



OPEN ACCESS

EDITED BY

Zijun Zhang,
City University of Hong Kong, Hong
Kong SAR, China

REVIEWED BY

Hong Liu,
City University of Hong Kong, Hong
Kong SAR, China
Minh Quan Duong,
The University of Danang, Vietnam

*CORRESPONDENCE

Imran Aziz,
✉ imran.aziz@physics.uu.se

[†]These authors have contributed equally
to this work

RECEIVED 16 March 2024

ACCEPTED 07 May 2024

PUBLISHED 30 May 2024

CITATION

Ahmed U, Muhammad R, Abbas SS, Aziz I and
Mahmood A (2024), Short-term wind power
forecasting using integrated boosting
approach.

Front. Energy Res. 12:1401978.

doi: 10.3389/fenrg.2024.1401978

COPYRIGHT

© 2024 Ahmed, Muhammad, Abbas, Aziz and
Mahmood. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Short-term wind power forecasting using integrated boosting approach

Ubaid Ahmed¹, Rasheed Muhammad^{1†}, Syed Sami Abbas^{1†},
Imran Aziz^{1,2*} and Anzar Mahmood¹

¹Department of Electrical Engineering, Mirpur University of Science and Technology, Mirpur, Pakistan,

²Department of Physics and Astronomy, Uppsala University, Uppsala, Sweden

Rapidly increasing global energy demand and environmental concerns have shifted the attention of policymakers toward the large-scale integration of renewable energy resources (RERs). Wind energy is a type of RERs with vast energy potential and no environmental pollution is associated with it. The sustainable development goals: affordable and clean energy, climate action, and industry, innovation and infrastructure, can be achieved by integrating wind energy into the existing power systems. However, the integration of wind energy will bring instability challenges due to its intermittent nature. Mitigating these challenges necessitates the implementation of effective wind power forecasting models. Therefore, we have proposed a novel integrated approach, Boost-LR, for hour-ahead wind power forecasting. The Boost-LR is a multilevel technique consisting of non-parametric models, extreme gradient boosting (XgBoost), categorical boosting (CatBoost), and random forest (RF), and parametric approach, linear regression (LR). The first layer of the Boost-LR uses the boosting algorithms that process the data according to their tree development architectures and pass their intermediary forecast to LR which is deployed in layer two and processes the intermediary forecasts of layer one models to provide the final predicted wind power. To demonstrate the generalizability and robustness of the proposed study, the performance of Boost-LR is compared with the individual models of CatBoost, XgBoost, RF, deep learning networks: long short-term memory (LSTM) and gated recurrent unit (GRU), Transformer and Informer models using root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE) and normalized root mean square error (NRMSE). Findings demonstrate the effectiveness of the Boost-LR as its forecasting performance is superior to the compared models. The improvement in MAE of Boost-LR is recorded as to be 31.42%, 32.14%, and 27.55% for the datasets of Bruska, Jelinak, and Inland wind farm, respectively as compared to the MAE of CatBoost which is revealed as the second-best performing model. Moreover, the proposed study also reports a literature comparison that further validates the effectiveness of Boost-LR performance for short-term wind power forecasting.

KEYWORDS

wind power, forecasting, hybrid model, boosting algorithms, deep learning network, LSTM

1 Introduction

The daily surge in the world's population drives up the power demand continuously. At present, electric power requirements are mainly fulfilled by fossil fuels as they cover 60% of global electricity demand (Tiseo, 2023). According to the report of Reuters, the United States (US) generated 59% of its electricity from fossil fuels in 2023 (Gavin, 2023). The findings of the global carbon project (GCP) state that the carbon emissions were 36.8 billion tons in 2022 which have increased by 1.1% in 2023. These rising numbers of carbon emissions will cause the temperature to increase by 1.5°C for the next 7 years (Canadell, 2023). With the growth of the world economy and technological advancements, renewable energy is becoming the center of attention in fulfilling the global energy requirements (Liu et al., 2019). Wind power has grown rapidly in recent years as a significant and potential source of energy to address issues such as shortage of energy and environmental pollution (Liu et al., 2018). The superior economics, minimal emissions of greenhouse gases, and endless wind resources have made wind energy a more attention-drawing resource than other energy transition options (Sari and Yalcin, 2024). However, uncertainty is the key issue associated with wind energy. The erratic nature of atmospheric elements such as wind direction and speed, leads to volatility and unreliable integration of wind energy into the present power system (Pinazo and Martinez, 2022). The uncertainties due to meteorological variables can be minimized by using an accurate wind power forecasting model. Based on a time scale and applications, wind power forecasting can be classified into ultra-short-term, short-term, medium-term, and long-term categories. Ultra-short-term forecasting has applications in transient analysis and covers a period of a few minutes to an hour. Short-term forecasting (STF) has applications in unit commitment and economic load dispatch, with a time horizon ranging from 1 hour to 1 week. Applications of medium-term forecasting include reserve requirement determining and maintenance planning which ranges from 1 week to 1 month. Long-term forecasting applications incorporate sustainability planning and proceed for 1 month to 1 year or thereafter (Ahmed et al., 2023). Various studies present different wind power forecasting techniques. In this proposed study, our main focus is on the STF of wind power.

1.1 Related work

Various statistical, physical, and artificial intelligence-based models have been reported in the literature to accomplish the task of wind power forecasting. In (Che et al., 2023), a probabilistic wind power forecasting model is developed to manage grid-tied wind power. The K-forward neighbors based sparse dynamic weighting (K-FSDW), spatial-temporal multi-scale characteristics, and kernel density estimation (KDE) are used in the model to quantify uncertainty. Datasets from offshore wind farms of Penglai District, Shandong, China are used to confirm the model's viability and efficacy. In Dowell and Pinson (2015), an innovative approach for fitting sparse vector auto-regressive is used to model the location parameters which have a significant numerical advantage. In He et al. (2015), the vector auto-regressive (VAR) model's parameters are improved and over-fitting problems are

reduced by creating a sparsified auto-regressive coefficient matrix. This method uses data on farm layout, wind direction, and speed. Turbine-level correlation shows a notable improvement as compared to auto-regressive (AR) and multi-auto-regressive models (multi-AR). Results indicate that the proposed approach, AR, and multi-AR record the MAPE of 5.88, 6.24, and 6.17, respectively.

The authors in Wahdany et al. (2023) have proposed a neural network architecture to anticipate wind power while adjusting for different energy systems. By implementing the optimization approach to train the network, system expenses are minimized. The error deviation and system cost lapses are mitigated based on an account study. In (Ponkumar et al., 2023), a short-term forecasting model leveraging machine learning techniques, such as extreme gradient boosting (XgBoost), categorical boosting (CatBoost), and light gradient boosting machine (LGBM), is presented. The metrics, R-squared, mean square error (MSE), and mean absolute error (MAE) are used to gauge the model's efficacy. Further, the CatBoost model beats the RF and shows better outcomes with an RMSE of 13.84 for the test set. In Karim et al. (2023), the recurrent neural network (RNN) based forecasting model is presented. The RNN model uses the Dynamic Fitness AI-Biruni Earth Radius (DFBER) algorithm to predict wind power data patterns. When it comes to dataset forecast and modelling, the optimized RNN-DFBER performs better than other models due to its accurate data visualizations and low root mean square error (RMSE). The self-attention-based neural network (SANN) is designed for online learning in Dai et al. (2023). By the self-attention mechanism, the SANN model directly represents time relations with power sequences.

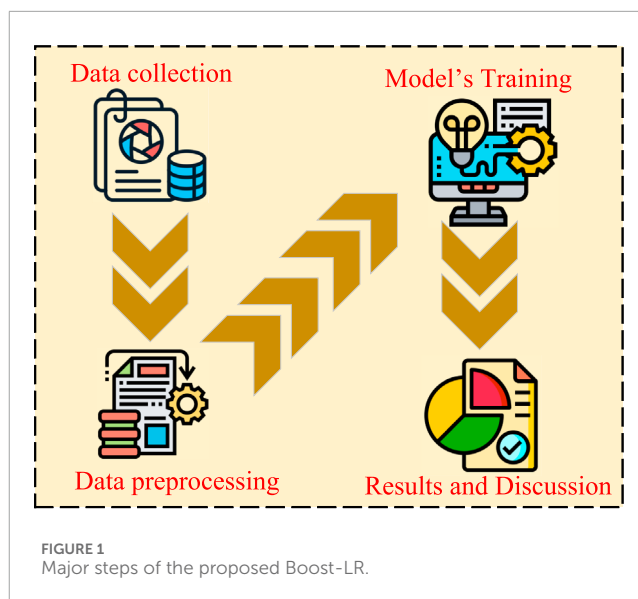
In Khazaei et al. (2022), a high-accuracy hybrid method combining numerical weather prediction data with historical wind farm data has been suggested for short-term wind power forecasting. The technique entails feature selection, prediction, wavelet transform decomposition, and outlier identification. A federated deep reinforcement learning (FedDRL) based model that combines federated learning with deep reinforcement learning (DRL) for ultra-short-term wind power forecasting is presented in Li et al. (2023). The model is decentralized, removing sensitive privacy issues and the deep deterministic policy gradient (DDPG) algorithm is employed to increase predictive accuracy. According to simulation results, FedDRL performs better than traditional methods while maintaining data privacy. The study in Xing and He (2023) explained a multi-modal, multi-step wind power forecasting model that uses stacked deep learning, low-rank matrix fusion, k-dimensional tree, and density-based spatial clustering of applications with noise (DBSCAN). The inland and offshore wind farms' datasets are used to ensure the precision and stability of the model. In Huang et al. (2023), a hybrid methodology is implemented that employs the algorithm of fuzzy C-means clustering to identify weather parameters across different places. A three-tiered hierarchical structure has been designed to use both the gathered data and prior knowledge. It gets trained on the data of wind speed, directs patterns of wind power generation, and predicts power based on historical data and anticipated wind conditions.

In Zhao et al. (2023), a hybrid model named variable mode decomposition convolutional neural network and gated recurrent unit (VMD-CNN-GRU) is developed to predict wind power forecasting in Shanxi, China. The VMD reduces the uncertainty

of wind speed, CNN extracts the spatial parameter from wind power data, and GRU extracts temporal features from the historical data. Findings demonstrate that the proposed VMD-CNN-GRU has better short-term forecasting performance when compared to the other deep-learning models. The model reports the RMSE, MAE, MAPE, and R-squared of 1.5651, 0.8161, 11.62%, and 0.9964, respectively. In [Sheng et al. \(2023\)](#), NWP-based feature classification is achieved with an accurate deep clustering model using a categorical generative adversarial network (CGAN). In addition to providing an enhanced TCN prediction model, a gating mechanism is introduced to increase residual block activation. The empirical mode decomposition and random forest (EMD-RF) model is implemented on wind farms in the US in [Shen et al. \(2018\)](#). In contrast to RF and SVM, it generates better results and minimizes forecasting errors. Findings indicate that the proposed approach, RF, and SVM report the RMSE of 7.86, 11.23, and 10.97, respectively. The authors in [Liu et al. \(2023\)](#) proposed a graph attention convolutional recurrent (GACR) method that coupled GCN and LSTM models. To train the model, high dimensional data consisting of multivariate time series and geoinformation of wind farms is collected. To extract the most relevant features, a double-channel feature extraction algorithm is used in the study. Moreover, the proposed model is also compared with different techniques using RMSE, MAE, accept rate, symmetric MAPE (sMAPE), and NMAE. In [Abou Houran et al. \(2023\)](#), a hybrid model of CNN and LSTM networks is proposed for the short-term forecasting of solar irradiance and wind power. The hyper-parameters of the CNN-LSTM network are optimized using the swarm intelligence (SI) algorithm. The performance of the optimized CNN-LSTM network is compared with individual models of CNN and LSTM and also with their hybrid architectures that are optimized by other meta-heuristic algorithms using different error measurement techniques. To overcome the data privacy issue in forecasting the renewable power output, a bi-party engaged data-driven modelling framework (BEDMF) is presented in [Liu and Zhang \(2022\)](#). In the first stage, spatial and temporal data is collected from multiple sites. The proposed BEDMF technique consists of two stages and involves local and global models. In the pre-training stage, the local model learns the local latent features and they are aggregated to form global latent features. The fine-tuning stage involves learning local and global latent features from local and global data of the previous iteration, respectively. The findings of the study demonstrate that the proposed approach records an average improvement of 3% over other state-of-the-art techniques. Two different ANN networks optimized by the PSO and GA algorithm are presented in [Viet et al. \(2020\)](#) for wind power forecasting. In the study, the dataset is collected from the Tuy Phong wind power plant in Binh Thuan, Vietnam. In the first stage, the ANN network's parameters are optimized by the PSO then another PSO or GA algorithm is implemented to further tune the hyper-parameters resulting in two different architectures: PSO-PSO-ANN and GA-PSO-ANN. The metrics, MSE and MAPE, are used to evaluate the forecasting performance of the models.

1.2 Motivation and contribution

It can be concluded from the above discussion that deep learning networks perform better for short-term wind power forecasting



than other techniques. The hidden layers feature of these networks enables them to learn the data pattern effectively. Various hybrid methodologies integrated into deep learning networks, have also been proposed to enhance forecasting performance. However, for wind power forecasting limited work is performed using boosting algorithms. Therefore, in this study, we proposed a novel integrated boosting algorithm, Boost-LR. The proposed Boost-LR is implemented on four different wind farms' datasets and its performance is evaluated through different metrics. The main contributions of the paper are listed as:

1. A novel integrated Boost-LR network is proposed for hour-ahead wind power forecasting. The proposed Boost-LR is a multi-step network and its first phase consists of three boosting algorithms: RF, CatBoost, and XgBoost. These level 1 boosting algorithms process the input data and give intermediary forecasts, individually. The LR is implemented as a level 2 learner which extracts the data patterns by learning intermediary forecast and input data and provides the final hour-ahead wind power forecasting.
2. To demonstrate the robustness of the study, a comparative analysis is performed for Boost-LR with the individual models of RF, CatBoost, XgBoost, and state-of-the-art deep learning networks, LSTM and GRU, Transformer and Informer models.
3. For generalizability and scalability analysis, the proposed Boost-LR is implemented on different geographical wind farms and its performance is also compared with the findings of studies already reported in the literature.

1.3 Article layout

The rest of the paper is organized as follows, [Section 2](#) describes the methodology, results and discussion are reported in [Section 3](#), and conclusions with future work are summarized in [Section 4](#). This paper is an extended version of our work

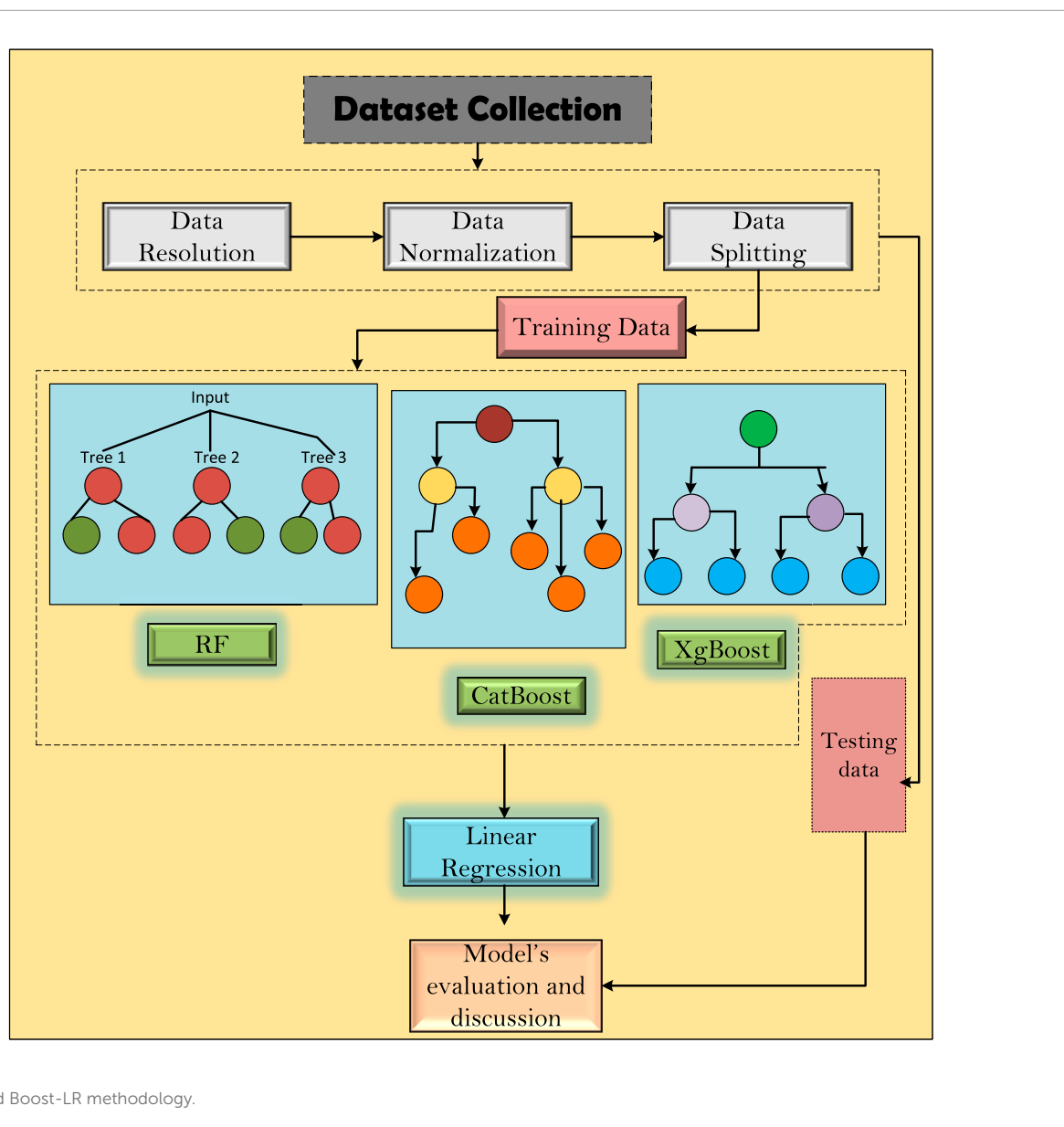


FIGURE 2
The proposed Boost-LR methodology.

accepted as conference proceedings in the 7th International Conference on Energy Conservation and Efficiency (ICECE), 2024 (Ahmed et al., 2024).

2 Methodology

The methodology implemented in this study consists of the following steps and is depicted in Figure 1.

1. The first step involves the collection of datasets. Datasets of different geographical locations are collected to demonstrate the diversity and scalability of the study.
2. In the second step, we preprocess the data which includes data cleaning, normalization, and splitting into training and testing sets.
3. The third step involves models' training and forecasting of hour-ahead wind power.

4. In the last step, the performance of the models is evaluated using different error measurement techniques, and key findings are documented.

These steps are elaborated in the subsequent sections.

2.1 Categorical boosting

Prokhorenkova and Dorogush et al. introduced the sequential gradient boosting technique called CatBoost. It has unique characteristics including categorical features support with little loss and sequentially building decision trees to eliminate gradient bias. During the execution, the model's performance is improved by taking lessons from previous trees and applying them to the next one (Zhang et al., 2023). The distinguished attribute that differentiates CatBoost from other gradient-boosting algorithms is the sequential development of oblivious trees. During the development of the

oblivious tree, one feature is selected that governs all of the splitting criteria (Wang and Qian, 2023). The oblivious tree prevents the overfitting of the curve and improves model execution speed.

2.2 Extreme gradient boosting (XgBoost)

The XgBoost is also a variant of Gradient Boosting Decision Tree (GBDT) technique used for regression problems. It builds novice learners, primarily classification regression trees to train the vulnerable ones. After training, it executes a weighted summation resulting in the final regression model. A new learner is created on the gradient and residual error of the previous learner to reduce the overall network's error (Asselman et al., 2023). Some of the characteristics that distinguish XgBoost from other GBDT algorithms are the incorporation of the L1 and L2 regularization terms, second-order Taylor expansion on loss function, column sampling, and allocating memory to leaf nodes (Li et al., 2022). The XgBoost loss function can be calculated through second-order Taylor expansion as described in Eqs 1–3.

$$L(t) \approx \sum_{i=1}^n \left(L(y_i, y'_{(i-1)} + a_i e_i(x_i) + (1/2) b_i e_i^2(x_i)) \right) \quad (1)$$

$$a_i = e'(t) = \frac{\partial L(y_i, y'_{(i-1)})}{\partial y'_{(i-1)}} \quad (2)$$

$$b_i = e''(x) = \frac{\partial^2 L(y_i, y'_{(i-1)})}{\partial y'_{(i-1)}^2} \quad (3)$$

Whereas, the first and second-order gradient statistics on loss function are denoted by a_i and b_i , respectively. The $y'_{(i-1)}$ corresponds to the prediction at the $(i-1)$ instance.

2.3 Random forest (RF)

The RF is an ensemble learning algorithm used for regression and classification problems. It takes advantage of decision trees in the model's training. It labels test items using decision trees compiled from a random training set. The votes from a lot of trees are summed up to pick the final class (Malakouti, 2023). The bedrock for forecasting is the individual feature and the mean value of the top area spanned by the training set (Gatera et al., 2023).

2.4 Long short-term memory (LSTM)

The LSTM network is an RNN variant that addresses vanishing gradients or outburst issues by switching memory cells for hidden nodes in the RNN architecture. Selective state alteration is made feasible by the network through the configuration of three control gate units: input, output, and forget (Berhich et al., 2023). The dependency information is established by the input gate, unworthy data is discarded by the forget gate, and stored information is sent to the next neuron by the output gate (Xin et al., 2024). The Eqs 4–9 describe the working of LSTM.

$$f(\tau) = \sigma[U_f x(\tau) + V_f h(\tau-1) + a_f] \quad (4)$$

$$i(\tau) = \sigma[U_i x(\tau) + V_i h(\tau-1) + a_i] \quad (5)$$

$$c_o(\tau) = \varphi[U_c x(\tau) + V_c h(\tau-1) + a_c] \quad (6)$$

$$o(\tau) = \sigma[U_o x(\tau) + V_o h(\tau-1) + a_o] \quad (7)$$

$$c(\tau) = f(\tau) \odot c(\tau-1) + i(\tau) \odot c_o(\tau) \quad (8)$$

$$h(\tau) = o(\tau) \odot \varphi[c(\tau)] \quad (9)$$

The symbols ($U_f, U_i, U_o, U_c, V_f, V_i, V_o, V_c$) and (a_f, a_i, a_o, a_c) represents the weights and biases, respectively. The Sigmoid and tanh activation functions are denoted by σ and φ , correspondingly. The \odot symbol is used to denote element-wise multiplication.

2.5 Gated recurrent unit (GRU)

Another RNN variant, the GRU was developed as a memory cell in 2014. By combining the input and forget gates into a single update gate, the number of gates and parameters of the network is minimized to accelerate the convergence. This model can handle abrupt changes in gradient during training and provide better forecasting output by storing the information over a longer period (Lafraxo et al., 2023). It also works very well for memorizing temporal elements that have shifted in external historical data (Thanh et al., 2022). The working of GRU network is explained through Eqs 10–13 (Lafraxo et al., 2023).

$$m_t = (1 - n_t) m_{t-1} + n_t m'_t \quad (10)$$

$$n_t = \sigma(o_z x_t + p_z (m_{t-1})) \quad (11)$$

$$m'_t = \tanh(o_h x_t + p (q_t \odot m_{t-1})) \quad (12)$$

$$q_t = \sigma(o_r x_t + p_q m_{t-1}) \quad (13)$$

The m_t and m'_t denote output and candidate output, respectively. The symbols o_z, o_r, o_h, p_z and p_r represent the metrics in GRU. Whereas, n_t and q_t , correspondingly, represent update and reset gates.

2.6 Transformer

The Transformer model is a transduction model with an attention mechanism and parallel operating capabilities. The encoder-decoder design is the basis of the Transformer. Six encoder blocks make up the encoder portion of the Transformer model, and each decoder's input is determined by the output of the final encoder. There is no way for them to communicate the parameters. However, the encoders are identical in architecture. The encoder block contains self-attention layers and feed-forward neural networks (Zhao et al., 2021). First, the input data is fed to the self-attention layer where a series of matrices, corresponding to

TABLE 1 Datasets description and references.

Dataset location	Duration (year)	External parameters	Ref
Texas, United States of America	1	Wind speed, wind direction, pressure, air temperature	DOBREV (2021)
Bruska	2	Wind speed, wind direction, air density	Zou et al. (2022b)
Jelinak	2	Wind speed, wind direction, air density	Zou et al. (2022b)
Inland wind farm	1	Wind speed	Ding (2019)

TABLE 2 Hyper-parameter tuning with search space and optimized values.

Models	Hyper-parameter	Search space	Optimized values
XgBoost	Tree method	["gpu hist", "approx"]	"approx"
	No. of estimators	[100,200,300]	300
	Maximum depth	[6,9,12]	9
	Learning rate	[0.01,0.03,0.1,0.5]	0.1
	Colsample bytree	[0.6,0.8,0.9,1]	0.9
CatBoost	No. of estimators	[1000,1250,1,500]	1,000
	Maximum depth	[2,4,6,8]	2
	Learning rate	[0.1,0.3,0.01,0.001]	0.3
	L2 leaf regularization	[0.2,0.5,1,3]	0.2
RF	No. of estimators	[20, 60, 100, 120]	120
	Minimum sample split	[2, 5]	2
	Minimum sample leaf	[1,2]	2
	Maximum samples	[0.5, 0.75]	0.5
	Maximum features	[0.2, 0.6, 1.0]	1
	Bootstrap	[True, False]	TRUE
LSTM	No. of layers	[1,2,3]	1
	Optimizers	["Adam", "RMSprop", "SDG"]	Adam
	Learning rate	[0.1, 0.005, 0.001, 0.0001]	0.001
	Drop out	[0.1, 0.2, 0.3, 0.4, 0.5]	0.3
	No. of units	[32, 64, 128, 256]	32

the source data, are generated. These matrices are then multiplied by the weight matrix and the product is the final matrix which is the output of the self-attention layer and can be described by Eq. 14 (Wen et al., 2022).

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (14)$$

Where, Q, K, and V represent query, key, and value vectors, respectively.

The decoder layer has three sub-layers as compared to the encoder layer which has two sub-layers. The additional sub-layer of the decoder is used to map the relation between the current forecasted value and the encoded feature vector. The feed-forward neural network consists of two layers with linear and rectified

linear activation functions which can be described by the Eqs 15, 16 (Zhao et al., 2021).

$$R(\text{Attention}(Q, K, V)) = \max(0, \text{Attention}(Q, K, V) W_1 + m_1) \quad (15)$$

$$\text{FeedNeural} = R(\text{Attention}(Q, K, V) W_2 + m_2) \quad (16)$$

2.7 Informer

The Informer network is a variant of the Transformer model, developed especially for the time series forecasting problems consisting of an encoder and decoder blocks. The distinctive feature of the Informer model is the addition of positional encoding information to vector inputs to address time correlation phenomena. Its multi-head attention mechanism focuses on degradation trends to solve long-term dependence problems. The decoder input consists of hidden intermediate data features and the original vector. To pay attention to the information of position which is to be anticipated in the future and preventing prior positions, the value to be forecasted is set to 0 (Wang et al., 2023).

2.8 Boost-LR

The proposed Boost-LR is a multi-step forecasting approach as elaborated in Algorithm 1. In the first step, we have three different boosting algorithms: XgBoost, CatBoost, and RF. These boosting algorithms process the input data according to their tree development mechanisms. Each of the models in step 1 provides the intermediary forecast. In the next step, LR is used to process the output of level 1 and input training data. The LR adjusts the weights and biases by extracting the information from the intermediary forecast and input training data. To effectively capture the non-linear and complex patterns in the dataset, the Boost-LR leverages different boosting algorithms and uses LR to optimally combine the output of level 1 models and provide a more accurate hour-ahead wind power forecast. Figure 2 illustrates the working of the proposed Boost-LR.

Integrating multiple models into the Boost-LR makes it effective in learning the nuanced and hidden patterns of the datasets. By leveraging the strengths of individual models and overcoming the weaknesses inherited in the single model, the ensemble structure enables the Boost-LR to handle the variability of the datasets accurately. The multilevel approach allows Boost-LR to effectively handle the inherent variability and complexity, present in wind power datasets, ultimately resulting in more accurate hour-ahead forecasts.

2.9 Simulation setup

2.9.1 Dataset description

To demonstrate the diversity and generalizability of the proposed Boost-LR network, wind power datasets of four different geographical locations are used in this study. All these datasets contain external weather parameters, vary in length, and are recorded at the resolution of 1 h. Further description of each dataset

```

Importing functions for data reading, regressors,
neural network, and mathematical operations
Level 1 Models: CatBoost, XgBoost, RF
for each model in [CatBoost, XgBoost, RF]:
model.fit(training_data)
intermediary_forecasts = []
for each model in [CatBoost, XgBoost, RF]:
intermediary_forecasts.append(model.predict(training_data))
Level2Model: LinearRegression
level2_model.fit(intermediary_forecasts, training_data)
new_intermediary_forecasts = []
foreach model in [CatBoost, XgBoost, RF]:
new_intermediary_forecasts.append(model.predict(new_data))
final_prediction = level2_model.predict(new_intermediary_forecasts)

```

Algorithm 1. Proposed Boost-LR

is provided in Table 1. In the data preprocessing steps, we first clean the data by removing the missing values then the MinMax scaler is applied to normalize the data in the range of 0 and 1. The highest value of the data transforms to 1, the lowest to 0 and the remaining values are adjusted between 0 and one accordingly. For splitting the data into training and testing sets, we use the 80, 20 ratio splitting principle where 80% of the data is dedicated to the model's training while 20% is for its evaluation.

2.9.2 Hyper-parameter tuning

The performance of machine learning models is usually dependent on its hyper-parameter settings. Choosing the appropriate values for the hyper-parameters not only improves the forecasting performance of the model but the computational burden is also reduced which enhances the execution speed of the model. Therefore, we tune the hyper-parameters for different models in this proposed study. In Table 2, the hyper-parameters with search space and optimized values are presented. For XgBoost, tree method, number of estimators, maximum depth, learning rate, and colsample bytree are the hyper-parameters that are tuned. For CatBoost, learning rate, number of iterations, L2 lead regularization, and maximum depth are the tuned hyper-parameters. For RF, the number of estimators, maximum depth, minimum sample leaf, minimum sample split, maximum sample, maximum features, and bootstrap are key parameters that are tuned. In the case of deep learning networks, number of layers, units, dropout, batch size, learning rate, and optimizers are tuned parameters. All the hyper-parameters are tuned on the Texas turbine dataset by a randomized search algorithm. Moreover, the optimized values are also used on other datasets.

2.9.3 Performance indicators

Four error metrics: RMSE, MSE, MAE, and NRMSE, are used for the evaluation of models' performances. These metrics are defined by the Eqs 17–20 (Wang et al., 2021):

$$E_{RMSE} = \sqrt{\frac{1}{O} \sum_{I=1}^O (A_I - B_I)^2} \quad (17)$$

$$E_{NRMSE} = \frac{RMSE}{\max(B_I) - \min(B_I)} * 100 \quad (18)$$

TABLE 3 Model's performance evaluation.

Location	Models	RMSE (kW)	MAE (kW)	MSE (kW) ²	NRMSE (%)
	CatBoost	5.23	3.66	27.55	0.17
	XgBoost	3.74	2.52	14	0.12
	RF	87.88	66.1	7,724.39	3.14
Texas	LSTM	21.16	17.85	447.96	0.67
	GRU	44.78	40.33	2005.98	1.43
	Transformer	29.01	18.2	841.58	1.01
	Informer	29.85	18.97	891.02	1.03
	Boost-LR	3.22	2.23	10.38	0.11
	CatBoost	0.052	0.035	0.003	0.144
	XgBoost	0.27	0.095	0.073	0.752
	RF	1.52	1.07	2.3	4.57
Bruska	LSTM	0.47	0.37	0.23	1.28
	GRU	0.32	0.22	0.1	0.84
	Transformer	3.9	2.4	15.21	1.13
	Informer	3.7	2.3	13.69	1.4
	Boost-LR	0.037	0.024	0.001	0.103
	CatBoost	0.045	0.028	0.002	0.152
	XgBoost	0.396	0.181	0.158	1.34
	RF	1.24	0.92	1.53	4.49
Jelinak	LSTM	0.46	0.366	0.214	1.55
	GRU	0.34	0.23	0.19	1.18
	Transformer	3.7	2.6	13.69	1.21
	Informer	3.4	2.1	11.56	1.19
	Boost-LR	0.034	0.019	0.001	0.113
	CatBoost	0.144	0.098	0.021	0.146
	XgBoost	1.45	1.03	2.11	1.48
	RF	0.415	0.258	0.172	0.42
Inland farm	LSTM	0.306	0.265	0.094	0.308
	GRU	0.69	0.498	0.485	0.72
	Transformer	2.2	1.7	4.84	0.86
	Informer	1.5	1.23	2.25	0.35
	Boost-LR	0.119	0.071	0.014	0.021

Bold values shows results of the proposed model.

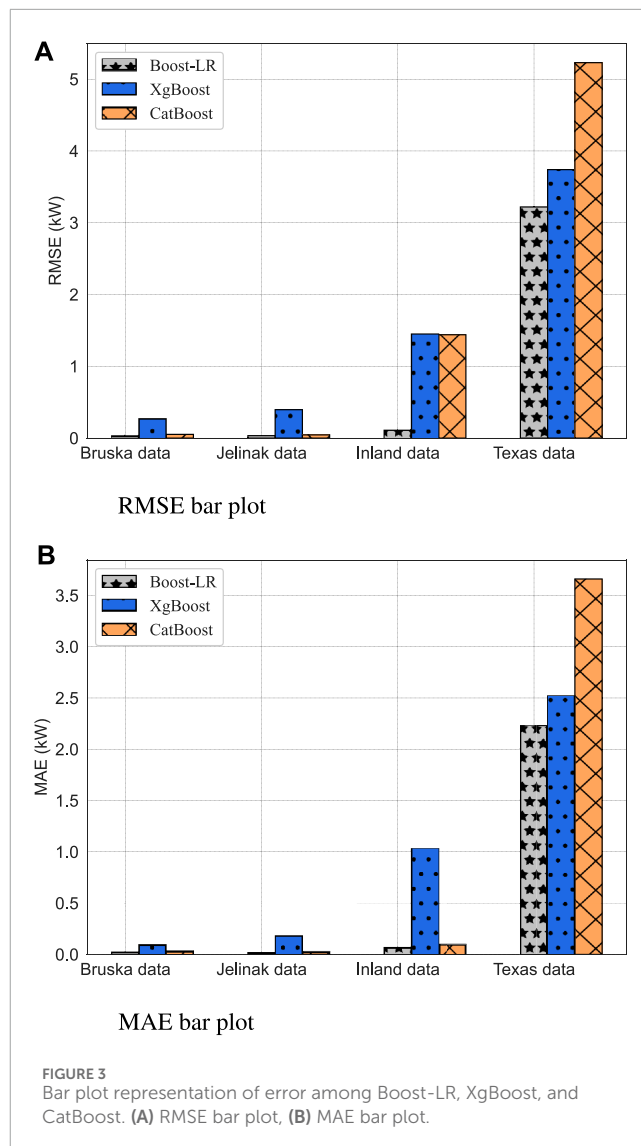


FIGURE 3 Bar plot representation of error among Boost-LR, XgBoost, and CatBoost. (A) RMSE bar plot, (B) MAE bar plot.

$$E_{MAE} = \frac{1}{O} \sum_{I=1}^O |A_I - B_I| \tag{19}$$

$$E_{MSE} = \frac{1}{O} \sum_{I=1}^O (A_I - B_I)^2 \tag{20}$$

Whereas, the measured and predictive values are denoted by A and B, respectively.

3 Results and discussion

In this section, the results of the proposed study are presented. The Boost-LR network is implemented on four different wind power datasets and compared with CatBoost, XgBoost, RF, LSTM, GRU, Transformer, and Informer models. In Table 3, the performance evaluation of each model is presented. The findings of Table 3 report that the proposed Boost-LR outperforms other models on every dataset. The Boost-LR records the RMSE, MAE, MSE, and NRMSE of 3.22, 2.23, 10.38, and 0.11, respectively, for the dataset of Texas Turbine. For Bruska dataset, the RMSE, MAE, MSE, and NRMSE of the

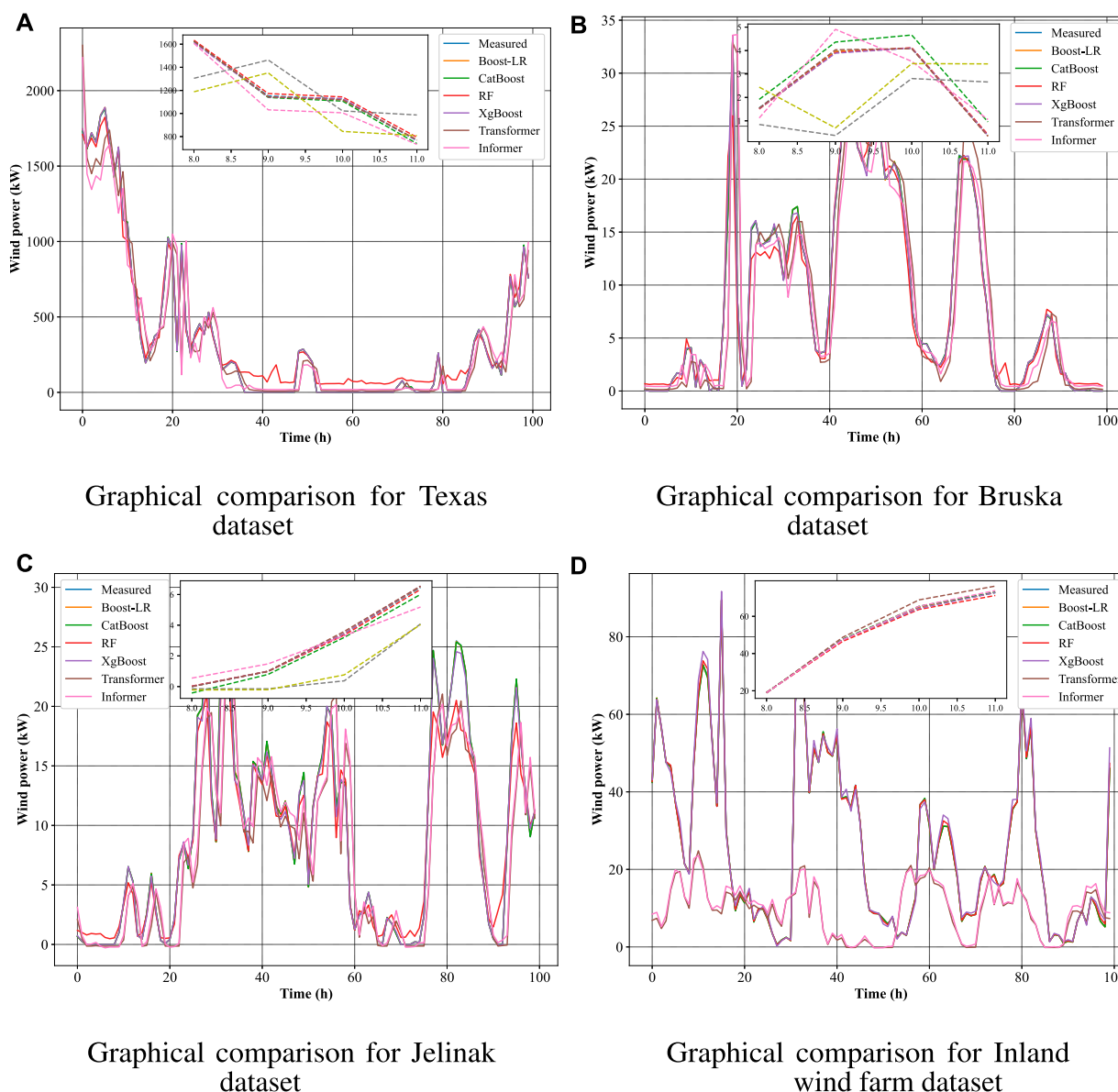


FIGURE 4 Graphical comparison among the measured and Boost-LR predictive curves. (A) Graphical comparison for Texas dataset, (B) Graphical comparison for Bruska dataset, (C) Graphical comparison for Jelinak dataset, (D) Graphical comparison for Inland wind farm dataset.

Boost-LR are found to be 0.037, 0.024, 0.001, and, 0.103, respectively. For the dataset of Jelinak, the Boost-LR records the RMSE, MAE, MSE, and NRMSE of 0.034, 0.019, 0.001, and 0.113, respectively. In the case of inland wind farm, the proposed Boost-LR produces better forecasting results than other techniques and records the RMSE, MAE, MSE, and NRMSE of 0.119, 0.071, 0.014, and 0.021, respectively. The findings also demonstrate that the second-best performing model in comparative analysis is CatBoost for Bruska, Jelinak, and inland wind farm. In the case of the Texas turbine dataset, XgBoost is the second-best-performing model. In Figure 3, the RMSE and MAE of Boost-LR, XgBoost, and CatBoost are illustrated as bar plots.

In Figure 4, the measured curves of wind power are compared with the forecasted curves of the proposed Boost-LR network. This further demonstrates the superiority of the proposed Boost-LR as its prediction curves accurately fit the measured curves.

3.1 Literature comparison

To demonstrate the scalability and generalizability of the proposed Boost-LR, we compare the findings of the proposed study with the work already reported in the literature. Different machine learning and deep learning methodologies are used in (Zou et al., 2022a; Malakouti, 2023) for wind power forecasting on the datasets of the Texas, Bruska, and Jelinak. The comparison of the proposed Boost-LR with the techniques reported in (Zou et al., 2022a; Malakouti, 2023) is presented in Table 4. The findings of Table 4 illustrate the superior performance of Boost-LR as it produces the least forecasting errors. In this study, we propose an integrated network Boost-LR for hour-ahead forecasting of wind power. The Boost-LR's superior performance is due to its two-stage forecasting mechanism.

TABLE 4 Literature comparison of Boost-LR with reported work.

Ref	Journal	Methodology	Dataset	RMSE	MAE
Malakouti (2023)	Case Studies in Chemical and Environmental Engineering	Random forest	Texas	4.64	2.56
//	//	Boost-LR	Texas	3.22	2.23
Zou et al. (2022a)	IEEE Transactions on Sustainable Energy	CNN-BiLSTM and	Bruska	0.133	0.07
//	//	Boost-LR	Bruska	0.037	0.024
Zou et al. (2022a)	IEEE Transactions on Sustainable Energy	CNN-BiLSTM	Jelinak	0.126	0.08
//	//	Boost-LR	Jelinak	0.034	0.019

Bold values shows results of the proposed model.

In the first stage, Boost-LR has multiple tree-based learners: CatBoost, XgBoost, and RF. Each learner provides an intermediary forecast by learning the pattern in the dataset. In the second stage, LR is implemented which measures the linear relationship between the intermediary forecasts and training data. The integration of multiple models makes the Boost-LR effective in learning the nuanced and hidden patterns of the datasets. By leveraging the strengths of individual models and overcoming the weaknesses inherited in the single model, the ensemble structure enables the Boost-LR to handle the variability of the datasets accurately. The superior performance of Boost-LR across five different geographical datasets demonstrates its generalizability and robustness. However, our study has some limitations. First, the performance of Boost-LR is not evaluated for offshore wind turbine datasets. Second, the hyperparameters are tuned on one dataset and the optimized values are used for other datasets as well. Further exploration of hyperparameters can improve the performance of the model. Despite the limitations, the proposed Boost-LR provides a significant advancement in wind power forecasting by providing more accurate results.

4 Conclusion

In this proposed study, we have presented a novel integrated Boost-LR network for hour-ahead wind power forecasting. The proposed approach is an ensemble of CatBoost, XgBoost, RF, and LR techniques. Level 1 of Boost-LR consists of boosting algorithms that process the input data and provide the intermediary forecast. The LR model is fused in level 2 of the Boost-LR which interprets the data and intermediary forecast of level 1 models and provides the final hour-ahead wind power's forecasted value. Wind power datasets of four different geographical locations are tested on Boost-LR and its performance is compared with the individual models of XgBoost, CatBoost, RF, LSTM, GRU Transformer, and Informer networks for the robustness analysis. Findings demonstrate that the proposed approach outperforms other techniques on each of the datasets. The MAE of Boost-LR for the datasets of Bruska, Jelinak, and Inland wind farm is 31.42%, 32.14%, and 27.55%, respectively, better

than the MAE of CatBoost which is the second-best performing model. For the dataset of the Texas turbine, the MAE of Boost-LR is 11.51% better than that of the XgBoost model which is found to be the second-best performing model in comparative analysis. In future work, we aim to integrate Boost-LR with neural network architectures and evaluate the model's performance. We also envision implementing some dimensionality reduction algorithms that highlight the important external features regarding wind power forecasting.

Data availability statement

Information for existing publicly accessible datasets is contained within the article.

Author contributions

UA: Conceptualization, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing–original draft, Writing–review and editing. RM: Investigation, Methodology, Visualization, Writing–original draft. SA: Investigation, Methodology, Visualization, Writing–original draft. IA: Methodology, Resources, Validation, Writing–review and editing. AM: Investigation, Methodology, Resources, Supervision, Validation, Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abou Houran, M., Bukhari, S. M. S., Zafar, M. H., Mansoor, M., and Chen, W. (2023). Coa-cnn-lstm: coati optimization algorithm-based hybrid deep learning model for pv/wind power forecasting in smart grid applications. *Appl. Energy* 349, 121638. doi:10.1016/j.apenergy.2023.121638
- Ahmed, U., Abbas, S. S., Razzaq, S., and Mahmood, A. (2024). "A novel integrated approach for short-term wind power forecasting," in Accepted for publication in 7th International Conference on Energy Conservation and Efficiency (ICECE), Lahore, Pakistan (IEEE).
- Ahmed, U., Khan, A. R., Razzaq, S., and Mahmood, A. (2023). "Comparison of memory-less and memory-based models for short-term solar irradiance forecasting," in 2023 7th International Multi-Topic ICT Conference (IMTIC), Jamshoro and Karachi, May, 2023, 1–6.
- Asselman, A., Khaldi, M., and Aammou, S. (2023). Enhancing the prediction of student performance based on the machine learning xgboost algorithm. *Interact. Learn. Environ.* 31, 3360–3379. doi:10.1080/10494820.2021.1928235
- Berhich, A., Belouadha, F.-Z., and Kabbaj, M. I. (2023). An attention-based lstm network for large earthquake prediction. *Soil Dyn. Earthq. Eng.* 165, 107663. doi:10.1016/j.soildyn.2022.107663
- Canadell, D. P. (2023) *Australia's nation science agency*. Canberra, Australia: CSIRO.
- Che, J., Yuan, F., Deng, D., and Jiang, Z. (2023). Ultra-short-term probabilistic wind power forecasting with spatial-temporal multi-scale features and k-fsdw based weight. *Appl. Energy* 331, 120479. doi:10.1016/j.apenergy.2022.120479
- Dai, X., Liu, G.-P., and Hu, W. (2023). An online-learning-enabled self-attention-based model for ultra-short-term wind power forecasting. *Energy* 272, 127173. doi:10.1016/j.energy.2023.127173
- Ding, Y. (2019) *Data science for wind energy*. Boca Raton, Florida, United States: CRC Press.
- Dobrev, P. (2021). Texas wind turbine dataset - simulated. Available at: <https://www.kaggle.com/datasets/pravdomirdobrev/texas-wind-turbine-dataset-simulated>.
- Dowell, J., and Pinson, P. (2015). Very-short-term probabilistic wind power forecasts by sparse vector autoregression. *IEEE Trans. Smart Grid* 7, 763–770. doi:10.1109/TSG.2015.2424078
- Gatera, A., Kuradusenge, M., Bajpai, G., Mikeka, C., and Shrivastava, S. (2023). Comparison of random forest and support vector machine regression models for forecasting road accidents. *Sci. Afr.* 21, e01739. doi:10.1016/j.sciaf.2023.e01739
- Gavin, M. (2023) *Fossil fuels still dominate global power systems*. London, United Kingdom: Reuters.
- He, M., Vittal, V., and Zhang, J. (2015). "A sparsified vector autoregressive model for short-term wind farm power forecasting," in 2015 IEEE Power and Energy Society General Meeting, Denver, CO, USA, July, 2015 (IEEE), 1–5.
- Huang, Y., Liu, G.-P., and Hu, W. (2023). Priori-guided and data-driven hybrid model for wind power forecasting. *ISA Trans.* 134, 380–395. doi:10.1016/j.isatra.2022.07.028
- Karim, F. K., Khafaga, D. S., Eid, M. M., Towfek, S., and Alkahtani, H. K. (2023). A novel bio-inspired optimization algorithm design for wind power engineering applications time-series forecasting. *Biomimetics* 8, 321. doi:10.3390/biomimetics8030321
- Khazaei, S., Ehsan, M., Soleymani, S., and Mohammadnezhad-Shourkaci, H. (2022). A high-accuracy hybrid method for short-term wind power forecasting. *Energy* 238, 122020. doi:10.1016/j.energy.2021.122020
- Lafraxo, S., El Ansari, M., and Koutti, L. (2023). Computer-aided system for bleeding detection in wce images based on cnn-gru network. *Multimedia Tools Appl.* 83, 21081–21106. doi:10.1007/s11042-023-16305-w
- Li, X., Ma, L., Chen, P., Xu, H., Xing, Q., Yan, J., et al. (2022). Probabilistic solar irradiance forecasting based on xgboost. *Energy Rep.* 8, 1087–1095. doi:10.1016/j.egyr.2022.02.251
- Li, Y., Wang, R., Li, Y., Zhang, M., and Long, C. (2023). Wind power forecasting considering data privacy protection: a federated deep reinforcement learning approach. *Appl. Energy* 329, 120291. doi:10.1016/j.apenergy.2022.120291
- Liu, H., Chen, C., Lv, X., Wu, X., and Liu, M. (2019). Deterministic wind energy forecasting: a review of intelligent predictors and auxiliary methods. *Energy Convers. Manag.* 195, 328–345. doi:10.1016/j.enconman.2019.05.020
- Liu, H., Wu, H., and Li, Y. (2018). Smart wind speed forecasting using ewt decomposition, gwo evolutionary optimization, relm learning and lewt reconstruction. *Energy Convers. Manag.* 161, 266–283. doi:10.1016/j.enconman.2018.02.006
- Liu, H., Yang, L., Zhang, B., and Zhang, Z. (2023). A two-channel deep network based model for improving ultra-short-term prediction of wind power via utilizing multi-source data. *Energy* 283, 128510. doi:10.1016/j.energy.2023.128510
- Liu, H., and Zhang, Z. (2022). A bi-party engaged modeling framework for renewable power predictions with privacy-preserving. *IEEE Trans. Power Syst.* 38, 5794–5805. doi:10.1109/tpwrs.2022.3224006
- Malakouti, S. M. (2023). Estimating the output power and wind speed with ml methods: a case study in Texas. *Case Stud. Chem. Environ. Eng.* 7, 100324. doi:10.1016/j.cscee.2023.100324
- Pinazo, M. A., and Martinez, J. L. R. (2022). Intermittent power control in wind turbines integrated into a hybrid energy storage system based on a new state-of-charge management algorithm. *J. Energy Storage* 54, 105223. doi:10.1016/j.est.2022.105223
- Ponkumar, G., Jayaprakash, S., and Kanagarathinam, K. (2023). Advanced machine learning techniques for accurate very-short-term wind power forecasting in wind energy systems using historical data analysis. *Energies* 16, 5459. doi:10.3390/en16145459
- Sari, F., and Yalcin, M. (2024). Investigation of the importance of criteria in potential wind farm sites via machine learning algorithms. *J. Clean. Prod.* 435, 140575. doi:10.1016/j.jclepro.2024.140575
- Shen, W., Jiang, N., and Li, N. (2018). "An emd-rf based short-term wind power forecasting method," in 2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS), Enshi, China, May, 2018, 283–288.
- Sheng, Y., Wang, H., Yan, J., Liu, Y., and Han, S. (2023). Short-term wind power prediction method based on deep clustering-improved temporal convolutional network. *Energy Rep.* 9, 2118–2129. doi:10.1016/j.egyr.2023.01.015
- Thanh, P. N., Cho, M.-Y., Chang, C.-L., and Chen, M.-J. (2022). Short-term three-phase load prediction with advanced metering infrastructure data in smart solar microgrid based convolution neural network bidirectional gated recurrent unit. *IEEE Access* 10, 68686–68699. doi:10.1109/access.2022.3185747
- Tiseo, I. (2023) *Annual carbon dioxide (co) emissions worldwide from 1940 to 2023*. Hamburg: Statista.
- Viet, D. T., Phuong, V. V., Duong, M. Q., and Tran, Q. T. (2020). Models for short-term wind power forecasting based on improved artificial neural network using particle swarm optimization and genetic algorithms. *Energies* 13, 2873. doi:10.3390/en13112873
- Wahdany, D., Schmitt, C., and Cremer, J. L. (2023). More than accuracy: end-to-end wind power forecasting that optimises the energy system. *Electr. Power Syst. Res.* 221, 109384. doi:10.1016/j.epr.2023.109384
- Wang, D., and Qian, H. (2023). Catboost-based automatic classification study of river network. *ISPRS Int. J. Geo-Information* 12, 416. doi:10.3390/ijgi12100416
- Wang, K., Zhang, J., Li, X., and Zhang, Y. (2023). Long-term power load forecasting using lstm-informer with ensemble learning. *Electronics* 12, 2175. doi:10.3390/electronics12102175
- Wang, Z., Zhang, T., Shao, Y., and Ding, B. (2021). Lstm-convolutional-blstm encoder-decoder network for minimum mean-square error approach to speech enhancement. *Appl. Acoust.* 172, 107647. doi:10.1016/j.apacoust.2020.107647
- Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., et al. (2022). Transformers in time series: a survey. Available at: <https://arxiv.org/abs/2202.07125>.
- Xin, Z., Liu, X., Zhang, H., Wang, Q., An, Z., and Liu, H. (2024). An enhanced feature extraction based long short-term memory neural network for wind power

forecasting via considering the missing data reconstruction. *Energy Rep.* 11, 97–114. doi:10.1016/j.egy.2023.11.040

Xing, Z., and He, Y. (2023). Multi-modal multi-step wind power forecasting based on stacking deep learning model. *Renew. Energy* 215, 118991. doi:10.1016/j.renene.2023.118991

Zhang, S., Lu, X., and Lu, Z. (2023). Improved cnn-based catboost model for license plate remote sensing image classification. *Signal Process.* 213, 109196. doi:10.1016/j.sigpro.2023.109196

Zhao, Z., Xia, C., Chi, L., Chang, X., Li, W., Yang, T., et al. (2021). Short-term load forecasting based on the transformer model. *information* 12, 516. doi:10.3390/info12120516

Zhao, Z., Yun, S., Jia, L., Guo, J., Meng, Y., He, N., et al. (2023). Hybrid vmd-cnn-gru-based model for short-term forecasting of wind power considering spatio-temporal features. *Eng. Appl. Artif. Intell.* 121, 105982. doi:10.1016/j.engappai.2023.105982

Zou, M., Holjevac, N., aković, J., Kuzle, I., Langella, R., Di Giorgio, V., et al. (2022a). Bayesian cnn-bilstm and vine-gmcm based probabilistic forecasting of hour-ahead wind farm power outputs. *IEEE Trans. Sustain. Energy* 13, 1169–1187. doi:10.1109/tste.2022.3148718

Zou, M., Holjevac, N., aković, J., Kuzle, I., Langella, R., Di Giorgio, V., et al. (2022b). Bayesian cnn-bilstm and vine-gmcm based probabilistic forecasting of hour-ahead wind farm power outputs (input dataset). Available at: <https://dx.doi.org/10.21227/y2rj-xb73>