X) with N solution is generated, where the position is X_i, $(i = 1, 2, \ldots, N)$. Then the opposition-based learning (OBL) mechanism is used to improve the exploration and performance of the whales by computing their opposite positions $(\overline{X})$, given as:

$$\overline{X} = u + l - X \quad (4)$$

where u and l denote the upper and lower bounds of the search domain, respectively. Then, the best N positions (i.e., solutions) are retrieved from $X \cup \overline{X}$ to produce a new population.

In phase two, steps of the traditional WOA are used to update the whale positions using coefficient vectors

TABLE 1. Summary of reviewed wind speed (*WS*) forecasting studies, where the forecast models from four major categories are outlined in bold fonts. Key: ♣ represents multivariate data, ♠ depicts univariate data, ◇ indicates that the mean absolute percentage error (*MAPE* in %) or mean absolute error (*MAE* in ms^{-1}) are presented as average value(s) (considering all sites, forecast horizons, and seasons), and ♡ conveys the standard *MAPE* or *MAE* of individual site or forecast horizon (mainly for single-step-ahead forecasts).

Model acronym	Method summary	Data type	Study sites	Forecast horizon results	Studies
Physical-based predictive methods					
GPR-**NWP**	NWP data corrected via Gaussian process regression (GPR) model	♣	15 sites in United Kingdom	3, 6, 12, 24, 48, 72 h (*MAPE* = 36.62%, ◇)	[10]
KF-**NWP**	Kalman filter (KP) served as an error compensation mechanism to update NWP	♣	1 wind farm site in China	4, 8, 12, 24 h (*MAPE* = 48.95%, ◇)	[11]
Statistical-based predictive methods					
DWT-**ARMA**	DWT used to decompose data and ARMA used for *WS* prediction	♠	1 site in New Zealand	1 h (*MAPE* = 2.70%, ♡)	[14]
WT-**ARIMA**	ARIMA used to forecast *WS* data decomposed via Repeated WT	♠	1 site in Ireland	1, 3, 5, 7, 10 min (*MAPE* = 3.58%, ◇)	[15]
Machine learning (ML)-based predictive methods					
EMD-**ANN**	EMD decomposed original time series and ANN performed forecasts	♠	1 wind farm site in China	1, 2, 3 h (*MAPE* = 1.33%, ◇)	[19]
WT-GA-**SVR**	WT denoised the time series and SVR optimized by genetic algorithm (GA) performed forecasts	♠	1 wind farm site in China	30 min (*MAPE* = 14.79%, ♡)	[20]
WPT-**ABR**	Wavelet packet transform (WPT) decomposed the data and ABR predicted seasonal *WS*	♣	Seasonal data from 1 site in China	5 min (*MAE* = 0.84 ms^{-1}, ◇)	[21]
EMD-**ENN**	EMD yielded stationary sub-series and ENN obtained accurate forecasts	♠	Seasonal data from 1 site in China	1 h (*MAPE* = 20.0%, ◇)	[44]
Deep learning (DL)-based predictive methods					
CEEMDAN-PE-**GRU**	Combined CEEMDAN and permutation entropy (PE) decomposed data and GRU performed forecasts	♠	1 site in China with high *WS*	15 min, 1 h (*MAPE* = 3.70%, ◇)	[28]
VMD-ICEEMDAN-**LSTM**	VMD and ICEEMDAN performed double decomposition, PACF selected features, and LSTM predicted *WS*	♣	1 site in Pakistan	10 min (*MAPE* = 4.70%, ♡)	[29]
ICEEMDAN-GWO-**LSTM**	ICEEMDAN optimized with GWO used for decomposition and LSTM used for forecasting	♠	5 sites in Turkey	10 h (*MAPE* = 5.26%, ◇)	[30]
FS-BO-**BiLSTM**	Three-stage method used for FS and BiLSTM optimized with BO for *WS* prediction	♣	6 sites in Fiji	10 min (*MAPE* = 14.57%, ◇)	[18]
Boruta-**BiLSTM**	Boruta and BiLSTM used for FS and *WS* prediction, respectively	♣	1 site in United Kingdom	1 h (*MAE* = 0.53 ms^{-1}, ♡)	[31]
CLSTM	Combined CNN and LSTM used as the main forecast model	♣	1 wind farm site in China	12-step-ahead from 5 to 60 min (*MAPE* = 5.32%, ◇)	[33]
SSA-MEMD-**AM-CBiLSTM**	SSA and MEMD performed double decomposition. CNN (optimized via AM) fused with BiLSTM for forecasting	♣	4 sites in United States of America	10 min, 1 h (*MAE* = 0.41 ms^{-1}, ◇)	[34]

$\overrightarrow{A}$ and $\overrightarrow{C}$, which are defined as:

$$\overrightarrow{A} = 2\overrightarrow{\alpha} \cdot \overrightarrow{r} - \overrightarrow{\alpha} \tag{5}$$

$$\overrightarrow{C} = 2 \cdot \overrightarrow{r} \tag{6}$$

where $\overrightarrow{r}$ is a random vector whose coordinates are uniformly distributed between $[0, 1]$ and $\overrightarrow{\alpha}$ is also a vector whose coordinates are linearly decreased from 2 to 0.

Furthermore, the OBL mechanism is employed again to create the opposite positions $\overline{X}_i$, $(i = 1, 2, \ldots, N)$. After which, the OBWOA selects the best N positions from $X \cup \overline{X}$ based on the fitness function values of the updated solution $f(X)$ and the opposite solution $f(\overline{X})$. These steps are repeated until the stop condition is reached or the best solution is obtained.

C. GATED ADDITIVE TREE ENSEMBLE (GATE)

The GATE network [36] is an efficient DL architecture for tabular data. Similar to GRU, GATE uses a gating mechanism as a feature representation learning unit with an integrated FS mechanism. The gated feature learning unit (GFLU) (Figure 2a) is utilized for sequential feature learning [71]. Unlike GRU, weights are not shared between different GFLUs since each unit applies different transformations in each stage of feature representation. For instance, at stage n, the hidden feature representation (H_n) is a linear interpolation between the current candidate feature representation ($\widetilde{H}_n$) and the previous feature representation (H_{n-1}), given as:

$$H_n = (1 - z_n) \odot H_{n-1} + z_n \odot \widetilde{H}_n \tag{7}$$

where $\odot$ denotes element-wise multiplication and z_n is the update gate, which determines how much information is needed to update its internal feature representation, and is defined as:

$$z_n = \sigma(W_n^z \cdot [H_{n-1}; X_n]) \tag{8}$$

where σ is the sigmoid activation function, W_n^z is a learnable parameter, and $[H_{n-1}; X_n]$ denotes a concatenation

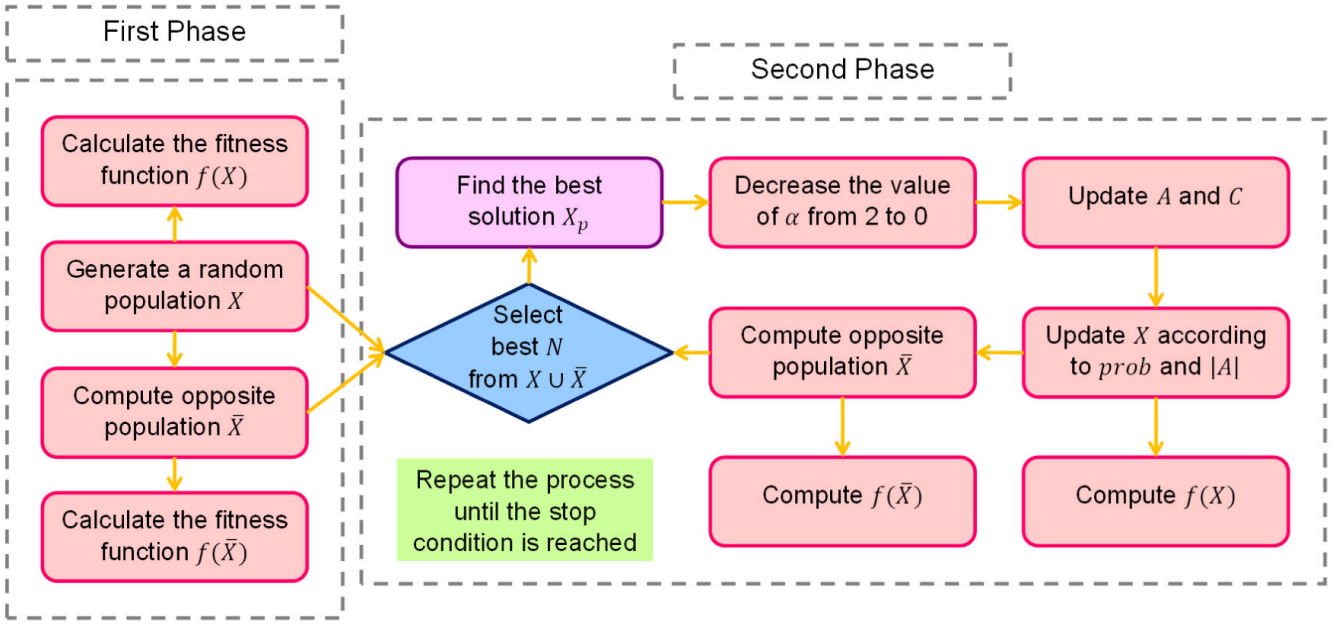


FIGURE 1. A schematic showing the opposition-based whale optimization algorithm (OBWOA) used in the proposed H-GATE predictive framework.

operation between H_{n-1} and X_n. The current candidate feature representation ($\widetilde{H}_n$) and the FS operation (X_n) via masking (M_n) are obtained as follows:

$$\widetilde{H}_n = tanh(W_n^o \cdot [r_n \odot H_{n-1}; X_n]) \tag{9}$$

$$X_n = M_n \odot X \tag{10}$$

where W_n^o is a learnable parameter and r_n is the reset gate that determines how much information is needed to forget from previous feature representation, and is computed as:

$$r_n = \sigma(W_n^r \cdot [H_{n-1}; X_n]) \tag{11}$$

In practice, any number of GFLUs can be ensembled to facilitate hierarchical learning of features. The GFLUs are combined with an ensemble of differentiable, nonlinear DTs. The DT model employs decisions made via a series of hierarchical nodes, where each node acts as a binary classifier that splits the samples either to the left or right leaf. As shown in Figure 2b, the node takes feature (X_i) and learns a cutoff (b) from data to form a hard routing to the left or right leaf. This gating function ($g(X_i, b)$) is called the soft binning function [72]. The overall structure describing this operation using a single feature (Figure 2b) is termed an individual soft decision stump.

The nonlinear DTs in GATE use all the features for a split. Here, instead of a linear combination of features, a nonlinear function is used for a split. A learnable feature mask $\mathcal{M} \in R^{\widetilde{d}}$ with dimension $\widetilde{d}$ is used to combine the outputs o_i of multiple soft decision stumps (Figure 2c). This enables the model to scale effectively to a large number of features and use the information from all variables in making split decisions. This computation is achieved using:

$$o = \left[\sum_{i=0}^{\widetilde{d}} \mathcal{M}_i \odot R_i^L, \sum_{i=0}^{\widetilde{d}} \mathcal{M}_i \odot R_i^R\right] \tag{12}$$

where the output (o) of an individual decision stump is a vector of two leaf values, which is obtained by combining the left and right leaf responses of each variable via a learned feature mask. The feature mask enables the model to determine which feature(s) to give higher priority when generating the predictions.

III. METHODOLOGY

A. STUDY REGION AND DATA DESCRIPTION

The study area is situated in Viti Levu, which is Fiji's largest and most populated island with the highest energy demand. In support of the Paris Agreement commitments and SDG 7, Fiji has set an ambitious target of 100% RE by 2036 [73]. Currently, 65% of this target has been accomplished mainly through hydro- and bio-based RE sources [74]. To meet the actual target, Fiji needs to add more RE capacities in the form of wind since its availability is more compared to solar. This is because, during the wet season (November - April), the availability of solar radiation is uncertain due to extended periods of rainfall.

The nation has a good wind regime, but the current status of wind energy in the electricity mix is negligible, with only one wind farm that has not performed to its full capacity due to poor commissioning [75]. To harness maximum wind energy, *WS* forecasting is an important pre-requisite. Hence, three sites: Rakiraki (RK), Navua (NV), and Sigatoka (SG) (Figure 3) were selected to perform the forecast experiments. RK site has one of the windiest climates in Fiji, which makes

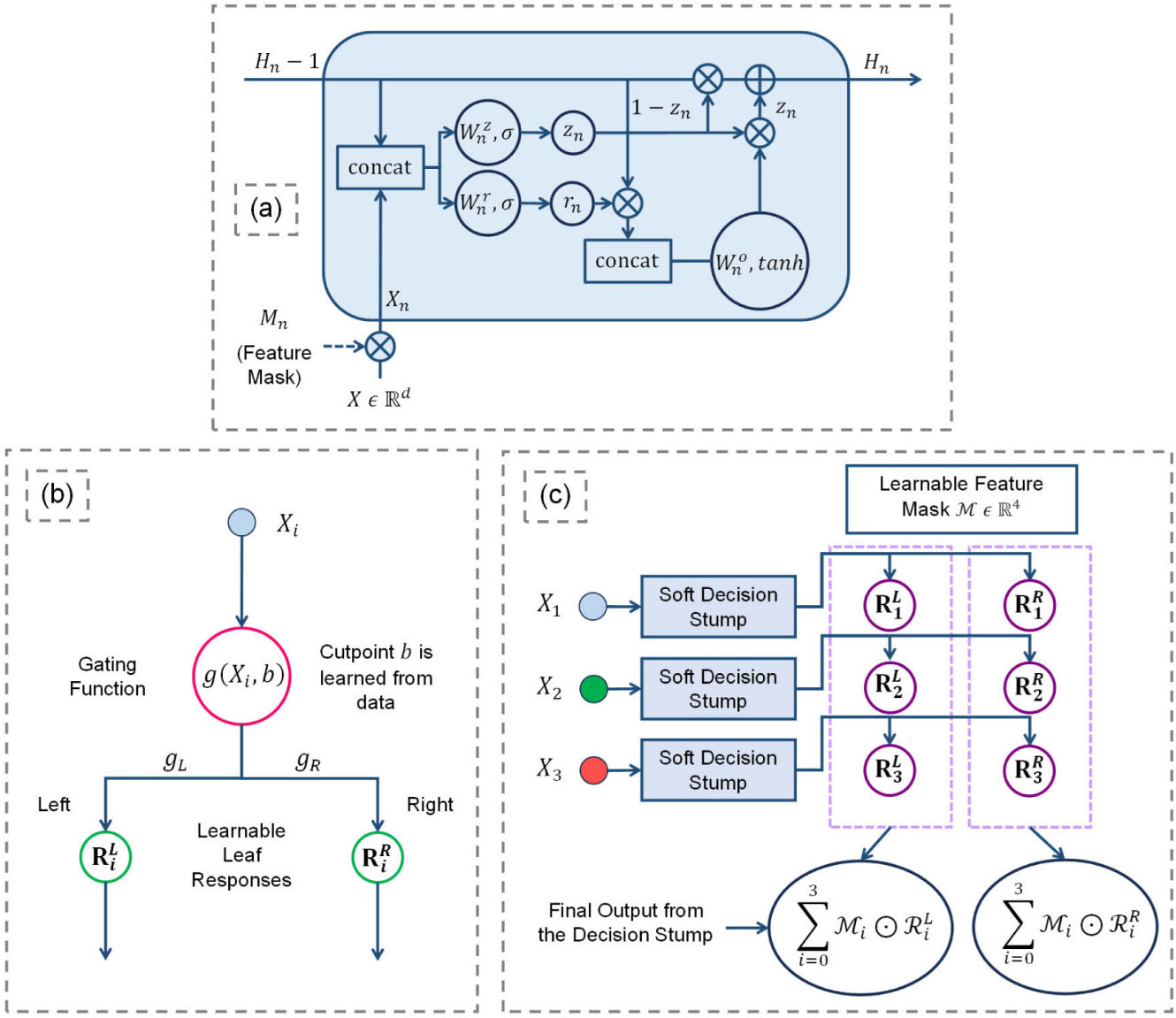


FIGURE 2. Components of gated additive tree ensemble (GATE) model, where (a) is the gated feature learning unit (GFLU), (b) is the soft decision stump for a single feature, and (c) is feature masking and aggregation for multiple soft decision stumps using three features as an example.

it a feasible site for future wind farm commissioning. NV site is about 38 km away from Suva, which is the capital city of Fiji. Although the wind regime of NV is not like RK, its location makes it a good site to explore further. SG is close to the Butoni wind farm site, which makes it a good benchmark site. Another reason for selecting these sites was due to the data availability (Table 2) because most monitored stations had $> 20\%$ of data missing.

The forecast models devised in this research were based on historical data with 1 hour temporal resolution, recorded at a height of 10 m above ground level. The data comprised of wind speed (*WS*), wind direction (*WD*), maximum temperature (*Tmax*), minimum temperature (*Tmin*), relative humidity (*RH*), mean sea-level pressure (*Pmsl*), total rainfall (*Rain*), and total solar radiation (*Radn*). The *Radn* data for RK and SG were not available. All data used were acquired from Fiji Meteorological Services. For more details on these attributes, Table 3 summarizes the statistics. The hourly *WS* data were symmetrically distributed with $0.02 \leq$ skewness ≤ 0.49, which closely depicted a normal distribution at all three sites. The distribution was further elaborated using the kurtosis measure, which revealed that the target *WS* data for all three sites had a platykurtic distribution.

B. CONSTRUCTION OF THE PROPOSED HYBRID WS FORECASTING MODEL

This study develops a hybrid GATE-based (H-GATE) multi-step-ahead *WS* forecasting framework integrated with MEMD for data decomposition, OBWOA for FS, and BO for hyperparameter optimization. All algorithms were implemented using Python under the Google Colaboratory platform. The models: GATE, CLSTM, and ABR were

TABLE 2. Geographical location and general description of three selected wind speed (*WS*) datasets obtained from Fiji Meteorological Services. (Note: The hourly *WS* data are recorded from 01-01-2019 to 31-12-2021.)

Site no.	Site name	Site acronym	Latitude	Longitude	Elevation (m)	Expected data points	% of missing data	Training data split (%)	Testing data split (%)
1	Rakiraki	RK	17.34 °S	178.22 °E	8.1		3.73		
2	Navua	NV	18.22 °S	178.17 °E	6.2	26,304	2.28	66.67	33.33
3	Sigatoka	SG	18.14 °S	177.50 °E	6.7		8.42		

TABLE 3. Statistical description in form of mean (Avg), range (Ran), standard deviation (Std), Skewness (Skew), and Kurtosis (Kurt) of wind speed (*WS*) and other selected meteorological variables used for the three study sites.

	Site 1: RK					Site 2: NV					Site 3: SG				
Variable	Avg	Ran	Std	Skew	Kurt	Avg	Ran	Std	Skew	Kurt	Avg	Ran	Std	Skew	Kurt
WS (ms^{-1})	5.98	14.28	2.89	0.02	-0.87	2.96	8.53	1.75	0.41	-0.84	1.96	5.50	1.25	0.49	-0.70
WD (deg)	153.19	354.83	50.03	1.88	3.50	136.56	351.50	79.09	0.97	0.13	125.02	354.67	66.36	1.30	1.89
Tmax (°C)	26.09	15.57	2.39	0.21	-0.14	25.21	18.22	2.74	0.12	-0.05	24.90	21.42	3.24	-0.02	-0.20
Tmin (°C)	25.53	15.35	2.33	0.19	-0.06	24.65	17.97	2.68	0.08	-0.03	24.28	21.00	3.16	-0.07	-0.16
RH (%)	79.25	41.97	8.56	-0.07	-0.49	83.29	47.18	9.96	-0.57	-0.65	84.07	54.43	11.47	-0.48	-0.84
Pmsl (hpa)	1010.92	21.97	3.63	-0.65	0.41	1011.93	32.69	4.81	-0.74	0.56	1011.72	21.38	3.96	-0.36	-0.36
Rain (mm)	0.04	6.86	0.24	13.50	241.14	0.06	13.25	0.34	13.94	297.90	0.03	8.25	0.24	17.91	419.02
Radn (kWm^{-2})	NA	NA	NA	NA	NA	0.09	0.69	0.14	1.70	2.05	NA	NA	NA	NA	NA

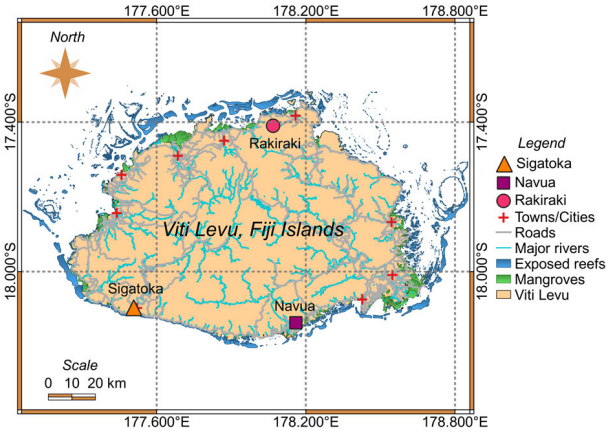


FIGURE 3. Geographical map of Viti Levu, Fiji showing the selected case study sites - Rakiraki (RK), Navua (NV), and Sigatoka (SG).

developed using PyTorch Tabular [36], Tensorflow [76], and Sklearn [77] libraries, respectively. Model optimization via BO was executed using the Optuna library [78]. The overall structure of the proposed H-GATE model is illustrated in Figure 4, and the steps are discussed as follows:

Step 1: The raw datasets had 3.73 - 8.42% missing values (Table 2), which were backfilled with calendar-averaged values [18], [79]. As part of data cleaning, all extreme outliers were substituted with the median values of the respective time series for better model learning. The data were separated into two partitions (Table 2), where 2019 to 2020 data (66.67%) were used for model training and validation; and 2021 data (33.33%) were used for model testing. The training data was also used for model optimization and FS. Data partitioning took place before the decomposition step to prevent the leakage of training data into the testing data [80], [81]. This was to avoid adding unintentional bias to the forecast.

Step 2: The MEMD technique was applied to simultaneously decompose the target (*WS*) and input data (*WD*, *Tmax*, *Tmin*, *RH*, *Pmsl*, *Rain*, and *Radn*) (Figure 5) into respective IMFs and the residual components. Figure 6 shows the decomposed *WS* for RK. To ensure reliable decomposition results, two key parameters: ensemble number (N) and amplitude of the added white noise (ε), were respectively set as 500 and 0.2, based on previous studies [42], [45], [82]. Sites RK and SG had 14 IMFs and a residual component, while NV had 13 IMFs with one residual component (Table 4). Before FS and model development, all IMFs and residual components were pooled into respective sets. For instance, the IMF-1 pool had IMF-1 of *WS* as the target variable and all corresponding IMF-1 sub-series of the predictors (e.g., IMF-1 of *WD* till IMF-1 of *Radn*). The same pooling strategy was carried out for all other IMFs (i.e., IMF-2, . . . ,IMF-n) and the residual component.

Step 3: As part of the feature extraction process, partial auto-correlation function (PACF) and cross-correlation function (CCF) coefficients were used to retrieve the significant lagged inputs of the respective IMF and residual component pools. PACF was used to get the significant lags of the decomposed *WS* variable. CCF was applied to obtain the significant lags of other decomposed meteorological variables. Only 5 lags (i.e., past 5 hours) were considered to limit the number of inputs and because the significance dropped for antecedent lags > 5. The lags which exceeded the 95% confidence band of PACF and CCF plots were considered significant, while the remainder were excluded. The significant inputs of the respective IMF and residual pools were then fed separately to OBWOA.

Training data was used during OBWOA-based FS. For optimal dimensionality reduction, the following OBWOA parameters were utilized: Linearly decreased variable $\alpha = 2$; Constant variable to define the spiral shape $b = 1$; Number of independent runs $NRuns = 10$; Number of iterations $Iter = 60$, and Population size $N \in \{10, 50, 100, 200, 300, 500\}$. A range of values was used to select N, where the best

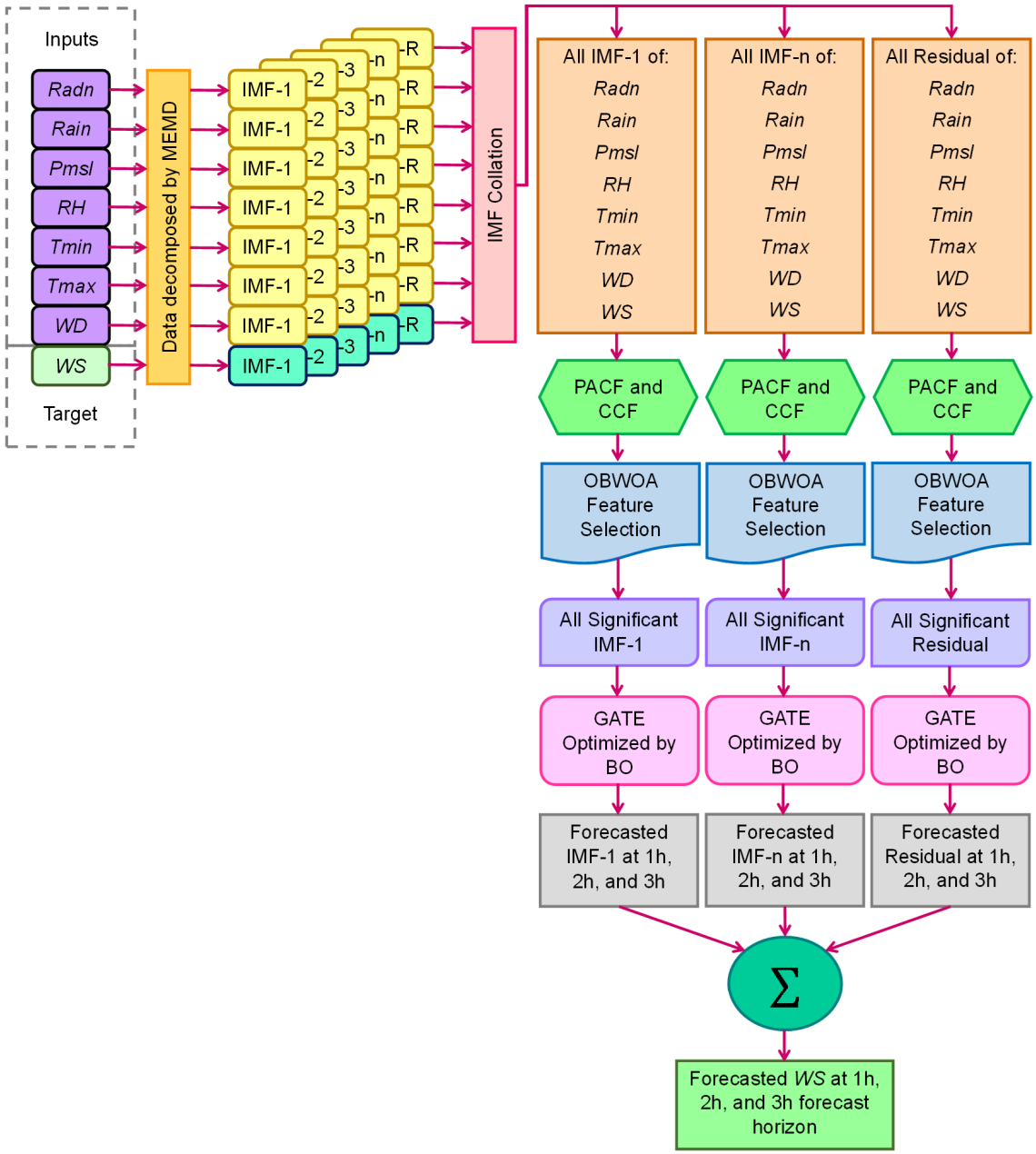


FIGURE 4. Schematic of the proposed multi-stage hybrid gated additive tree ensemble (i.e., H-GATE) model coupled with multivariate empirical model decomposition (MEMD), opposition-based whale optimization algorithm (OBWOA), and Bayesian optimization (BO) used for multi-step-ahead wind speed (*WS*) (ms^{-1}) forecasting.

solution was obtained at the lowest fitness function value (*FV*) (i.e., *RMSE*). The *FV* was determined using a K-nearest neighbour (KNN) regressor with $K = 5$ and a time series split (TSS) 5-fold cross-validation (*CV*). The choice of K is crucial for achieving optimal *FV*. A smaller K (e.g., 1) may lead to overfitting, making KNN sensitive to noise in the data, whereas a larger K (e.g., 10) could introduce bias through excessive smoothing of the decision boundary, resulting in underfitting [83]. Hence, studies [84], [85] on meta-heuristic algorithms for FS advocate for a balanced K value of 5 as a reasonable number of neighbours. The selection of parameters α and b were suggested in [65] for robust results. In addition, there is no general rule for determining optimal values for parameters *NRuns*, *Iter*, and *CV*. Following the suggestion in [85], this study adopted an experimental trial-and-error approach to identify the ideal values for these parameters within the specified search spaces: $NRuns \in \{5,\ 10,\ 15\}$, $Iter \in \{50,\ 60,\ 70\}$, and $CV \in \{5,\ 10\}$.

The FS convergence plots for only IMF-1 and residual are depicted in Figure 7 for RK. The convergence performance of the proposed OBWOA is benchmarked against the standard WOA in terms of the *FV*. For all panels in Figure 7, OBWOA

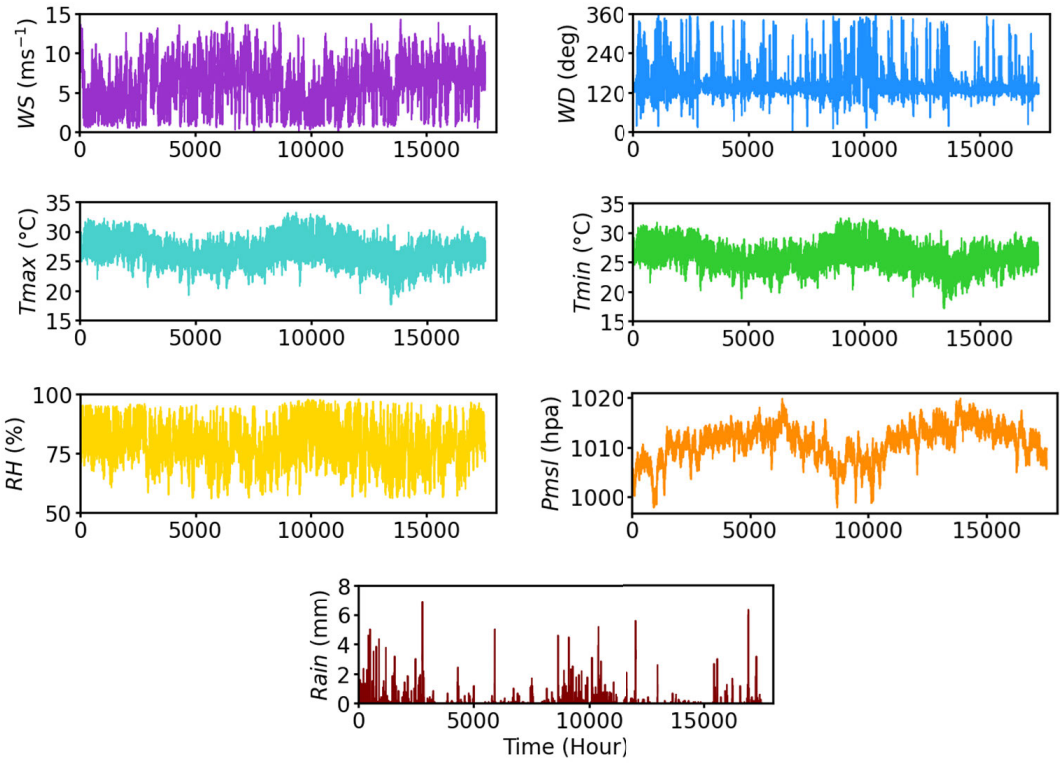


FIGURE 5. Time series plots of the target variable - wind speed (*WS*) and input variables - wind direction (*WD*), maximum temperature (*Tmax*), minimum temperature (*Tmin*), relative humidity (*RH*), mean sea-level pressure (*Pmsl*), and total rainfall (*Rain*) in the training phase for Rakiraki (RK) site as an example.

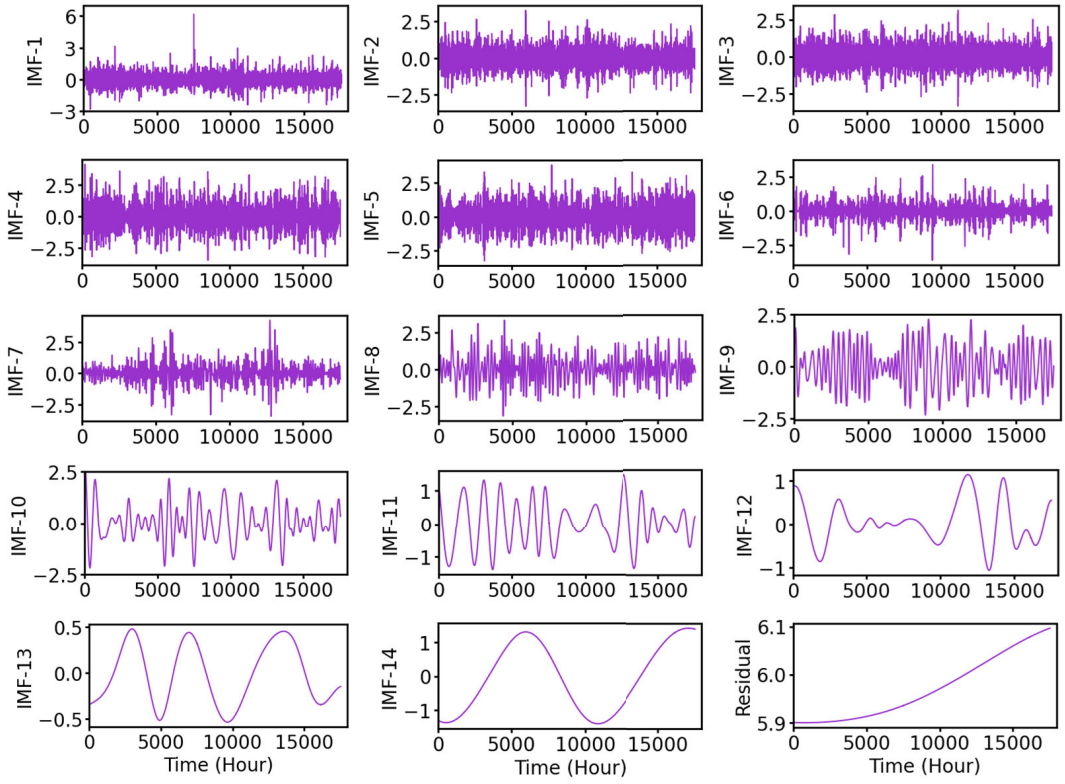


FIGURE 6. An example of time series plots of the decomposed wind speed (*WS*) target variable into several intrinsic mode functions (IMFs) and a residual in the training phase for Rakiraki (RK) site.

TABLE 4. The selected lags of input variables for the undecomposed (i.e., standalone model) and decomposed data (i.e., hybrid model). Decomposed data lags are selected using the partial auto-correlation function (PACF), cross-correlation function (CCF), and the opposition-based whale optimization algorithm (OBWOA), whereas undecomposed data lags are selected only using PACF and CCF. Key: * represents lags selected using PACF, NA = Not Available, and NS = Not Significant.

Site 1: RK	Significant lag numbers (i.e., 1 – 5) based on PACF*, CCF, and OBWOA analysis								
Standalone model: Undecomposed data									
	*WS**	*WD*	*Tmax*	*Tmin*	*RH*	*Pmsl*	*Rain*	*Radn*	Total
Intact time series	1 – 3	1 – 5	2 – 5	2 – 5	1 – 5	1 – 5	1 – 5	NA	31
Hybrid model: Decomposed data									
IMF-1	1, 2	1, 2	1	1	NS	1	1, 2	NA	9
IMF-2	1 – 5	2	2, 5	2, 5	2	2, 3	4	NA	14
IMF-3	1 – 4	1	1, 4, 5	1, 4, 5	4, 5	1 – 3	1	NA	17
IMF-4	1 – 4	1 – 3	1 – 3	1 – 3	1, 2	1, 2	5	NA	18
IMF-5	1, 2, 4	1 – 3	1 – 3	1 – 3	1, 2	1 – 3	NS	NA	17
IMF-6	1, 2, 4, 5	3 – 5	5	5	1 – 3	2	3	NA	14
IMF-7	1, 2	3, 4	1, 2	1, 2	1 – 3	1 – 3	NS	NA	14
IMF-8	1 – 3	1 – 3	5	4	1, 2	1, 2	5	NA	13
IMF-9	1, 3, 4	1 – 3	1	1	1	1 – 3	5	NA	13
IMF-10	1	2 – 5	1, 2	1	1, 2	1	1	NA	12
IMF-11	1	3 – 5	1	1	1 – 3	3 – 5	1, 2	NA	14
IMF-12	1	3 – 5	1, 2	1, 2	4, 5	1, 2	1	NA	13
IMF-13	1, 2	3 – 5	4, 5	3 – 5	4, 5	1 – 3	4, 5	NA	17
IMF-14	1, 2	1 – 4	1 – 3	1 – 3	1 – 4	1 – 3	1 – 3	NA	22
Residual	1, 2	1	1 – 3	1, 2	1	1	1 – 3	NA	13
Site 2: NV									
Standalone model: Undecomposed data									
	*WS**	*WD*	*Tmax*	*Tmin*	*RH*	*Pmsl*	*Rain*	*Radn*	Total
Intact time series	1 – 5	1 – 5	1 – 5	1 – 5	1 – 5	1 – 5	1 – 5	1 – 5	40
Hybrid model: Decomposed data									
IMF-1	1 – 3	2	1	1	1, 2	1	NS	1	10
IMF-2	1 – 4	4	2, 3	2, 3	1, 3	2	1	2, 4	15
IMF-3	1 – 4	5	1, 4, 5	1, 4, 5	1, 5	1 – 3	2	3 – 5	20
IMF-4	1 – 4	2 – 4	1 – 4	1 – 4	1 – 4	2 – 4	2	1, 2	25
IMF-5	1 – 4	2 – 4	1 – 4	1 – 4	1 – 4	1, 2	5	1 – 3	25
IMF-6	1 – 3	1, 2	1 – 3	1 – 3	1 – 3	1	5	1 – 3	19
IMF-7	1 – 4	4, 5	1 – 3	1 – 3	2 – 5	1 – 4	NS	1, 2	22
IMF-8	1 – 3	2 – 5	3 – 5	3 – 5	2 – 5	1 – 4	1	5	23
IMF-9	1 – 3	2 – 5	NS	NS	1 – 4	1 – 4	2 – 4	1	19
IMF-10	1, 3 – 5	5	1	1	1 – 4	1 – 4	5	3 – 5	19
IMF-11	1 – 3	2 – 5	5	5	1 – 5	1 – 4	5	5	20
IMF-12	1	1 – 4	1 – 5	1 – 5	1 – 4	1	1 – 3	1	24
IMF-13	1 – 3	1 – 4	1 – 5	1 – 5	2 – 5	1 – 3	5	1 – 3	28
Residual	1, 2	1 – 5	2 – 5	2 – 5	1 – 4	NS	1 – 5	1 – 3	27
Site 3: SG									
Standalone model: Undecomposed data									
	*WS**	*WD*	*Tmax*	*Tmin*	*RH*	*Pmsl*	*Rain*	*Radn*	Total
Intact time series	1 – 5	1 – 5	1 – 5	1 – 5	1 – 5	NS	NS	NA	25
Hybrid model: Decomposed data									
IMF-1	1, 3	1	1, 2	1	NS	1	NS	NA	7
IMF-2	1 – 4	1, 2	2	2	1	1	2	NA	11
IMF-3	1 – 4	3, 4	1, 4, 5	1, 4, 5	1, 5	1 – 4	1	NA	19
IMF-4	1, 2	3 – 5	1 – 4	1 – 4	1 – 4	1 – 3	NS	NA	20
IMF-5	1, 2	2, 3	1 – 4	1 – 4	1 – 4	1, 2	NS	NA	18
IMF-6	1 – 3	3 – 5	1 – 4	1 – 4	1 – 5	NS	NS	NA	19
IMF-7	1 – 4	5	1	1	1 – 5	NS	NS	NA	12
IMF-8	1, 3 – 5	1 – 5	5	5	1 – 5	NS	5	NA	17
IMF-9	1 – 3	2 – 5	1	1	1 – 5	1	1	NA	16
IMF-10	1	1 – 3	4, 5	3 – 5	1 – 5	1 – 3	NS	NA	17
IMF-11	1, 2	1	1	1	1 – 5	NS	2 – 5	NA	14
IMF-12	1	1, 2	1	1	1 – 5	1, 2	1 – 3	NA	15
IMF-13	1 – 3	1 – 5	1	1	1 – 5	1	5	NA	17
IMF-14	1, 2	1 – 5	1, 2	1 – 3	1 – 5	1 – 4	1, 2	NA	23
Residual	1, 2	1 – 5	1 – 4	1 – 4	1 – 4	NS	1 – 5	NA	24

converged better than WOA during FS as it had the lowest *FV* at every *Iter*. For OBWOA, the *FV* decreased with an increase in N, where the best solution was obtained at $N = 500$ for both IMF-1 and residual. The total features remaining after the application of OBWOA are furnished in Table 4 for all IMFs and the residual.

Step 4: The selected inputs for training and testing phases were normalized between 0 - 1 [42], [79] using:

$$X_n = \frac{X_{act} - X_{min}}{X_{max} - X_{min}} \tag{13}$$

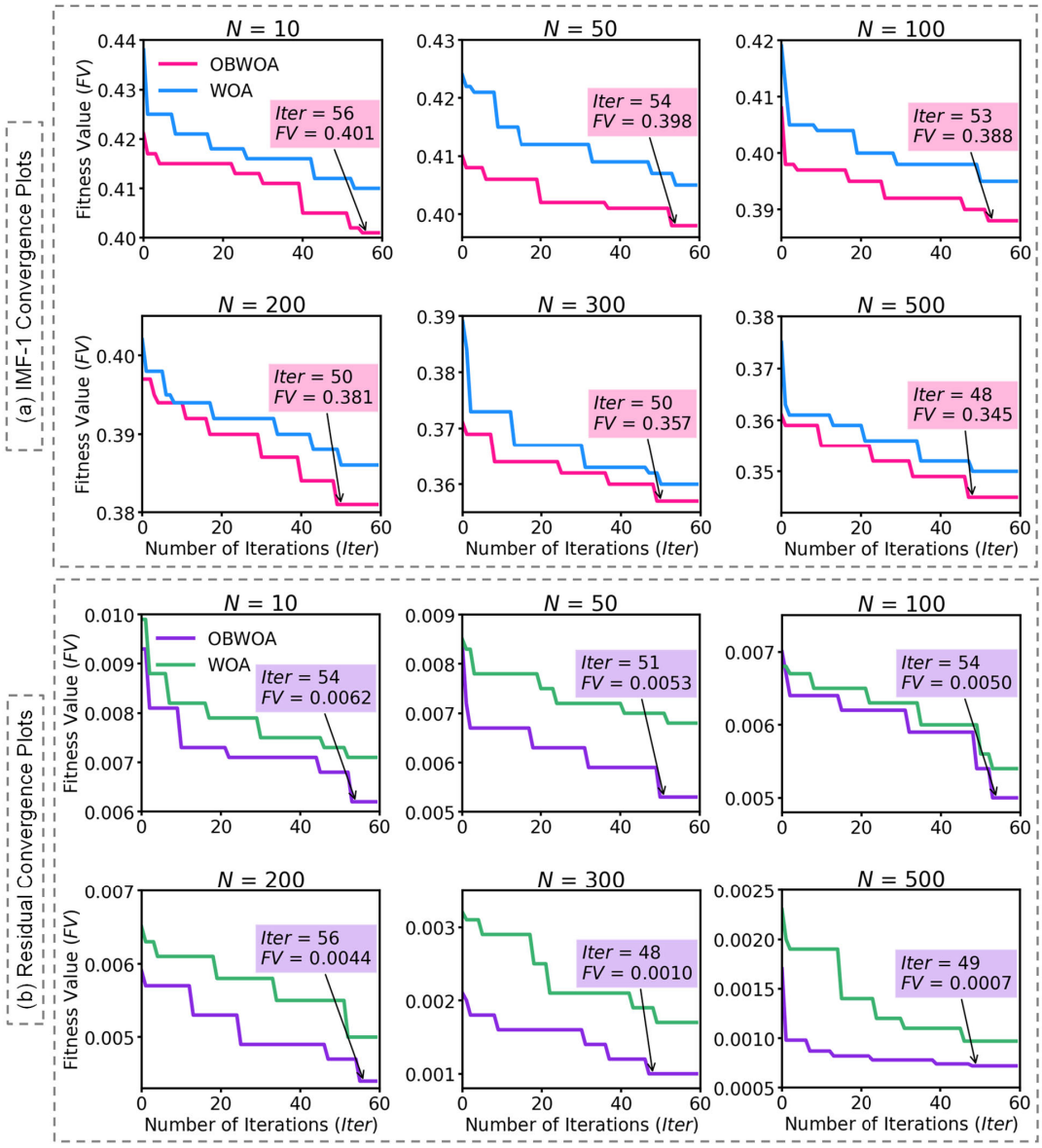


FIGURE 7. Convergence plots for the standard whale optimization algorithm (WOA) and the proposed opposition-based WOA (OBWOA) feature selection on the predictors of the decomposed (a) intrinsic mode function 1 (IMF-1) and the (b) residual component for Rakiraki (RK) site as an example.

where X_n, X_{act}, X_{min}, and X_{max} represent the normalized, actual, minimum, and maximum value of the input.

Step 5: The training data with optimal normalized features were fed to the GATE model for hyperparameter optimization via BO, done using a TSS-based 5-fold *CV* approach. BO effectively selected the best hyperparameters with the lowest computational complexity, which guaranteed optimal forecasting accuracy. The optimal hyperparameters for the standalone models are summarized in Table 5 and the best parameters for the proposed and other hybrid models are furnished in Table 6, where results for only IMF-1 and residual are shown.

The optimized GATE was used to forecast the respective IMF and residual components using a multi-input multi-output (MIMO) multi-step-ahead forecasting strategy [86]. The MIMO approach developed the entire multi-step-ahead prediction sequence in a single go without accumulation of prediction errors. The predicted outputs included a forecast horizon of one-step-ahead (t_{L+1}), two-step-ahead (t_{L+2}), and three-step-ahead (t_{L+3}). Lastly, the multi-step-ahead forecasted *WS* of the respective IMF and residual components were summed to compute the final multi-step-ahead forecasted *WS*.

Step 6: A total of nine statistical metrics and five visual plots were used to assess the predictive accuracy of the proposed H-GATE model against the counterpart hybrid (H-CLSTM and H-ABR) and standalone (S-GATE, S-CLSTM, and S-ABR) models.

TABLE 5. Model hyperparameter search space and the selected hyperparameters obtained via Bayesian optimization (BO) for the undecomposed data (i.e., standalone models).

Predictive models	Model hyperparameters	Hyperparameter acronym	Hyperparameter search space	RK	NV	SG
GATE	Number of layers in the feature abstraction layer	gflu_stages	{3, 4, 5, 6}	4	4	4
	Feature mask function	feature_mask_function	{'entmax', 'sparsemax', 'softmax'}	'sparsemax'	'entmax'	'sparsemax'
	Learning rate	learning_rate	{0.001, 0.01, 0.1, 1.0}	0.001	0.001	0.001
	Head used for the model	head	{'None', 'LinearHead', 'MixtureDensityHead'}	'LinearHead'	'LinearHead'	'None'
	Depth of the tree	tree_depth	{4, 5, 6, 7, 8, 9, 10}	5	7	4
	Number of trees	num_trees	{10, 12, 14, 16, 18, 20}	12	10	12
	Batch size	batch	{256, 512}	512	512	512
	Epochs	epoch	{64, 128}	64	64	128
CLSTM	CNN Filter 1	filter_1	{9, 18, 36, 72, 144}	72	36	72
	CNN Filter 2	filter_2	{9, 18, 36, 72, 144}	36	36	18
	CNN Filter 3	filter_3	{9, 18, 36, 72, 144}	9	18	18
	Hidden neuron 1	neuron_1	{8, 16, 32, 64, 128}	64	64	32
	LSTM cell 1	cell_1	{8, 16, 32, 64, 128}	64	32	32
	LSTM cell 2	cell_2	{8, 16, 32, 64, 128}	16	16	16
	Hidden neuron 2	neuron_2	{8, 16, 32, 64, 128}	32	16	8
	Activation function	activation_function	{'ELU', 'ReLU', 'SELU', 'TanH'}	'ReLU'	'ReLU'	'SELU'
	Batch size	batch	{256, 512}	512	512	512
	Epochs	epoch	{64, 128}	64	64	64
	Optimizer	optimizer	{'Adam', 'Adamax'}	'Adam'	'Adam'	'Adam'
ABR	Maximum number of estimators	max_estimators	{50, 100, 150, 200}	100	150	100
	Learning rate	learning_rate	{0.001, 0.01, 0.1, 1.0}	0.01	0.01	0.001
	Loss function	loss	{'linear', 'square', 'exponential'}	'linear'	'linear'	'linear'

C. MODEL PERFORMANCE EVALUATION

To compare the performance of the H-GATE model against other competing models, a total of nine statistical metrics were used, including Pearson's correlation coefficient (r), Nash-Sutcliffe Efficiency (E_{NS}), Willmott's Index of agreement (WI), Legates and McCabe Index (LM), mean absolute error (MAE; ms^{-1}), root mean square error ($RMSE$; ms^{-1}), absolute percentage bias (APB; %), mean absolute percentage error ($MAPE$; %), and relative root mean square error ($RRMSE$; %). These were computed using:

$$r = \frac{\sum_{i=1}^{N}\left(WS_i^O - \overline{WS}^O\right)\left(WS_i^F - \overline{WS}^F\right)}{\sqrt{\sum_{i=1}^{N}\left(WS_i^O - \overline{WS}^O\right)^2}\sqrt{\sum_{i=1}^{N}\left(WS_i^F - \overline{WS}^F\right)^2}} \tag{14}$$

$$E_{NS} = 1 - \left[\frac{\sum_{i=1}^{N}\left(WS_i^F - WS_i^O\right)^2}{\sum_{i=1}^{N}\left(WS_i^O - \overline{WS}^O\right)^2}\right] \tag{15}$$

$$WI = 1 - \frac{\sum_{i=1}^{N}\left(WS_i^F - WS_i^O\right)^2}{\sum_{i=1}^{N}\left(\left|WS_i^F - \overline{WS}^O\right| + \left|WS_i^O - \overline{WS}^O\right|\right)^2} \tag{16}$$

$$LM = 1 - \left[\frac{\sum_{i=1}^{N}\left|WS_i^F - WS_i^O\right|}{\sum_{i=1}^{N}\left|WS_i^O - \overline{WS}^O\right|}\right] \tag{17}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\left(WS_i^F - WS_i^O\right)\right| \tag{18}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(WS_i^F - WS_i^O\right)^2} \tag{19}$$

$$APB = \frac{\sum_{i=1}^{N}\left(WS_i^F - WS_i^O\right) \times 100}{\sum_{i=1}^{N} WS_i^O} \tag{20}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{WS_i^F - WS_i^O}{WS_i^O}\right| \times 100 \tag{21}$$

$$RRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(WS_i^F - WS_i^O\right)^2}}{\frac{1}{N}\sum_{i=1}^{N}\left(WS_i^O\right)} \times 100 \tag{22}$$

where WS_i^O = observed i^{th} WS, WS_i^F = forecasted i^{th} WS, $\overline{WS}^O$ = average of observed WS, $\overline{WS}^F$ = average of forecasted WS, and N = total number of samples.

Model selection based on diverse metrics can present challenges due to their inherent trade-offs [87]. A model excelling in one metric may be less favourable in another. Consequently, a global performance indicator (GPI) [88] was employed to amalgamate the results of nine metrics into a unified index, eliminating the need for analysing individual metrics. The GPI measure assigns equal weights to all statistical metrics involved in the computation, where a larger GPI indicates a more accurate model. For i^{th} model, the GPI is computed as follows:

$$GPI = \sum_{j=1}^{N}\alpha_j\left(\overline{y_j} - y_{ij}\right) \tag{23}$$

where N is the total number of metrics used, $\alpha_j = -1$ for metrics with ideal value = 1 (i.e., r, E_{NS}, WI, and LM) and $+1$ for metrics with ideal value = 0 (i.e., MAE, $RMSE$, APB,

TABLE 6. Model hyperparameter search space and the selected hyperparameters obtained via Bayesian optimization (BO) for the decomposed data (i.e., hybrid models). (Note: The selected hyperparameters of only intrinsic mode function 1 (IMF-1) and residual components are summarized as an example.)

Predictive models	Hyperparameter acronym	RK		NV		SG	
		IMF-1	Residual	IMF-1	Residual	IMF-1	Residual
GATE	gflu_stages	4	4	3	4	4	3
	feature_mask_function	'sparsemax'	'sparsemax'	'sparsemax'	'entmax'	'sparsemax'	'entmax'
	learning_rate	0.001	0.01	0.001	0.001	0.01	0.001
	head	'LinearHead'	'None'	'LinearHead'	'LinearHead'	'None'	'None'
	tree_depth	5	7	5	5	4	5
	num_trees	10	14	12	12	10	10
	batch	512	512	512	512	512	256
	epoch	64	64	64	128	64	64
CLSTM	filter_1	72	144	72	36	72	72
	filter_2	36	72	18	18	36	36
	filter_3	18	36	18	18	36	18
	neuron_1	64	64	32	64	32	128
	cell_1	64	64	32	32	16	64
	cell_2	32	16	16	32	8	16
	neuron_2	32	32	8	16	16	32
	activation_function	'SELU'	'SELU'	'ReLU'	'ReLU'	'SELU'	'ReLU'
	batch	512	512	512	256	512	512
	epoch	64	64	64	64	128	64
	optimizer	'Adam'	'Adam'	'Adam'	'Adam'	'Adam'	'Adam'
ABR	max_estimators	100	100	150	100	200	100
	learning_rate	0.01	0.01	0.001	0.001	0.001	0.01
	loss	'linear'	'linear'	'linear'	'linear'	'linear'	'linear'

MAPE, and *RRMSE*), y_{ij} is the scaled value of metric j for model i, and $\overline{y_j}$ is the median of scaled values of metric j.

The performance metrics and *GPI* only assess the forecast models' predictions against the observed *WS* data. These measures do not offer a detailed distinction between the forecast models. Hence, a two-sample t-test [89] was conducted to ascertain whether a significant difference exists between the means of the proposed and the benchmark models. The null hypothesis (H_0) of this statistical test indicates that the mean performance of the proposed model is equal to that of the benchmark model, while the alternative hypothesis (H_1) suggests the presence of a significant difference. The t-statistic for a two-sample t-test is calculated as follows:

$$t\text{-statistic} = \frac{\overline{WS}_1^F - \overline{WS}_2^F}{Std_{pool}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \tag{24}$$

where $\overline{WS}_1^F$ and $\overline{WS}_2^F$ are the means of forecasted *WS* using model 1 (i.e., H-GATE) and model 2 (i.e., a benchmark model), N_1 and N_2 are the sample sizes, and Std_{pool} is the pooled standard deviation, which is given as:

$$Std_{pool} = \sqrt{\frac{\left((N_1 - 1)\, Std_1^2\right) + \left((N_2 - 1)\, Std_2^2\right)}{N_1 + N_2 - 2}} \tag{25}$$

where Std_1^2 and Std_2^2 are the squared standard deviations of forecasted *WS* using models 1 and 2.

The resulting t-statistic is compared with the critical value from the t-distribution table to determine a two-tailed p-value. The null hypothesis is rejected if the p-value is less than the significance level, indicating a statistically significant difference in performance between the two models. This statistical test is conducted at a significance level of 0.05, which is the common choice in most studies [90], [91].

IV. RESULTS AND DISCUSSION

This section presents and discusses the performance of the proposed H-GATE model against comparative models in the testing phase.

A. PERFORMANCE OF H-GATE BASED ON STATISTICAL METRICS

The merits of the proposed H-GATE model were rigorously assessed against benchmark models using nine statistical metrics. Table 7 evaluates the model performances in terms of *r*, *MAE*, and *RMSE* measures. The *r* indicator captured the degree of variance between the observed (WS^O) and forecasted (WS^F) data [92], while the error metrics *MAE* and *RMSE* helped determine the model biases [93]. Table 7 shows that the best results are obtained for a one-step-ahead forecast horizon, where H-GATE registered the highest *r* (0.976 - 0.994), and the lowest *MAE* (0.160 - 0.261) and *RMSE* (0.215 - 0.355) for all three sites. The predictive performance slightly decreases at higher forecast horizons (e.g., two-step-ahead and three-step-ahead). Since the MIMO strategy was adopted, 5 significant lags of all selected predictors were considered as model inputs and 3 leads of *WS* were used as multi-outputs. The correlation of *WS* at higher leads was lower with the predictors at higher lags given the stochastic nature of *WS*. Hence, the forecasting accuracy of *WS* at two-step-ahead and three-step-ahead is lower compared to one-step-ahead WS^F in terms of *r*, *MAE*, and *RMSE* indicators. However, these metrics have a few limitations. For example, *r* is scale and offset invariant; hence, can allocate higher correlation scores to mediocre models [94]. Also, both *MAE* and *RMSE* are non-normalized error measures expressed in absolute units; hence, are unreliable to gauge model performances in geographically

TABLE 7. Multi-step-ahead forecasting test performance of the proposed hybrid gated additive tree ensemble (i.e., H-GATE) model against other comparative models in terms of correlation coefficient (*r*), mean absolute error (*MAE*) in ms^{-1}, and root mean squared error (*RMSE*) in ms^{-1}. The best results are shown in italics. (Key: t_{L+1} = one-step-ahead, t_{L+2} = two-step-ahead, and t_{L+3} = three-step-ahead forecast horizon.)

Sites	Models	t_{L+1}			t_{L+2}			t_{L+3}		
		r	*MAE*	*RMSE*	*r*	*MAE*	*RMSE*	*r*	*MAE*	*RMSE*
RK	*H-GATE*	*0.994*	*0.261*	*0.355*	*0.978*	*0.375*	*0.506*	*0.953*	*0.502*	*0.677*
	H-CLSTM	0.948	0.623	0.810	0.929	0.700	0.909	0.904	0.793	1.027
	H-ABR	0.920	0.841	1.054	0.903	0.905	1.130	0.879	0.984	1.229
	S-GATE	0.904	0.975	1.227	0.889	1.025	1.288	0.863	1.104	1.388
	S-CLSTM	0.855	1.223	1.519	0.837	1.276	1.584	0.810	1.345	1.670
	S-ABR	0.815	1.424	1.755	0.801	1.464	1.801	0.776	1.525	1.873
NV	H-GATE	0.986	0.188	0.256	0.961	0.271	0.368	0.922	0.364	0.493
	H-CLSTM	0.943	0.417	0.530	0.908	0.492	0.631	0.864	0.570	0.735
	H-ABR	0.911	0.580	0.723	0.884	0.632	0.791	0.841	0.700	0.883
	S-GATE	0.863	0.772	0.937	0.832	0.823	1.002	0.789	0.885	1.081
	S-CLSTM	0.828	0.918	1.109	0.803	0.954	1.156	0.761	1.010	1.230
	S-ABR	0.816	1.035	1.232	0.787	1.077	1.286	0.743	1.133	1.360
SG	H-GATE	0.976	0.160	0.215	0.939	0.231	0.310	0.889	0.302	0.405
	H-CLSTM	0.940	0.330	0.409	0.897	0.392	0.492	0.843	0.456	0.576
	H-ABR	0.897	0.491	0.592	0.861	0.534	0.649	0.808	0.589	0.723
	S-GATE	0.850	0.650	0.765	0.802	0.701	0.831	0.742	0.755	0.903
	S-CLSTM	0.821	0.782	0.906	0.775	0.827	0.964	0.717	0.880	1.030
	S-ABR	0.781	0.914	1.057	0.740	0.951	1.105	0.681	0.998	1.168

different sites with different *WS* [95]. Therefore, more robust percentage error measures (*APB*, *MAPE*, and *RRMSE*) were used to reliably evaluate model bias.

APB gave the error of predicted values as a percentage, where lower values closer to zero are preferred [96]. *MAPE* and *RRMSE* also used percentage criteria to categorize the models as excellent (bias $<$ 10%), good (10% $<$ bias $<$ 20%), fair (20% $<$ bias $<$ 30%), and poor (bias $\geq$ 30%) [97]. *APB* in Table 8 shows that the proposed H-GATE model recorded the lowest error at all three forecast horizons i.e., the error range for the respective forecast horizons was between 4.673 - 8.196% (t_{L+1}), 6.701 - 11.818% (t_{L+2}), and 8.973 - 15.461% (t_{L+3}) for the three studied stations. *MAPE* in Table 8 classifies the proposed H-GATE as excellent for RK (at t_{L+1} and t_{L+2}), NV (at t_{L+1}), and SG (at t_{L+1}); and good for RK (at t_{L+3}), NV (at t_{L+2} and t_{L+3}), and SG (at t_{L+2} and t_{L+3}). For the accurately predicted one-step-ahead *WS*, the optimal H-GATE achieved 73.35%, 67.74%, and 63.45% reduction in *MAPE* over the worst-performing S-ABR for RK, NV, and SG sites, respectively. Furthermore, *RRMSE* categorizes H-GATE as excellent for RK (at t_{L+1} and t_{L+2}) and NV (at t_{L+1}); good for RK (at t_{L+3}), NV (at t_{L+2} and t_{L+3}), and SG (at t_{L+1} and t_{L+2}); and fair for SG (at t_{L+3}) (Table 8). The proposed H-GATE registered a high percentage reduction in *RRMSE* over S-ABR at all three forecast horizons for all sites. The percentage-based error measures in Table 8 validate the superior predictive performance of the GATE model over CLSTM and ABR models. Additional improvements to the GATE algorithm were achieved through the MEMD of the predictor signals and OBWOA-based FS of optimal inputs.

Table 9 further demonstrates the efficacy of the proposed H-GATE against other benchmark models using E_{NS}, *WI*, and *LM* measures. Similar to the coefficient of determination (R^2), E_{NS} measures how well the plot of WS^O versus WS^F fits the 1:1 line. This scaled variant of *MSE* helped determine the relative magnitude of the residual variance compared to the observed data variance [98].

As presented in Table 9, H-GATE obtained the highest E_{NS} at all three forecast horizons for all sites. Better results were shown for one-step-ahead *WS* forecasting with E_{NS} being 0.989 (for RK), 0.973 (for NV), and 0.952 (for SG). E_{NS} is a commonly applied criterion for model assessment, but it overestimates the larger *WS* values while neglecting the lower values [94]. *WI* was designed to resolve this issue by taking into account the ratio of *MSE* instead of the differences [99], [100].

The *WI* in Table 9 reconfirms the superior predictive performance of H-GATE at multi-step-ahead forecast horizons; for instance, *WI* = (t_{L+1}: 0.986 - 0.996, t_{L+2}: 0.976 - 0.987, and t_{L+3}: 0.962 - 0.980) for all three sites. While *WI* is generally favoured over E_{NS}, it has a limitation. It is oversensitive to peak residuals and the squaring of residuals often leads to mediocre models registering large *WI* values like *r* metric (e.g., $\geq$ 0.65) [101].

Moreover, the *LM* measure is not overestimated like E_{NS} and *WI* as it allocates appropriate weights to errors and discrepancies [94]. This metric is insensitive to extreme *WS* outliers as it is not inflated by the squared values. Hence, *LM* is one of the most reliable statistical metrics, which justified in Table 9 that H-GATE hybridized with MEMD and OBWOA accomplished the best predictive performance. Thus, when benchmarked against the standalone S-GATE, the hybrid H-GATE recorded 24.3%, 29.27%, and 30.72% (at t_{L+1}), 21.15%, 27.43%, and 29.86% (at t_{L+2}), and 19.36%, 25.71%, and 29.62% (at t_{L+3}) improvements in *LM* for RK, NV, and SG, respectively.

Moreover, the results from all nine metrics, as furnished in Tables 7, 8, and 9, consistently indicate a preference for the proposed H-GATE model in multi-step-ahead *WS*

TABLE 8. Multi-step-ahead forecasting test performance of the proposed hybrid gated additive tree ensemble (i.e., H-GATE) model against other comparative models in terms of absolute percentage bias (*APB* in %), mean absolute percentage error (*MAPE* in %), and relative root mean square error (*RRMSE* in %). The best results are shown in italics. (Key: t_{L+1} = one-step-ahead, t_{L+2} = two-step-ahead, and t_{L+3} = three-step-ahead forecast horizon.)

Sites	Models	t_{L+1}			t_{L+2}			t_{L+3}		
		APB	*MAPE*	*RRMSE*	*APB*	*MAPE*	*RRMSE*	*APB*	*MAPE*	*RRMSE*
RK	*H-GATE*	*4.673*	*6.134*	*6.351*	*6.701*	*8.668*	*9.045*	*8.973*	*11.599*	*12.104*
	H-CLSTM	9.347	12.000	12.701	10.731	13.695	14.461	12.388	15.863	16.579
	H-ABR	11.464	14.349	15.264	12.604	15.732	16.623	14.023	17.602	18.390
	S-GATE	12.067	16.656	16.576	12.967	17.519	17.658	14.373	19.323	19.463
	S-CLSTM	14.715	20.176	20.013	15.666	21.217	21.162	16.889	22.818	22.701
	S-ABR	16.523	23.018	22.438	17.245	23.586	23.255	18.329	24.973	24.556
NV	H-GATE	6.419	7.967	8.748	9.246	11.305	12.556	12.421	15.221	16.825
	H-CLSTM	10.803	13.127	14.677	13.391	16.209	18.136	16.035	19.504	21.661
	H-ABR	12.986	17.170	17.836	14.758	19.006	20.185	17.070	21.892	23.305
	S-GATE	16.124	18.836	21.735	17.854	20.787	23.960	19.981	23.455	26.673
	S-CLSTM	17.669	24.384	24.208	18.906	25.183	25.811	20.816	27.434	28.334
	S-ABR	18.271	24.698	24.984	19.689	25.955	26.828	21.605	28.260	29.354
SG	H-GATE	8.196	9.927	11.011	11.818	14.072	15.848	15.461	18.368	20.711
	H-CLSTM	11.748	13.921	15.782	14.942	17.537	20.021	18.204	21.376	24.348
	H-ABR	14.869	18.667	20.022	17.085	21.002	22.971	19.907	24.330	26.741
	S-GATE	17.916	21.075	23.779	20.476	23.834	27.159	23.245	27.036	30.807
	S-CLSTM	19.514	21.735	25.890	21.832	24.148	28.818	24.541	27.269	32.220
	S-ABR	21.170	27.162	28.460	23.077	28.876	30.944	25.480	31.577	34.142

TABLE 9. Multi-step-ahead forecasting test performance of the proposed hybrid gated additive tree ensemble (i.e., H-GATE) model against other comparative models in terms of Nash-Sutcliffe coefficient (E_{NS}), Willmott's Index (*WI*), and Legates and McCabe index (*LM*). The best results are shown in italics. (Key: t_{L+1} = one-step-ahead, t_{L+2} = two-step-ahead, and t_{L+3} = three-step-ahead forecast horizon.)

Sites	Models	t_{L+1}			t_{L+2}			t_{L+3}		
		E_{NS}	*WI*	*LM*	E_{NS}	*WI*	*LM*	E_{NS}	*WI*	*LM*
RK	*H-GATE*	*0.989*	*0.996*	*0.890*	*0.957*	*0.987*	*0.842*	*0.909*	*0.980*	*0.789*
	H-CLSTM	0.898	0.964	0.780	0.863	0.959	0.747	0.816	0.952	0.708
	H-ABR	0.846	0.946	0.730	0.815	0.941	0.703	0.773	0.934	0.670
	S-GATE	0.816	0.933	0.716	0.791	0.928	0.695	0.745	0.921	0.661
	S-CLSTM	0.731	0.909	0.653	0.700	0.904	0.631	0.657	0.897	0.602
	S-ABR	0.664	0.889	0.611	0.641	0.884	0.594	0.603	0.878	0.568
NV	H-GATE	0.973	0.989	0.870	0.924	0.982	0.813	0.850	0.972	0.748
	H-CLSTM	0.889	0.962	0.781	0.824	0.952	0.729	0.747	0.940	0.675
	H-ABR	0.830	0.944	0.737	0.781	0.937	0.701	0.708	0.925	0.654
	S-GATE	0.745	0.919	0.673	0.692	0.910	0.638	0.623	0.897	0.595
	S-CLSTM	0.686	0.903	0.642	0.645	0.895	0.617	0.579	0.883	0.578
	S-ABR	0.666	0.889	0.630	0.619	0.880	0.601	0.551	0.868	0.562
SG	H-GATE	0.952	0.986	0.834	0.882	0.976	0.761	0.790	0.962	0.687
	H-CLSTM	0.883	0.961	0.763	0.804	0.949	0.698	0.710	0.934	0.632
	H-ABR	0.804	0.941	0.699	0.741	0.930	0.655	0.653	0.916	0.598
	S-GATE	0.723	0.916	0.638	0.643	0.901	0.586	0.550	0.883	0.530
	S-CLSTM	0.674	0.895	0.606	0.601	0.881	0.559	0.514	0.863	0.504
	S-ABR	0.610	0.879	0.572	0.547	0.867	0.534	0.463	0.849	0.485

forecasting, showcasing superior accuracy. However, while these performance metrics are beneficial in evaluating the forecast models' predictions against the observed *WS* data, they lack detailed distinctions between different predictive models. For instance, as evidenced in Table 7, the r values of the proposed H-GATE and the second-best H-CLSTM model are close, revealing slight differences of 4.85% (RK), 4.56% (NV), and 3.83% (SG) at a one-step-ahead forecast horizon. To further justify the H-GATE model's predictive superiority, the two-sample t-test results (i.e., t-statistic and p-value) are summarized in Table 10 to assess the presence of significant mean differences between the proposed H-GATE and benchmark models.

The t-statistic values are employed to derive corresponding p-values, where a p-value $\leq$ the chosen significance level results in the rejection of the null hypothesis, indicating no statistical significance between the means of the proposed and benchmark models. The choice of the significance level is pivotal, influencing the decision to accept or reject the null hypothesis. A lower significance level of 0.01 is generally preferred for large samples with low variance to bolster result confidence, and a higher significance level of 0.1 is favoured for small samples with high variance to enhance test sensitivity [102]. Since the *WS* data used in this study is neither too large nor too small, a significance level of 0.05 was deemed appropriate, which is also a commonly used value in most studies [90], [91].

The p-values summarized in Table 10 are less than 0.05, rejecting the null hypothesis for all tests conducted at multi-step-ahead forecast horizons. It indicates that the mean of WS^F obtained using H-GATE is different from the means of WS^F retrieved using all benchmark models evaluated

TABLE 10. Results of a two-sample t-test evaluating the statistical significance of the differences between the proposed model and comparative model predictions at multi-step-ahead forecast horizons. (Note: All tests are performed at a 5% statistical significance level, where a p-value < 5% leads to rejection of the null hypothesis (H_O). Key: t_{L+1} = one-step-ahead, t_{L+2} = two-step-ahead, and t_{L+3} = three-step-ahead forecast horizon.)

Sites	Two-sample t-test	H-GATE vs. H-CLSTM	H-GATE vs. H-ABR	H-GATE vs. S-GATE	H-GATE vs. S-CLSTM	H-GATE vs. S-ABR
t_{L+1}						
RK	t-statistic	2.008	2.358	2.483	2.557	2.682
	p-value	0.045	0.018	0.013	0.011	0.007
	H_O	Reject	Reject	Reject	Reject	Reject
NV	t-statistic	2.034	2.442	2.599	2.787	2.927
	p-value	0.042	0.015	0.009	0.005	0.003
	H_O	Reject	Reject	Reject	Reject	Reject
SG	t-statistic	2.054	2.514	2.738	2.960	3.341
	p-value	0.040	0.012	0.006	0.003	0.001
	H_O	Reject	Reject	Reject	Reject	Reject
t_{L+2}						
RK	t-statistic	2.142	2.493	2.618	2.692	2.817
	p-value	0.032	0.013	0.009	0.007	0.005
	H_O	Reject	Reject	Reject	Reject	Reject
NV	t-statistic	2.179	2.588	2.746	2.950	3.388
	p-value	0.029	0.010	0.006	0.003	0.001
	H_O	Reject	Reject	Reject	Reject	Reject
SG	t-statistic	2.221	2.681	2.905	3.127	3.508
	p-value	0.026	0.007	0.004	0.002	0.000
	H_O	Reject	Reject	Reject	Reject	Reject
t_{L+3}						
RK	t-statistic	2.277	2.628	2.753	2.826	2.952
	p-value	0.023	0.009	0.006	0.005	0.003
	H_O	Reject	Reject	Reject	Reject	Reject
NV	t-statistic	2.325	2.731	2.890	3.077	3.534
	p-value	0.020	0.006	0.004	0.002	0.000
	H_O	Reject	Reject	Reject	Reject	Reject
SG	t-statistic	2.388	2.848	3.072	3.495	3.675
	p-value	0.017	0.004	0.002	0.000	0.000
	H_O	Reject	Reject	Reject	Reject	Reject

in this study. Another observation is that the t-statistic increases and p-values drop as the comparison extends from H-GATE (best-performing) vs. CLSTM (second-best) to H-GATE vs. ABR (worst-performing model) (Table 10). This demonstrates that the mean difference gets higher as benchmark model performance deteriorates. Therefore, the two-sample t-test thoroughly substantiates the superior predictive ability of H-GATE, even if few statistical metrics (e.g., r) register close results between H-GATE and counterpart models.

B. PERFORMANCE OF H-GATE BASED ON VISUAL PLOTS

For a more detailed view of model performances, five visual plots were used in addition to numerical model assessment. Only the one-step-ahead forecast horizon results are visualized based on their excellent accuracy. The Taylor plots in Figure 8 display the r and the standard deviation values on the radial and polar axes, respectively. These plots portray the magnitude of statistical correlation and the root mean square (RMS) centre difference between the WS^O and WS^F.

The proposed H-GATE model registered the highest r and the lowest centred RMS difference compared to the benchmark models for all three sites (Figure 8). Also, the r to standard deviation points for the benchmark models were highly deviated from the ideal WS^O point for all sites. The proposed H-GATE model had the closest r and standard deviation to the actual WS^O, which confirms its superior forecasting ability compared to the comparative models.

Similar to r, R^2 measures the strength of the relationship between the WS^O and WS^F. Figure 9 shows the scatter plots of the WS^O and one-step-ahead WS^F with the R^2 scores (ideal value $= +1$) for all three sites. The proposed H-GATE reported optimal R^2 of 0.989, 0.973, and 0.952 for sites RK, NV, and SG, respectively (Figure 9). This revealed that H-GATE had the lowest variance between WS^O and WS^F in comparison to other benchmark models. Therefore, H-GATE at all three sites demonstrated better and more reliable forecasting competence.

Absolute forecasting error $|FE|$, which is the absolute values of the differences between WS^O and WS^F were computed and the distributions are illustrated as box plots (Figure 10), ECDF plots (Figure 11), and histograms (Figure 12). The box plots in Figure 10 summarize important statistical information including the upper quartile, median, and lower quartile of the $|FE|$. The points above and below the whiskers represent the outliers of the extreme $|FE|$ in the test phase.

The distribution of the $|FE|$ attained by the proposed H-GATE exhibited the lowest interquartile range (IQR) at sites RK (0.283), NV (0.204), and SG (0.175). This indicates that H-GATE had the least $|FE|$ range compared to the competing models. The ECDF plots of $|FE|$ were also

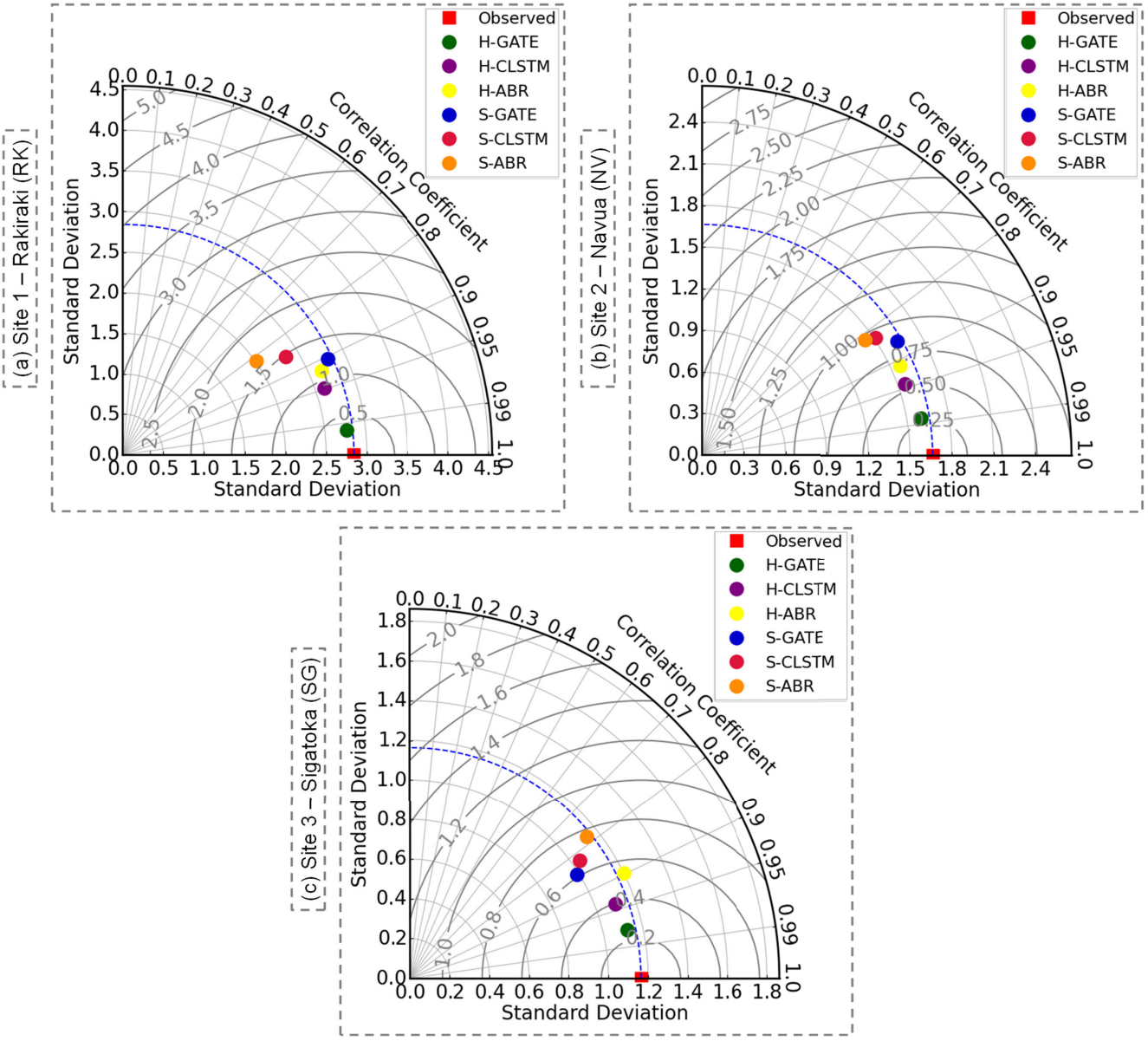


FIGURE 8. Taylor plots for one-step-ahead (t_{L+1}) forecasted wind speed (*WS*) indicating the correlation coefficient (*r*) and standard deviation (*Std*) in the testing phase generated by the hybrid gated additive tree ensemble (i.e., H-GATE) against other comparative models implemented at (a) Site 1 - Rakiraki, (b) Site 2 - Navua, and (c) Site 3 - Sigatoka.

displayed in Figure 11 to get additional details on the $|FE|$ distributions. All hybrid models (i.e., H-GATE, H-CLSTM, and H-ABR) outperformed the standalone counterparts.

When compared to all benchmark models, the proposed H-GATE had the least $|FE|$ range for all three sites. Additionally, Figure 12 shows the probability distribution of $|FE|$ yielded in error brackets of 0.5 step-sizes for all three sites. Similar to the information shown in Figures 10 and 11, the proposed H-GATE acquired the narrowest spreads in $|FE|$ over other competing models for all sites (Figure 12). For instance, H-GATE accumulated 85.8%, 92.9%, and 96.6% errors in the first bin ($0 \leq |FE| \leq 0.5$) for RK, NV, and SG, respectively. The higher proportion of errors in the lowest error band revealed that the combination of GATE with MEMD and OBWOA successfully outperformed the counterpart models for multi-step-ahead *WS* forecasting.

C. SEASONAL PERFORMANCE OF H-GATE

H-GATE exhibits strong performance in *WS* forecasting, as evidenced by various statistical metrics and visual plots. To further enhance its predictive capabilities, it is essential to incorporate seasonal forecasts, specifically for Fiji's dry (May - October) and wet (November - April) seasons. Seasonal forecasting becomes critical due to the country's susceptibility to climatic extremes, particularly cyclones during the wet season, impacting wind turbine

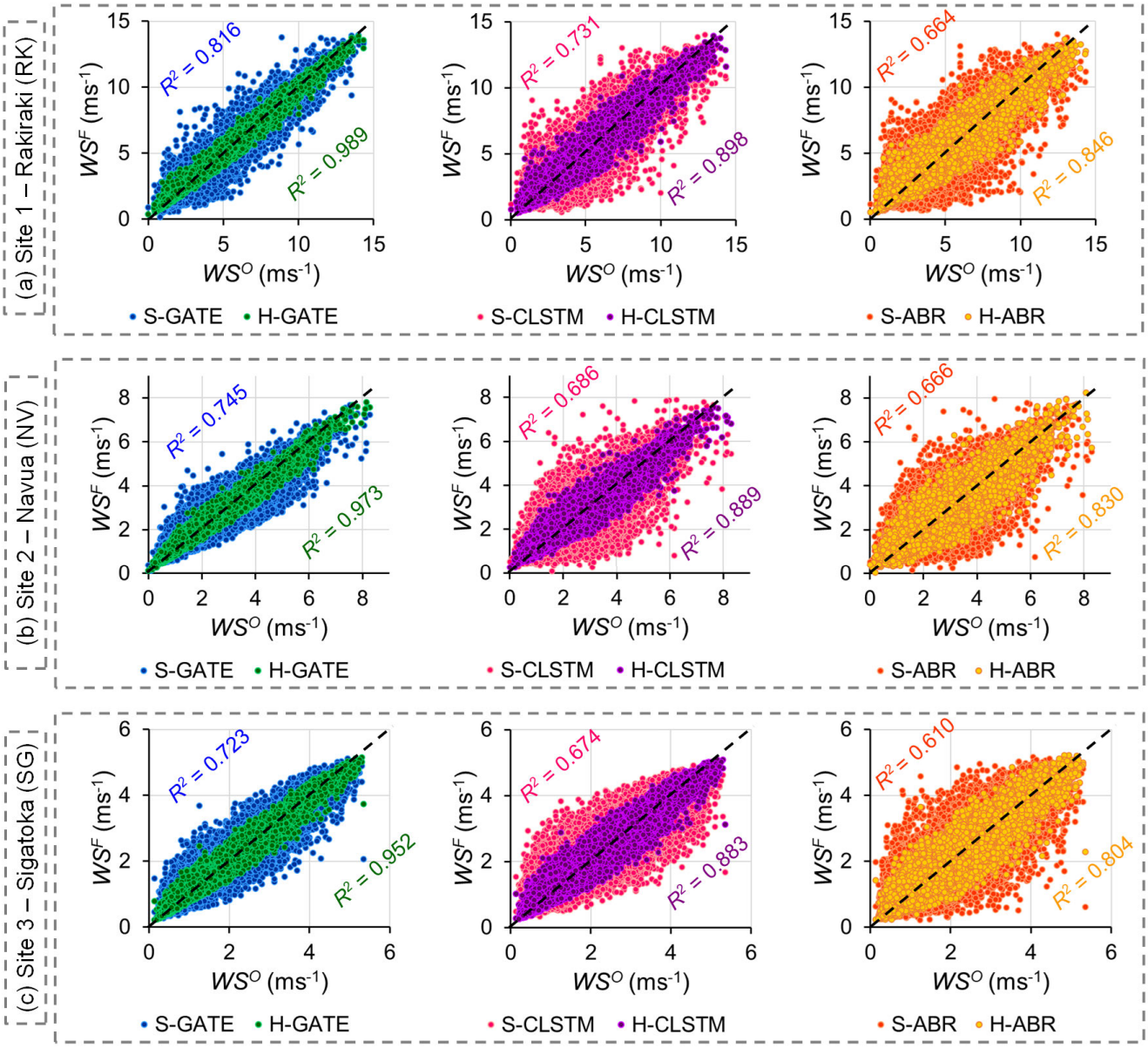


FIGURE 9. Scatter plots for one-step-ahead (t_{L+1}) forecasted wind speed (WS^F) (ms^{-1}) vs. observed wind speed (WS^O) for (a) Site 1 - Rakiraki, (b) Site 2 - Navua, and (c) Site 3 - Sigatoka in the testing phase. For each panel, both standalone (S) and hybrid (H) models are represented, where the highest coefficient of determination (R^2) is considered ideal. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.).

functionality. In 2020, wind energy contributed only 0.12% to Fiji's energy mix, influenced by non-functional turbines and tropical cyclones. The subsequent year saw the lowest generation at 0.03%, attributed to additional turbine damages exacerbated by cyclones and hindered repairs due to COVID-19 restrictions [74]. Although turbines at the Butoni wind farm can be lowered during cyclones, the absence of *WS* forecasting tools delays timely deployment and hampers pre-planning. Hence, implementing accurate seasonal forecasting tools would not just safeguard against turbine damage but also streamline strategic energy supply management, particularly during periods of low *WS* in the dry season, thereby mitigating the risk of power brownouts.

Table 11 provides an overview of the test phase performance evaluation of the proposed H-GATE and comparative models developed for dry and wet seasons. The results are presented only for the one-step-ahead forecast horizon in terms of *MAPE*, *LM*, and *GPI*, offering insights into the predictive performance of models across two different climatic conditions in Fiji. *MAPE*, a relative measure expressing errors in percentage form, facilitates effective model comparison across diverse geographical sites. Notably, *MAPE* underscores H-GATE's robust predictive ability for *WS* forecasting across both seasons, evident by the "darker green" colours. However, all models revealed better *MAPE* for the dry season, where H-GATE registered 29.718% (RK),

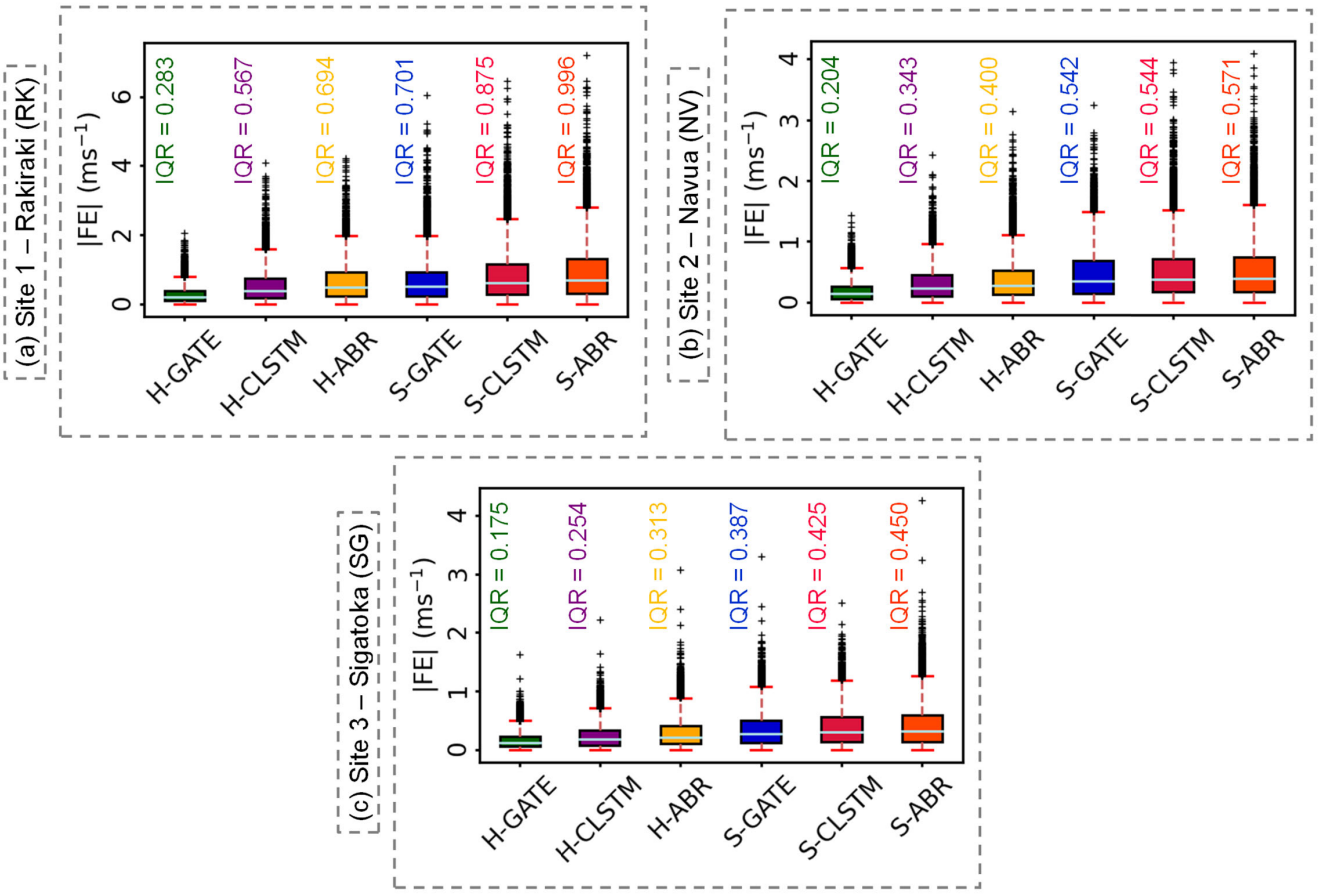


FIGURE 10. Box plots for one-step-ahead (t_{L+1}) forecasted wind speed (*WS*) indicating the absolute value of the forecasting errors $|FE|$ (ms^{-1}) in the testing phase generated by the hybrid gated additive tree ensemble (i.e., H-GATE) against other comparative models implemented at (a) Site 1 - Rakiraki, (b) Site 2 - Navua, and (c) Site 3 - Sigatoka. (Key: IQR is the interquartile range of $|FE|$.).

17.534% (NV), and 16.702% (SG) lower errors compared to the wet season. Site-wise, H-GATE's performance ranking during the dry season is RK (5.184%) > NV (6.754%) > SG (8.359%). Additionally, the *LM* metric, known for its reliable allocation of weights to errors, supports *MAPE* findings, reinforcing H-GATE's predictive superiority for both seasons. *LM* results (Table 11) further highlight the collective poor performance of all models in *WS* forecasting during the wet season.

Moreover, Table 11 summarizes the unified performance index, *GPI*, derived from the results of nine statistical metrics, including *MAPE* and *LM*. *GPI* is a robust indicator that eliminates the need to assess individual metrics. The *GPI* results of models developed for wet and dry seasons are also benchmarked against the original models (Tables 7, 8, and 9) developed for all seasons. Based on this, *GPI* yields greater and lower performance scores for H-GATE and all other models tested during the dry and wet seasons, respectively. It makes sense, given that the wet season has a higher occurrence of cyclones, which induce rapid and unpredictable changes in wind patterns. Hence, the variability in *WS* during cyclonic events makes it challenging to accurately predict the behaviour of wind. Conversely, the dry season has less variability in *WS*, providing better forecast accuracies. Accurate predictions during these distinct seasons are vital for optimizing energy production, enhancing grid stability, and minimizing potential damages to wind infrastructure. Therefore, incorporating seasonal forecasting into H-GATE's capabilities is crucial for ensuring a comprehensive and resilient approach to wind energy management in Fiji.

D. MERITS OF THE PROPOSED H-GATE MODEL

The statistical error metrics and diagnostic plots used for model evaluation supported the H-GATE model for its supreme predictive ability. We wish to clarify that the proposed H-GATE model is superior in the following ways:

1) The GATE architecture outperformed the CLSTM model due to its novel parameter-efficient design, which combines a gating mechanism inspired by GRU and differential nonlinear DTs inspired by tree-based ML models. The GFLUs in GATE are superior for feature learning and selection compared to the feature extraction ability of the 1D-CNN layer in CLSTM. The nonlinear DT-like structure in GATE captures as much of the inductive bias of a DT as possible while

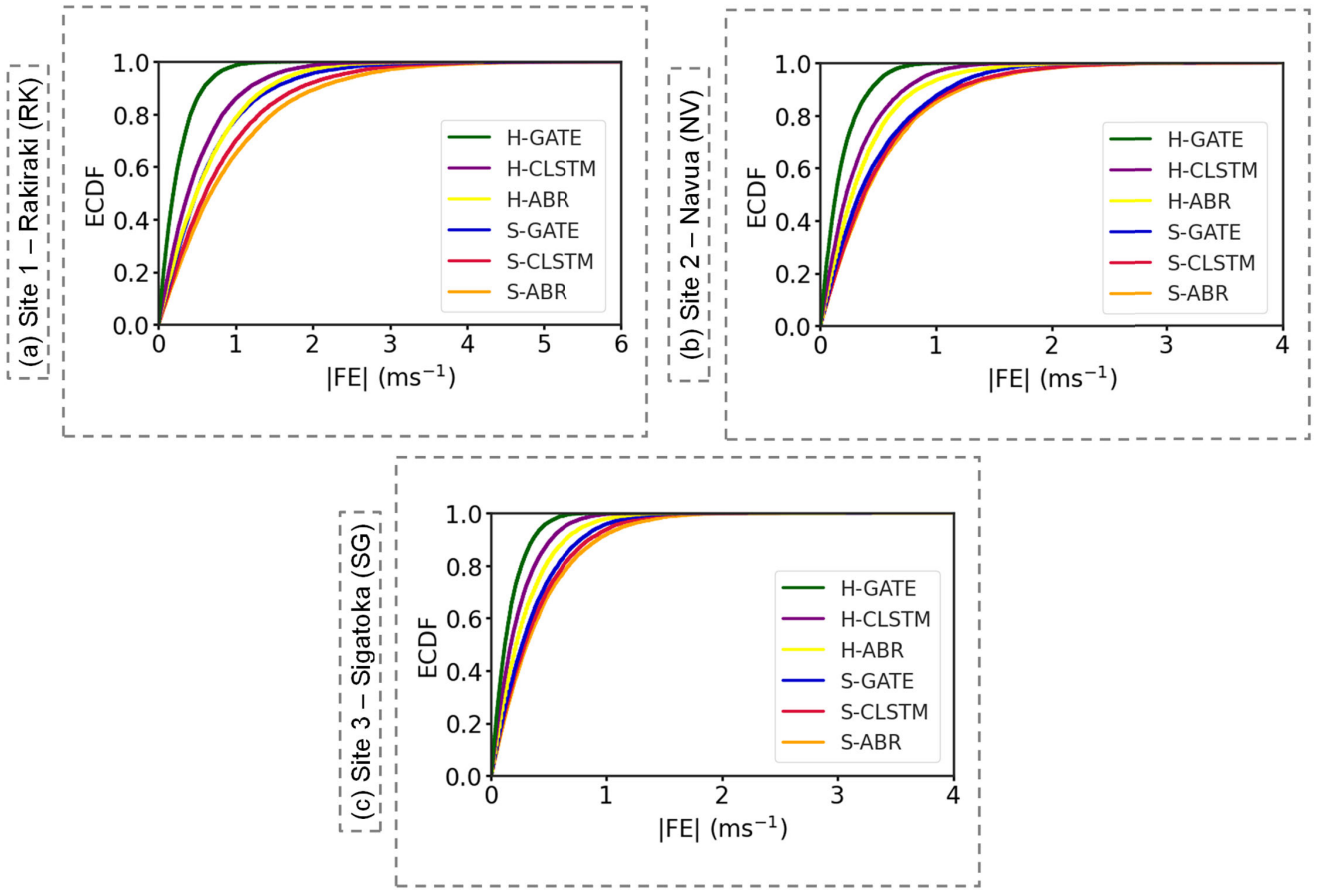


FIGURE 11. Empirical cumulative distribution function (ECDF) plots for one-step-ahead (t_{L+1}) forecasted wind speed (*WS*) indicating the absolute value of the forecasting errors |*FE*| (ms^{-1}) in the testing phase generated by the hybrid gated additive tree ensemble (i.e., H-GATE) against other comparative models implemented at (a) Site 1 - Rakiraki, (b) Site 2 - Navua, and (c) Site 3 - Sigatoka.

TABLE 11. Seasonal evaluation of the one-step-ahead (t_{L+1}) forecasting test performance of the proposed hybrid gated additive tree ensemble (i.e., H-GATE) model against other comparative models in terms of mean absolute percentage error (*MAPE* in %), Legates and McCabe index (*LM*), and global performance indicator (*GPI*) (Note: *GPI* is also tabulated for All seasons for comparison with Dry and Wet season predictive results. Key: Colours "darker green" denotes better model performance, while "darker red" represents poor-performing models.)

		MAPE		*LM*		*GPI*		
Sites	Models	Dry	Wet	Dry	Wet	Dry	Wet	All
RK	H-GATE	5.184	7.376	0.902	0.876	5.061	4.799	4.940
	H-CLSTM	11.131	13.077	0.792	0.767	1.967	1.857	1.929
	H-ABR	13.382	15.799	0.744	0.718	0.438	0.389	0.433
	S-GATE	14.422	17.894	0.729	0.705	-0.420	-0.457	-0.432
	S-CLSTM	17.876	23.651	0.668	0.639	-2.324	-2.673	-2.476
	S-ABR	21.368	24.422	0.625	0.598	-3.939	-4.201	-4.060
NV	H-GATE	6.754	8.187	0.881	0.857	5.624	5.533	5.572
	H-CLSTM	11.005	14.232	0.794	0.768	2.999	2.530	2.817
	H-ABR	15.268	18.002	0.751	0.724	1.149	1.011	1.064
	S-GATE	17.637	19.879	0.687	0.661	-1.014	-1.153	-1.065
	S-CLSTM	20.412	24.755	0.656	0.629	-2.571	-2.809	-2.727
	S-ABR	22.476	25.589	0.643	0.616	-3.376	-3.467	-3.428
SG	H-GATE	8.359	10.035	0.847	0.821	5.206	5.146	5.177
	H-CLSTM	12.233	14.122	0.778	0.751	3.118	2.939	3.049
	H-ABR	17.003	19.348	0.712	0.687	0.982	0.969	0.973
	S-GATE	20.397	23.023	0.652	0.624	-0.822	-1.089	-0.975
	S-CLSTM	20.834	25.899	0.620	0.592	-2.101	-2.349	-2.210
	S-ABR	25.774	28.277	0.587	0.561	-3.794	-3.854	-3.823

remaining parameter-efficient. This makes GATE scalable to large datasets. For better performance, GATE constructs multiple layers of differentiable DTs and ensembles the outputs to obtain accurate *WS* predictions. GATE also outweighed the ML-based ABR model, which could not capture the complex nonlinear

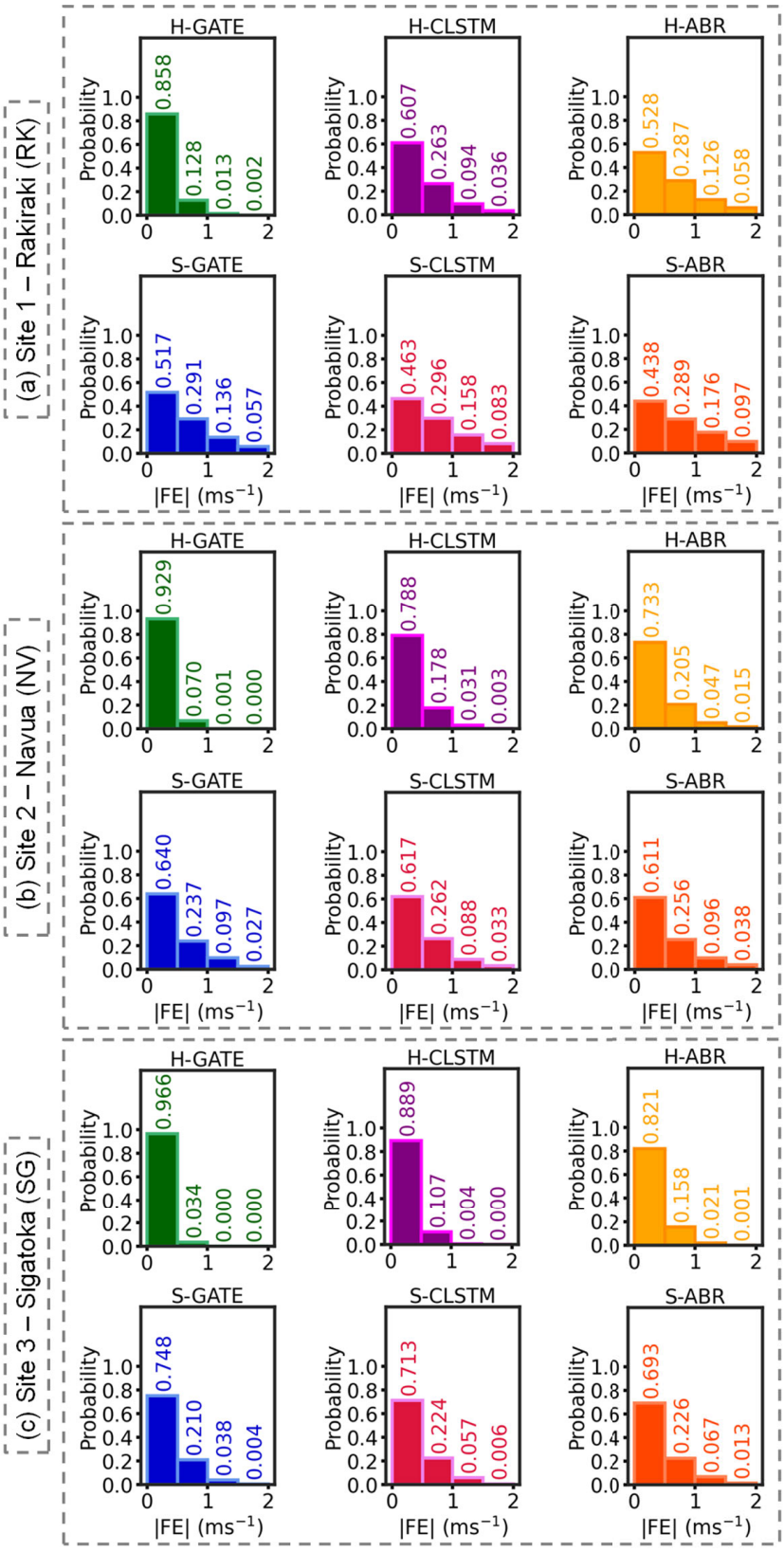


FIGURE 12. Histograms for one-step-ahead (t_{L+1}) forecasted wind speed (*WS*) illustrating the probability of the absolute value of the forecasting errors |*FE*| (ms^{-1}) in the testing phase generated by the hybrid gated additive tree ensemble (i.e., H-GATE) against other comparative models implemented at (a) Site 1 - Rakiraki, (b) Site 2 - Navua, and (c) Site 3 - Sigatoka.

patterns in the data effectively. Another limitation of ABR is its sensitivity to stochastic data. For instance, if the outlier data points get higher weights during model training, then the test performance suffers due to overfitting. This issue is avoidable in GATE due to its gating mechanism.

2) The application of MEMD helped denoise the stochastic *WS* and other meteorological data simultaneously. MEMD effectively captured the interdependencies and interactions between variables during the decomposition process by yielding IMF and residual components with meaningful information. Also, unlike the traditional Fourier and WT-based methods, MEMD did not require any predefined functions or filters, which made it suitable for nonlinear data used in this study. Thus, data decomposition helped further improve the forecasting accuracy of the GATE model.
3) The use of OBWOA helped reduce the risk of overfitting during model training as it enabled the selection of optimal features. The opposition-based learning strategy adopted in OBWOA helped improve the convergence rate during FS. For instance, OBWOA integrated the 'exploration' abilities of the WOA with opposition-based learning's 'exploitation' capabilities. This balance between 'exploration' and 'exploitation' was crucial for locating the optimal feature subset, which helped maximize the generalization performance of the GATE model.

V. CONCLUSION

In recent years, the wind power industry has grown rapidly. Therefore, reliable wind energy management requires accurate *WS* forecasting to encourage further wind energy expansion. Despite this, there is difficulty in obtaining precise short-term *WS* predictions due to the intermittent nature of wind. Accordingly, this paper proposes a novel hybrid GATE (H-GATE) approach to improve the accuracy of *WS* prediction at multi-step-ahead forecast horizons.

In the first step, target and input data are decomposed using MEMD into IMFs and residuals. From the respective IMF and residual pools, the best-lagged IMFs and residuals are selected using the efficient OBWOA. For each pool, optimal lagged IMFs and residuals are fed separately to the BO-optimized GATE model for forecasting each IMF and residual. Using the MIMO strategy, all forecasted IMF and residual components are summed to get the predicted *WS* at one-step-ahead, two-step-ahead, and three-step-ahead forecast horizons. The predictive performance of H-GATE is evaluated against H-CLSTM, H-ABR, S-GATE, S-CLSTM, and S-ABR benchmark models.

Numerous statistical metrics and diagnostic plots endorse the superior forecasting competence of H-GATE over other competing models. The H-GATE model obtained the best results for the one-step-ahead forecast horizon, registering the highest r (0.976 - 0.994), and the lowest *MAE* (0.160 - 0.261) and *RMSE* (0.215 - 0.355) for all three sites.

The error analysis reconfirms the excellent predictive capacity of H-GATE at t_{L+1} by accumulating 85.8%, 92.9%, and 96.6% errors in the lowest error bin ($0 \leq |FE| \leq 0.5$) for RK, NV, and SG, respectively. With this accurate multi-step-ahead prediction method, wind energy managers can produce more energy with less losses, integrate grids, participate in markets, and reduce risks. This tool can help the wind energy industry address the challenges associated with the variable nature of wind resources, ensuring a reliable and sustainable energy supply.

VI. LIMITATIONS AND FUTURE RESEARCH SCOPE

The proposed H-GATE model has acquired accurate multi-step-ahead predictive results at all tested sites. To further improve the forecast performance and the research scope, a few limitations are identified and potential solutions are recommended, which are listed as follows:

1) This study used a MIMO multi-step-ahead forecasting strategy to predict *WS*, which gave highly accurate results for one-step-ahead forecasted *WS*. For two-step-ahead and three-step-ahead predictions, performance dropped successively. In the future, different multi-step-ahead forecasting strategies should be compared, such as recursive, direct, combined direct and recursive (dirREC), and direct and MIMO (dirMO).
2) The wet season *WS* predictions exhibited suboptimal performance compared to the dry season. Future research should prioritize improving wet season *WS* predictions by focusing on advanced modelling strategies, including stacking-based ensemble models, anomaly detection tools, error correction methods, and robust data decomposition techniques. Addressing these aspects with focused efforts can significantly enhance the reliability and precision of *WS* forecasts during the wet season.
3) The forecasted outputs were conveyed as point forecasts. In the future, interval forecasting should be explored to quantify the uncertainty associated with a prediction, and probabilistic forecasting should be explored to fully interpret the distribution of possible outcomes.
4) While the GATE model is interpretable, model interpretability is not discussed in this study as the PyTorch Tabular library used for model execution did not support this feature. To explain the results of GATE, model-agnostic explainable artificial intelligence (*xAI*) methods like LIME and SHAP are recommended for future studies.
5) Data availability is another limitation. Case study sites were selected mainly based on data availability. Most monitored sites in Fiji have over 20% of missing data, which cannot be used for forecasting studies. Advanced data imputation techniques (e.g., transfer learning) should be explored to test if optimal forecasting results are obtained.

TABLE 12. List of acronyms.

Acronym	Full name
ABR	Adaptive Boosting Regressor
AI	Artificial Intelligence
ANN	Artificial Neural Network
APB	Absolute Percentage Bias
ARIMA	Autoregressive Integrated Moving Average
BiLSTM	Bidirectional LSTM
BO	Bayesian Optimization
CEEMDAN	Complete EEMD with Adaptive Noise
CLSTM	CNN-LSTM
CNN	Convolutional Neural Network
DL	Deep Learning
DT	Decision Tree
DTR	Decision Tree Regressor
EEMD	Ensemble EMD
ELM	Extreme Learning Machine
EMD	Empirical Mode Decomposition
E_{NS}	Nash-Sutcliffe Efficiency
FS	Feature Selection
GA	Genetic Algorithm
GATE	Gated Additive Tree Ensemble
GBR	Gradient Boosting Regressor
GFLU	Gated Feature Learning Unit
GPI	Global Performance Indicator
GPR	Gaussian Process Regression
GRU	Gated Recurrent Unit
ICEEMDAN	Improved CEEMDAN
IMF	Intrinsic Mode Function
KF	Kalman Filter
LM	Legates and McCabe Index
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MEMD	Multivariate EMD
MIMO	Multi-Input Multi-Output
ML	Machine Learning
MRA	Multiresolution Analysis
NAR	Nonlinear Autoregressive
NWP	Numerical Weather Prediction
OBWOA	Opposition-Based WOA
r	Pearson's Correlation Coefficient
R^2	Coefficient of Determination
RE	Renewable Energy
RFR	Random Forest Regressor
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RRMSE	Relative Root Mean Square Error
SSA	Singular Spectrum Analysis
SVR	Support Vector Regression
WPD	Wavelet Packet Decomposition
WS	Wind Speed
WS^F	Forecasted Wind Speed
WS^O	Observed Wind Speed
WI	Willmott's Index of Agreement
WOA	Whale Optimization Algorithm
WT	Wavelet Transform
XGBR	Extreme Gradient Boosting Regressor

APPENDIX
ACRONYMS LIST

Table 12 shows the list of acronyms used along the paper.

ACKNOWLEDGMENT

The authors acknowledge Fiji Meteorological Services (FMS) for providing data that enabled this study.

REFERENCES

[1] *Transforming Our World: The 2030 Agenda for Sustainable Development*, document UN Doc. A/RES/70/1, United Nations, New York, NY, USA, Sep. 2015.

[2] S.-X. Lv and L. Wang, "Multivariate wind speed forecasting based on multi-objective feature selection approach and hybrid deep learning model," *Energy*, vol. 263, Jan. 2023, Art. no. 126100.

[3] *Global Wind Report 2023*, Global Wind Energy Council (GWEC), Brussels, Belgium, 2023.

[4] Z. Shang, Y. Chen, Y. Chen, Z. Guo, and Y. Yang, "Decomposition-based wind speed forecasting model using causal convolutional network and attention mechanism," *Expert Syst. Appl.*, vol. 223, Aug. 2023, Art. no. 119878.

[5] S. Farahbod, T. Niknam, M. Mohammadi, J. Aghaei, and S. Shojaeiyan, "Probabilistic and deterministic wind speed prediction: Ensemble statistical deep regression network," *IEEE Access*, vol. 10, pp. 47063–47075, 2022.

[6] Y. Wang, R. Zou, F. Liu, L. Zhang, and Q. Liu, "A review of wind speed and wind power forecasting with deep neural networks," *Appl. Energy*, vol. 304, Dec. 2021, Art. no. 117766.

[7] E. Dokur, N. Erdogan, M. E. Salari, C. Karakuzu, and J. Murphy, "Offshore wind speed short-term forecasting based on a hybrid method: Swarm decomposition and meta-extreme learning machine," *Energy*, vol. 248, Jun. 2022, Art. no. 123595.

[8] M. H. Albadi and E. F. El-Saadany, "Overview of wind power intermittency impacts on power systems," *Electr. Power Syst. Res.*, vol. 80, no. 6, pp. 627–632, Jun. 2010.

[9] C. Lowery and M. O'Malley, "Impact of wind forecast error statistics upon unit commitment," *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 760–768, Oct. 2012.

[10] V. Hoolohan, A. S. Tomlin, and T. Cockerill, "Improved near surface wind speed predictions using Gaussian process regression combined with numerical weather predictions and observed meteorological data," *Renew. Energy*, vol. 126, pp. 1043–1054, Oct. 2018.

[11] H. Cai, X. Jia, J. Feng, Q. Yang, W. Li, F. Li, and J. Lee, "A unified Bayesian filtering framework for multi-horizon wind speed prediction with improved accuracy," *Renew. Energy*, vol. 178, pp. 709–719, Nov. 2021.

[12] M. Khodayar, J. Wang, and M. Manthouri, "Interval deep generative neural network for wind speed forecasting," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3974–3989, Jul. 2019.

[13] S. R. Moreno, R. G. da Silva, V. C. Mariani, and L. dos Santos Coelho, "Multi-step wind speed forecasting based on hybrid multi-stage decomposition model and long short-term memory neural network," *Energy Convers. Manage.*, vol. 213, Jun. 2020, Art. no. 112869.

[14] D. Kaur, T. Tjing Lie, N. K. C. Nair, and B. Vallès, "Wind speed forecasting using hybrid wavelet transform—ARMA techniques," *AIMS Energy*, vol. 3, no. 1, pp. 13–24, 2015.

[15] S. N. Singh and A. Mohapatra, "Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting," *Renew. Energy*, vol. 136, pp. 758–768, Jun. 2019.

[16] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, "A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks," *Future Internet*, vol. 15, no. 8, p. 255, Jul. 2023.

[17] H. Bashir, M. Sibtain, Ö. Hanay, M. I. Azam, Qurat-ul-Ain, and S. Saleem, "Decomposition and Harris hawks optimized multivariate wind speed forecasting utilizing sequence2sequence-based spatiotemporal attention," *Energy*, vol. 278, Sep. 2023, Art. no. 127933.

[18] L. P. Joseph, R. C. Deo, R. Prasad, S. Salcedo-Sanz, N. Raj, and J. Soar, "Near real-time wind speed forecast model with bidirectional LSTM networks," *Renew. Energy*, vol. 204, pp. 39–58, Mar. 2023.

[19] H. Liu, C. Chen, H.-Q. Tian, and Y.-F. Li, "A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks," *Renew. Energy*, vol. 48, pp. 545–556, Dec. 2012.

[20] D. Liu, D. Niu, H. Wang, and L. Fan, "Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm," *Renew. Energy*, vol. 62, pp. 592–597, Feb. 2014.

[21] H. Shao, X. Deng, and F. Cui, "Short-term wind speed forecasting using the wavelet decomposition and AdaBoost technique in wind farm of East China," *IET Gener., Transmiss. Distrib.*, vol. 10, no. 11, pp. 2585–2592, Aug. 2016.

[22] M. Neshat, M. M. Nezhad, E. Abbasnejad, S. Mirjalili, L. B. Tjernberg, D. Astiaso Garcia, B. Alexander, and 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.*, vol. 236, May 2021, Art. no. 114002.

[23] G. Santamaría-Bonfil, A. Reyes-Ballesteros, and C. Gershenson, "Wind speed forecasting for wind farms: A method based on support vector regression," *Renew. Energy*, vol. 85, pp. 790–809, Jan. 2016.

[24] D. E. Rumelhart and J. L. McClelland, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, vol. 1. Hoboken, NJ, USA: Wiley, 1987, pp. 318–362.

[25] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci. USA*, vol. 79, no. 8, pp. 2554–2558, 1982.

[26] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," 2014, *arXiv:1406.1078*.

[27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[28] T. Liang, G. Xie, S. Fan, and Z. Meng, "A combined model based on CEEMDAN, permutation entropy, gated recurrent unit network, and an improved bat algorithm for wind speed forecasting," *IEEE Access*, vol. 8, pp. 165612–165630, 2020.

[29] M. Sibtain, H. Bashir, M. Nawaz, S. Hameed, M. I. Azam, X. Li, T. Abbas, and S. Saleem, "A multivariate ultra-short-term wind speed forecasting model by employing multistage signal decomposition approaches and a deep learning network," *Energy Convers. Manage.*, vol. 263, Jul. 2022, Art. no. 115703.

[30] A. Altan, S. Karasu, and E. Zio, "A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer," *Appl. Soft Comput.*, vol. 100, Mar. 2021, Art. no. 106996.

[31] S. S. Subbiah, S. K. Paramasivan, K. Arockiasamy, S. Senthivel, and M. Thangavel, "Deep learning for wind speed forecasting using bi-LSTM with selected features," *Intell. Autom. Soft Comput.*, vol. 35, no. 3, pp. 3829–3844, 2023.

[32] A. Lawal, S. Rehman, L. M. Alhems, and M. M. Alam, "Wind speed prediction using hybrid 1D CNN and BLSTM network," *IEEE Access*, vol. 9, pp. 156672–156679, 2021.

[33] Q. Wu, F. Guan, C. Lv, and Y. Huang, "Ultra-short-term multi-step wind power forecasting based on CNN-LSTM," *IET Renew. Power Gener.*, vol. 15, no. 5, pp. 1019–1029, Apr. 2021.

[34] S. Zhang, Y. Chen, J. Xiao, W. Zhang, and R. Feng, "Hybrid wind speed forecasting model based on multivariate data secondary decomposition approach and deep learning algorithm with attention mechanism," *Renew. Energy*, vol. 174, pp. 688–704, Aug. 2021.

[35] K. Wang, K. Li, L. Zhou, Y. Hu, Z. Cheng, J. Liu, and C. Chen, "Multiple convolutional neural networks for multivariate time series prediction," *Neurocomputing*, vol. 360, pp. 107–119, Sep. 2019.

[36] M. Joseph and H. Raj, "GANDALF: Gated adaptive network for deep automated learning of features," 2022, *arXiv:2207.08548*.

[37] S. Popov, S. Morozov, and A. Babenko, "Neural oblivious decision ensembles for deep learning on tabular data," 2019, *arXiv:1909.06312*.

[38] J. Močkus, "On Bayesian methods for seeking the extremum," in *Proc. Optim. Techn. IFIP Tech. Conf.*, Novosibirsk, Russia. Springer, Jul. 1974, pp. 400–404.

[39] N. Bokde, A. Feijóo, D. Villanueva, and K. Kulat, "A review on hybrid empirical mode decomposition models for wind speed and wind power prediction," *Energies*, vol. 12, no. 2, p. 254, Jan. 2019.

[40] H. Bo, X. Niu, and J. Wang, "Wind speed forecasting system based on the variational mode decomposition strategy and immune selection multi-objective dragonfly optimization algorithm," *IEEE Access*, vol. 7, pp. 178063–178081, 2019.

[41] S. Pei, H. Qin, Z. Zhang, L. Yao, Y. Wang, C. Wang, Y. Liu, Z. Jiang, J. Zhou, and T. Yi, "Wind speed prediction method based on empirical wavelet transform and new cell update long short-term memory network," *Energy Convers. Manage.*, vol. 196, pp. 779–792, Sep. 2019.

[42] R. Prasad, M. Ali, P. Kwan, and H. Khan, "Designing a multi-stage multivariate empirical mode decomposition coupled with ant colony optimization and random forest model to forecast monthly solar radiation," *Appl. Energy*, vol. 236, pp. 778–792, Feb. 2019.

[43] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London. Ser. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[44] J. Wang, W. Zhang, Y. Li, J. Wang, and Z. Dang, "Forecasting wind speed using empirical mode decomposition and Elman neural network," *Appl. Soft Comput.*, vol. 23, pp. 452–459, Oct. 2014.

[45] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, Jan. 2009.

[46] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 4144–4147.

[47] M. A. Colominas, G. Schlotthauer, and M. E. Torres, "Improved complete ensemble EMD: A suitable tool for biomedical signal processing," *Biomed. Signal Process. Control*, vol. 14, pp. 19–29, Nov. 2014.

[48] D. Korkmaz, H. Acikgoz, and C. Yildiz, "A novel short-term photovoltaic power forecasting approach based on deep convolutional neural network," *Int. J. Green Energy*, vol. 18, no. 5, pp. 525–539, Apr. 2021.

[49] S. Monjoly, M. André, R. Calif, and T. Soubdhan, "Hourly forecasting of global solar radiation based on multiscale decomposition methods: A hybrid approach," *Energy*, vol. 119, pp. 288–298, Jan. 2017.

[50] Y. Li, N. Zhu, and Y. Hou, "A novel hybrid model for building heat load forecasting based on multivariate empirical modal decomposition," *Building Environ.*, vol. 237, Jun. 2023, Art. no. 110317.

[51] N. Rehman and D. P. Mandic, "Multivariate empirical mode decomposition," *Proc. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 466, no. 2117, pp. 1291–1302, 2117.

[52] D. Looney and D. P. Mandic, "Multiscale image fusion using complex extensions of EMD," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1626–1630, Apr. 2009.

[53] R. Prasad, M. Ali, Y. Xiang, and H. Khan, "A double decomposition-based modelling approach to forecast weekly solar radiation," *Renew. Energy*, vol. 152, pp. 9–22, Jun. 2020.

[54] P. Gupta and R. Singh, "Forecasting hourly day-ahead solar photovoltaic power generation by assembling a new adaptive multivariate data analysis with a long short-term memory network," *Sustain. Energy, Grids Netw.*, vol. 35, Sep. 2023, Art. no. 101133.

[55] W. J. M. L. P. Jayasinghe, R. C. Deo, A. Ghahramani, S. Ghimire, and N. Raj, "Deep multi-stage reference evapotranspiration forecasting model: Multivariate empirical mode decomposition integrated with the boruta-random forest algorithm," *IEEE Access*, vol. 9, pp. 166695–166708, 2021.

[56] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016.

[57] X.-S. Yang, *Nature-Inspired Metaheuristic Algorithms*. Luniver Press, 2010.

[58] R. E. J. Kennedy, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw.*, vol. 4, Oct. 1995, pp. 1942–1948.

[59] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014.

[60] N. F. Johari, A. M. Zain, M. H. Noorfa, and A. Udin, "Firefly algorithm for optimization problem," *Appl. Mech. Mater.*, vol. 421, pp. 512–517, Sep. 2013.

[61] A. Askarzadeh, "A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm," *Comput. Struct.*, vol. 169, pp. 1–12, Jun. 2016.

[62] A. Sheng, L. Xie, Y. Zhou, Z. Wang, and Y. Liu, "A hybrid model based on complete ensemble empirical mode decomposition with adaptive noise, GRU network and whale optimization algorithm for wind power prediction," *IEEE Access*, vol. 11, pp. 62840–62854, 2023.

[63] C. Zhang, C. Ji, L. Hua, H. Ma, M. S. Nazir, and T. Peng, "Evolutionary quantile regression gated recurrent unit network based on variational mode decomposition, improved whale optimization algorithm for probabilistic short-term wind speed prediction," *Renew. Energy*, vol. 197, pp. 668–682, Sep. 2022.

[64] Q. Zhang and L. Liu, "Whale optimization algorithm based on Lamarckian learning for global optimization problems," *IEEE Access*, vol. 7, pp. 36642–36666, 2019.

[65] M. A. Elaziz and D. Oliva, "Parameter estimation of solar cells diode models by an improved opposition-based whale optimization algorithm," *Energy Convers. Manage.*, vol. 171, pp. 1843–1859, Sep. 2018.

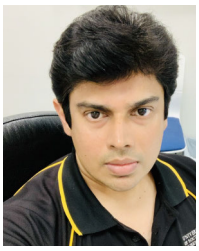
[66] H. S. Alamri, Y. A. Alsariera, and K. Z. Zamli, "Opposition-based whale optimization algorithm," *Adv. Sci. Lett.*, vol. 24, no. 10, pp. 7461–7464, Oct. 2018.
[67] D. P. Solomatine and D. L. Shrestha, "AdaBoost.RT: A boosting algorithm for regression problems," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jul. 2004, pp. 1163–1168.
[68] Y. Huang, N. Hasan, C. Deng, and Y. Bao, "Multivariate empirical mode decomposition based hybrid model for day-ahead peak load forecasting," *Energy*, vol. 239, Jan. 2022, Art. no. 122245.
[69] N. U. Rehman and D. P. Mandic, "Filter bank property of multivariate empirical mode decomposition," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2421–2426, May 2011.
[70] G. Rilling, P. Flandrin, and P. Goncalves, "On empirical mode decomposition and its algorithms," in *Proc. IEEE Workshop Nonlinear Signal Image Process. (EURASIP)*, vol. 3, Jun. 2003, pp. 8–11.
[71] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder–decoder approaches," in *Proc. 8th Workshop Syntax, Semantics Struct. Stat. Transl.*, 2014, p. 103.
[72] Y. Yang, I. Garcia Morillo, and T. M. Hospedales, "Deep neural decision trees," 2018, *arXiv:1806.06988*.
[73] *Fiji's Updated Nationally Determined Contribution*, Ministry Economy Fiji Islands, Suva, Republic of Fiji, 2020.
[74] *2021 Annual Report*, Energy Fiji Ltd., Suva, Republic of Fiji, 2021.
[75] E. Michalena, V. Kouloumpis, and J. M. Hills, "Challenges for Pacific small island developing states in achieving their nationally determined contributions (NDC)," *Energy Policy*, vol. 114, pp. 508–518, Mar. 2018.
[76] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," in *Proc. USENIX OSDI*, vol. 16, Savannah, GA, USA, 2016, pp. 265–283.
[77] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, and D. Cournapeau, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Aug. 2011.
[78] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 2623–2631.
[79] R. C. Deo, X. Wen, and F. Qi, "A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset," *Appl. Energy*, vol. 168, pp. 568–593, Apr. 2016.
[80] R. C. Deo, M. K. Tiwari, J. F. Adamowski, and J. M. Quilty, "Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model," *Stochastic Environ. Res. Risk Assessment*, vol. 31, no. 5, pp. 1211–1240, Jul. 2017.
[81] S. S. Prasad, R. C. Deo, N. Downs, D. Igoe, A. V. Parisi, and J. Soar, "Cloud affected solar UV prediction with three-phase wavelet hybrid convolutional long short-term memory network multi-step forecast system," *IEEE Access*, vol. 10, pp. 24704–24720, 2022.
[82] Y. Ren, P. N. Suganthan, and N. Srikanth, "A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 236–244, Jan. 2015.
[83] M. Premalatha, M. Jayasudha, R. Čep, J. Priyadarshini, K. Kalita, and P. Chatterjee, "A comparative evaluation of nature-inspired algorithms for feature selection problems," *Heliyon*, vol. 10, no. 1, Jan. 2024, Art. no. e23571.
[84] B. Nouri-Moghaddam, M. Ghazanfari, and M. Fathian, "A novel multi-objective forest optimization algorithm for wrapper feature selection," *Expert Syst. Appl.*, vol. 175, Aug. 2021, Art. no. 114737.
[85] M. Abdel-Basset, D. El-Shahat, I. El-henawy, V. H. C. de Albuquerque, and S. Mirjalili, "A new fusion of grey wolf optimizer algorithm with a two-phase mutation for feature selection," *Expert Syst. Appl.*, vol. 139, Jan. 2020, Art. no. 112824.
[86] G. Bontempi, "Long term time series prediction with multi-input multi-output local learning," in *Proc. 2nd ESTSP*, 2008, pp. 145–154.
[87] S. Ghimire, R. C. Deo, D. Casillas-Pérez, and S. Salcedo-Sanz, "Two-step deep learning framework with error compensation technique for short-term, half-hourly electricity price forecasting," *Appl. Energy*, vol. 353, Jan. 2024, Art. no. 122059.
[88] O. Behar, A. Khellaf, and K. Mohammedi, "Comparison of solar radiation models and their validation under Algerian climate – the case of direct irradiance," *Energy Convers. Manage.*, vol. 98, pp. 236–251, Jul. 2015.
[89] H. O. Posten, "Robustness of the two-sample T-test," in *Robustness of Statistical Methods and Nonparametric Statistics*. Dordrecht, The Netherland: Springer, 1984, pp. 92–99.
[90] S. Docimo, K. Seeras, R. Acho, A. Pryor, and K. Spaniolas, "Academic and community hernia center websites in the United States fail to meet healthcare literacy standards of readability," *Hernia*, vol. 26, no. 3, pp. 779–786, Jun. 2022.
[91] J. Parapar, D. E. Losada, and Á. Barreiro, "Testing the tests: Simulation of rankings to compare statistical significance tests in information retrieval evaluation," in *Proc. 36th Annu. ACM Symp. Appl. Comput.*, Mar. 2021, pp. 655–664.
[92] R. Prasad, R. C. Deo, Y. Li, and T. Maraseni, "Ensemble committee-based data intelligent approach for generating soil moisture forecasts with multivariate hydro-meteorological predictors," *Soil Tillage Res.*, vol. 181, pp. 63–81, Sep. 2018.
[93] V. Nourani, Ö. Kisi, and M. Komasi, "Two hybrid artificial intelligence approaches for modeling rainfall–runoff process," *J. Hydrol.*, vol. 402, nos. 1–2, pp. 41–59, May 2011.
[94] D. R. Legates and G. J. McCabe Jr., "Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation," *Water Resour. Res.*, vol. 35, no. 1, pp. 233–241, 1999.
[95] J. Hora and P. Campos, "A review of performance criteria to validate simulation models," *Expert Syst.*, vol. 32, no. 5, pp. 578–595, Oct. 2015.
[96] S. Ghimire, R. C. Deo, N. Raj, and J. Mi, "Deep learning neural networks trained with MODIS satellite-derived predictors for long-term global solar radiation prediction," *Energies*, vol. 12, no. 12, p. 2407, Jun. 2019.
[97] M.-F. Li, X.-P. Tang, W. Wu, and H.-B. Liu, "General models for estimating daily global solar radiation for different solar radiation zones in mainland China," *Energy Convers. Manage.*, vol. 70, pp. 139–148, Jun. 2013.
[98] J. E. Nash and J. V. Sutcliffe, "River flow forecasting through conceptual models Part I—A discussion of principles," *J. Hydrol.*, vol. 10, no. 3, pp. 282–290, Apr. 1970.
[99] C. J. Willmott, "On the evaluation of model performance in physical geography," in *Spatial Statistics and Models*. Dordrecht, The Netherland, 1984, pp. 443–460.
[100] C. J. Willmott, "On the validation of models," *Phys. Geography*, vol. 2, no. 2, pp. 184–194, 1981.
[101] P. Krause, D. P. Boyle, and F. Bäse, "Comparison of different efficiency criteria for hydrological model assessment," *Adv. Geosci.*, vol. 5, pp. 89–97, Dec. 2005.
[102] M. Jafari and N. Ansari-Pour, "Why, when and how to adjust your P values?," *Cell J.*, vol. 20, no. 4, pp. 604–607, 2019.

LIONEL P. JOSEPH received the B.Sc. degree in biology and chemistry and the M.Sc. degree in renewable energy management from The University of Fiji. He is currently pursuing the Ph.D. degree with the Advanced Data Analytics Laboratory, School of Mathematics, Physics and Computing, University of Southern Queensland, Australia. His research interests include data science, with a focus on artificial intelligence (including machine learning and deep learning), renewable energy systems, wind energy, climate mitigation, and environmental sciences.

RAVINESH C. DEO (Senior Member, IEEE) currently leads the Advanced Data Analytics Laboratory, University of Southern Queensland (USQ), Australia, as a Professor. He has published more than 270 articles, 170 journals, and seven books with a cumulative citation that exceeds 14,000 and an H-index of more than 67. He leads cross-disciplinary research in deep learning and artificial intelligence. He also supervising more than 20 Ph.D./M.Sc. degrees and has supervised more than 35 Ph.D./M.Sc. degrees/postdocs. He is a Clarivate Highly Cited Researcher with publications ranking in top 1% by citations for field and publication year in the Web of Science citation index and is among scientists and social scientists who have demonstrated significant broad influence, reflected in the publication of multiple papers frequently cited by peers. He received the Employee Excellence Award, the Elsevier Highly Cited Paper Award, and the Publication Excellence and Teaching Commendations, including U.S. $4.3 M research funding.

DAVID CASILLAS-PÈREZ received the B.S. degree in telecommunication engineering and the M.S. and Ph.D. degrees in electronic control systems from Universidad de Alcalá, Spain, in 2013, 2014, and 2019, respectively. He is currently an Assistant Professor with the Department of Signal Processing and Communications, Universidad Rey Juan Carlos, Madrid, Spain. His research interests include the development of machine learning algorithms with applications in different fields, such as computer vision, mobile communication systems, or renewable energy systems.

RAMENDRA PRASAD received the B.Sc. degree in mathematics and physics and the M.Sc. degree in physics from The University of the South Pacific, Fiji, and the Ph.D. degree in modeling and simulations from the University of Southern Queensland, Australia, in 2019. He is currently a Senior Lecturer with The University of Fiji. He has authored/coauthored 40 papers in peer-reviewed journals and reputed conferences. His research interests include modeling and simulations, advanced machine learning approaches, advanced data analytics, hydrological modeling, energy modeling, energy management, environmental and atmospheric modeling, and ocean wave modeling.



NAWIN RAJ received the B.Sc., B.Ed., PGDMA., and M.Sc. degrees in computational fluid dynamics from the University of the South Pacific, and the Ph.D. degree from the School of Maths, Physics and Computing, University of Southern Queensland (USQ), Queensland, Australia, in 2015. From 2007 to 2010, he was a Lecturer with Fiji National University. From 2015 to 2016, he was a Mathematics Learning Advisor with USQ, where he is currently a Senior Lecturer. He is a member of the Advanced Data Analytics Research Laboratory, USQ. His research interests include artificial intelligence, deep learning, non-linear oscillation, computational fluid dynamics, and oceanography.

SANCHO SALCEDO-SANZ was born in Madrid, Spain, in 1974. He received the B.S. degree in physics from Universidad Complutense de Madrid, Spain, in 1998, the Ph.D. degree in telecommunications engineering from Universidad Carlos III de Madrid, Spain, in 2002, and the Ph.D. degree in physics from Universidad Complutense de Madrid, in 2019. He spent one year with the School of Computer Science, University of Birmingham, U.K., as a Postdoctoral Research Fellow. He is currently a Full Professor with the Department of Signal Processing and Communications, Universidad de Alcalá, Spain. He has coauthored more than 220 international journal articles in the field of machine learning and soft-computing and its applications. His current research interests include soft-computing techniques, hybrid algorithms, and neural networks in different problems of science and technology.