scientific reports

OPEN

Enhanced framework embedded with data transformation and multi-objective feature selection algorithm for forecasting wind power

Yahya Z. Alharthi[1]✉, Haruna Chiroma[2]✉ & Lubna A. Gabralla[3]✉

The increasing global interest in utilizing wind turbines for power generation emphasizes the importance of accurate wind power forecasting in managing wind power. This paper proposed a framework that integrates a data transformation mechanism with a multi-objective none-dominated sorting genetic algorithm III (NSGA-III), coupled with a hybrid deep Recurrent Network (DRN) and Long Short-Term Memory (LSTM) architecture for modeling wind power. The feature selection algorithm, multi-objective NSGA-III, identifies the optimal subset features from wind energy datasets. These selected features undergo a data transformation process before being input into the hybrid DRN-LSTM for wind power forecasting. A comparative study demonstrates the proposal's superior effectiveness and robustness compared to existing frameworks with the proposal achieving 2.6593e−10 and 1.630e−05 in terms of MSE and RMSE respectively whereas the classical algorithm recorded 8.8814e−07 and 9.424e−04. The study's contributions lie in its approach integration of data transformation mechanism and the notable enhancements in wind power forecasting accuracy. Furthermore, the study offers valuable insights to guide research efforts in the future.

Keywords Long short term memory, Renewable energy, Data-transformation, Deep recurrent neural network, Feature selection, Wind turbine power

Using disposable, non-renewable conventional energy can result in ecological harm by producing issues with acid rain and the greenhouse effect. One crucial area for the growth of the new energy sector is wind energy. A wind turbine's aerodynamic performance and flow physical parameters are what determine how efficient it is[1,2]. GWEC's Global wind Report[3] reported that there are 742 GW of installed wind generating systems worldwide at the moment with capacity to reduce 1.1 billion tons carbon dioxide emissions. It is on the earth surface that the wind is generated by uneven heating[4].

Wind energy is recognized as a key renewable energy source that has garnered the attention of government authorities worldwide. It addresses the growing demand for energy and has become a focal point for energy authorities in many countries[5,6]. Regarded as an important energy resource capable of electricity generation[3], the kinetic energy associated with wind increases with its speed. Accurate forecasts of wind energy are essential for the effective utilization of wind power grid[2].

The increasing global interest in harnessing wind turbines for power generation is unprecedented, as noted by Donadio et al.[7]. These turbines, recognized widely as a promising source of renewable energy[8], contribute significantly to electricity production. However, the surge in wind power utilization brings about challenges, primarily stemming from the high-level uncertainties associated with wind power[7]. While generating wind power is a key aspect of utilizing wind energy, its effective control and scheduling pose challenges because of the volatile, uncontrollable, and irregular nature of this energy source[9]. In solving these defies, accurate forecasting

[1]Department of Electrical Engineering, College of Engineering, University of Hafr Albatin, 39524 Hafr Al Batin, Saudi Arabia. [2]College of Computer Science and Engineering, Applied College, University of Hafr Al Batin, Hafr Al Batin, Saudi Arabia. [3]Department of Computer Science, Applied College, Princess Nourah bint Abdulrahman University, P.O. Box 84428, 11671 Riyadh, Saudi Arabia. ✉email: yalharthi@uhb.edu.sa; freedonchi@yahoo.com; charuna@uhb.edu.sa; lagabralla@pnu.edu.sa

of wind power becomes crucial[7]. Thus, ensuring the accurate forecasting of wind power is deemed necessary to manage wind energy to address uncertainties effectively[9,10].

Recognizing the paramount importance of accurate wind energy forecasting, the research community has put forth numerous methodologies centered around intelligent algorithms, particularly artificial neural networks (ANN) and it is variants, to predict wind energy. Different architectures of the ANN, including new generation versions, have been employed to forecast wind energy from diverse perspectives. Shallow algorithms, such as backpropagation ANN, recurrent neural networks (RNN), support vector machine (SVM), genetic algorithm (GA) -optimized ANN, support vector regression (SVR), and particle swarm optimization (PSO)-optimized ANN, have been utilized. Similarly, deep neural networks (DNN) like auto-encoders, long short-term memory (LSTM), convolutional neural networks (CNN), graph neural networks (GNN), deep belief networks (DBN), etc., have been adopted in different ways for wind energy prediction. To enhance data quality, the research community widely incorporates data pre-processing stages. Notably, feature engineering methods play a crucial role, with prominent approaches including feature extraction using principal component analysis and employing deep learning architectures such as auto-encoders (AE) as extractors. Conventional feature selection methods, exemplified by the correlation coefficient and partial mutual information, remained the dominant methods in this domain.

Despite improvement in wind energy prediction using ANN models combined with traditional feature selection algorithms, there exists a noteworthy research gap that requires careful consideration and exploration. The current literature overlooks the incorporation of a data transformation mechanism for data normalization in the frameworks already discussed. Nevertheless, data normalization has demonstrated improved algorithm performance by reducing estimation errors and decreasing algorithm pre-training computational time[11].

It has been established that the effectiveness of modeling time series data with an algorithm is highly dependent on data normalization[12], and wind power data qualifies as time series data because the wind power data are typically gathered at regular period of time intervals. Passalis et al.[13] argue that time series data should be normalized before being fed into DNN for pre-training. A model is likely to perform better if the data is normalized to ensure all features are proportionally within equal range[14]. This consideration is due to the fact that activation functions, especially the sigmoid activation function, can become saturated if input features exceed a certain threshold (exp (-3) 0.05)[15].

Despite wind energy literature predominantly focuses on the deployment of different ANN architectures including DNN for wind energy prediction, often integrating conventional feature selection techniques to enhance model accuracy, a comprehensive exploring of critical aspects of multi-objective nature-inspired meta-heuristic algorithms is conspicuously absent. Published literature indicates that NSGA-III stands out as a robust evolutionary algorithm for feature selection, showcasing superior performance compared to other nature-inspired algorithms feature selection[16–18].

As far as the authors are aware, none of the frameworks proposed in the existing literature for predicting wind energy have integrated both a data transformation mechanism and a multi-objective non-dominated sorting genetic algorithm III (NSGA-III) feature selection algorithm concurrently with a hybrid deep RNN (DRNN) and LSTM.

The paper proposed to develop an enhance pioneering framework that combines a data transformation mechanism and a multi-objective NSGA-III in conjunction with a hybrid deep DRNN-LSTM for predicting wind power.

Research questions: What are the implications and advantages of incorporating a multi-objective NSGA-III in conjunction with the hybrid DRNN-LSTM for wind turbine power prediction? How does the synergistic utilization of a data transformation mechanism and a multi-objective NSGA-III contribute to the overall performance of the proposed framework? What are the comparative strengths of the proposed integrated framework combining hybrid DRNN-LSTM, NSGA-III feature selection and data transformation mechanism in the context of predicting wind turbine power compared to the baseline algorithms?

Additional segments of the paper encompassed the following: Section "Review of related literature on modeling wind energy": Conducts an exhaustive exploration of wind energy forecasting related to the current study. Section "Fundamentals of the framework algorithms": Outlines the fundamentals of the adopted algorithms. Section "Wind energy datasets and pre-processing": Details the collection and pre-processing of the datasets. Section "The proposed integrated framework": Describes the integrated framework proposed in this study. Section "Results and discussion": Presents results along with discussions, leading to conclusive remarks in Section "Conclusions and suggestion for future research".

The study makes several noteworthy contributions highlighted as follows:

- The research introduces a framework designed that enhanced wind power forecasting. This framework combines a data transformation mechanism with a multi-objective NSGA-III feature selection algorithm, integrated with a hybrid DRNN-LSTM.
- The proposed framework integrates diverse techniques, including combining DRNN with LSTM, embedding multi-objective NSGA-III, and incorporating a data transformation mechanism.
- The multi-objective NSGA-III is employed to select optimal wind power features, contributing to improved model performance. This feature selection process is a critical element for enhancing the performance of the framework to forecast wind power.
- Results from the study demonstrate that both the feature selection by multi-objective NSGA-III and the additional data transformation significantly improve the overall performance of the model.
- Comparative analysis reveals that the hybrid DRNN-LSTM outperforms individual algorithms and genetically optimized ANN. This finding underscores the advantage of the proposed framework compared to baseline frameworks.

- The study advocates for the adoption of evolutionary feature selection algorithms and the integration of data transformation in future wind energy forecasting systems by the research community. This recommendation aims to promote advancements in wind power prediction methodologies.
- The study suggests that future research should adopt multimodal transformer for the forecasting of wind energy incorporating multimodalities and external factors. This emphasis on multimodalities may address a practical concern in dealing with varying data and extreme weather conditions in such models.

Review of related literature on modeling wind energy

This section aims to conduct a comprehensive review exclusively focused on relevant works, placing our research problem within the broader context of the existing literature. We examined recent and highly pertinent papers that have presented related works, showcasing the latest advancements in the research field. Our review encompasses both published survey papers and empirical studies published subsequent to the review papers, thus encompassing insights not covered in existing surveys.

A comprehensive examining ANN in wind energy forecasting has been documented. The review highlights that ANN serves as a viable alternative to traditional methods for modeling wind energy, and the rate of publications in this domain is experiencing a notable surge. The studies covered in the review span various countries, incorporating wind data gathered from diverse contexts across nations[19]. Furthermore, an additional review, centering primarily on nonlinear algorithms—particularly ANN—and other forecasting methods for wind power generation, was conducted. This study delved into the diverse applications of ANN for predicting wind power generation, revealing a prevalent reliance on hybrid algorithms over individual algorithms. The trend in the utilization of ANN in wind energy forecasting was observed to be growing rapidly[20]. Similarly, another survey offered a concise review of the utilization of ANN in forecasting renewable energy. Special attention was given to the sources of renewable energy, case studies, major variables employed, and the diverse approaches utilized for forecasting renewable energy[21]. However, there is absence of focus on feature selection, with no identified studies employing feature selection on wind energy data based on the literature surveys. Additionally, the surveyed lacked exploration into the utilization of evolutionary, swarm, and other nature-inspired algorithms in conjunction with ANN to forecast wind power.

An extensive survey delving into the utilization of deep learning algorithms, including LSTM, CNN, Bidirectional LSTM, Gated RNN (GRU), Bidirectional GRU, DBN, and AE, has been presented for wind energy forecasting. The review encompasses the use of deep neural networks as feature extractors, employing algorithms such as AE, CNN, and restricted Boltzmann machine. Various aspects such as data processing, signal processing, outlier detection, and relationship learning are covered within the deep learning context[22]. In a parallel study, Yildiz et al.[23] used a hybrid model for the forecasting of wind power in Turkey. This framework integrates variational mode decomposition and a residual CNN, employing the former for feature extraction before transforming the features into images. These images are then input into a modified residual CNN for the subsequent wind power forecasting. It was found that the proposed framework was superior to SqueezeNet, AlexNet, VGG-16, GoogLeNet and ResNet-18. An alternative research put forth a transfer learning for multi-step wind power forecasting, employing a scheme that utilizes separate batches of data sources across multiple domains. The study incorporated the use of an AE, leveraging it to extract homogeneous characteristics from multiple wind turbines within the dataset. The proposed scheme, integrating AE with transfer learning, demonstrated superior performance compared to conventional methods[24]. A study suggested an intelligent framework that combines CNN, variable mode decomposition, and GRNN for forecasting short-term wind power in Shanxi. The variable mode decomposition is applied to mitigate wind speed volatility. The CNN serves as an extractor, while the GRNN extracts temporal features. Then, the proposed framework is employed for the forecasting of wind power at fifteen-minute intervals, demonstrating superiority over the constituent algorithms[25]. The framework connected various wind turbines from wind farms across different locations to construct a graph data. Utilizing GNN, features were extracted to get the most out of the utilization of spatial and temporal information, resulting in improved forecasting performance[26]. However, the drawback of the feature extraction lies in its inability to provide information about the extent to which the original feature contributes and the absence of interpretability in the linear combination of the raw features[27].

In their study, Puri and Kumar[28] utilized ANN to forecast wind energy in a mountainous region. Data spanning 30 days were collected from the mountaintop and employed to train the ANN for wind energy prediction. Evaluating the effectiveness and efficiency of the ANN proved challenging, as it was not benchmarked against various classical intelligent algorithms. Pasari et al.[29] conducted a study focusing on wind speed prediction using ANN and LSTM. The predictions were made on a monthly basis due to the monthly frequency of the collected data. The study forecast wind farms utilizing ANN, with the computed wind farm values demonstrating superior performance compared to other methods[30]. In a related work, three variants of ANN algorithm, namely, multilayer ANN, a hybrid of SVM with Radial Basis Function (RBF), and PSO optimized ANFIS—were employed for wind speed, wind direction, and turbine wind power forecasting. The results revealed that the hybrid of SVM-RBF outperformed the ANFIS optimized by PSO and multilayer ANN[31]. Another research attempt employed hybrids of numerical weather prediction and ANN for wind power prediction, focusing on a topology with high terrain[7]. Medina et al.[32] employed an ANN for wind power forecasting based on data gathered from various weather stations. Notably, these studies did not incorporate rigorous feature engineering. However, in many instances, not all features are relevant in providing meaningful information. Consequently, certain features in the dataset may be irrelevant and redundant, potentially diminishing the algorithm's performance[33].

In the research, a multi-objective gravity search algorithm was employed for predicting wind speed. The study uses a partial mutual information and a candidate input feature pool for selecting features from the wind speed dataset. This approach led to an enhance forecasting accuracy of wind speed[1]. It's noteworthy that the methods used, are conventional techniques for selecting features. Similarly, in a research endeavor, a deep

GRNN was employed for wind power forecasting. The study utilized a recursive feature elimination approach to select features from the wind dataset. The findings revealed that the GRNN demonstrated enhanced accuracy in wind power forecasting when compared to LSTM[34]. Another investigation utilized a DNN for forecasting wind turbine power. The study involved examining the correlation coefficient of features to select the optimal ones before inputting them into the DRNN for prediction[35]. A proposed multi-task framework combines maximal information coefficient and LSTM for wind power prediction. Correlation coefficient analysis is conducted on wind power data to select relevant features. Grid search is employed to explore the hyper-parameters. Wind speed is considered as the auxiliary prediction, while wind power is the primary task. The proposed multi-task framework outperforms individual algorithms like LSTM, ANN, and ARIMA[36]. However, If the correlation coefficient is too small, it may lead to the exclusion of relevant features, resulting in an inaccurate feature ranking[37]. It is argued that many conventional methods for choosing optimal feature subsets typically encounter significant complexity, premature convergence, and high computational expenses[33]. It is well-established that nature-inspired metaheuristic algorithms outperform conventional methods when it comes to feature selection in prediction tasks[38–40].

Peiris et al.[5] conducted a study utilizing ANN with diverse training algorithms for wind power prediction in Sri Lanka. The findings highlighted the superiority of the Levenberg-Marquardt algorithm in predicting wind power. In the United States, a study utilized ANN with wind data gathered from a 2.5 wind turbine hosted in Iowa to predict wind energy. The study demonstrated the sensitivity of ANN performance to hyper-parameter settings. The results indicated that ANN outperformed classical approaches in predicting wind energy[41]. Similarly, Zafirakis et al.[42] employed ANN and SVR for wind power prediction, reporting an improvement in prediction accuracy. Introducing a different variant of ANN, RNN, Shabbir et al.[43] applied it to predict the energy generated from wind sources. The study found that the predicted energy from wind sources surpassed the accuracy of the classical approach typically used by energy authorities. An examination delved into the effectiveness of RNN variants, including LSTM, temporal pattern attention LSTM, and attention RNN, in predicting wind power in Australia. The study achieved a contemporary record in short-term wind power prediction[44]. In a separate study, Karaman[45] employed LSTM to forecast wind power, comparing its performance with RNN and CNN. The findings indicated that LSTM surpassed the performance of the compared algorithms. These studies including Kisvari et al.[34] applied single algorithm whereas hybrid is proven to be more effective in this domain as evident in[20].

A study introduced a multi-task learning architecture based on a DNN to predict Wind Power. In the context of multi-tasking learning, irrelevant features for a specific task were removed. The DNN multi-tasking learning successfully predicted the wind power ramps events, and the study conducted comparisons with various architectural variations of the proposed framework[46]. In another study, both ANN and multi-linear regression were used for predicting wind speed. The predicted values generated by each algorithm were then compared, revealing that the wind speed values predicted by multi-linear regression were in closer proximity to the actual values compared to the corresponding values predicted by the ANN[4]. However, these studies comparison did not extend to algorithms outside the proposed framework's family, and the studies did not address feature engineering with evolutionary algorithms or any nature-inspired algorithms despite the fact that these feature engineering methods outperform the conventional methods[33].

In contrast to the use of ANNs in studies by Donadio et al.[7] and Peiris et al.[5], a spatial-temporal GNN based on transformers was introduced for wind speed prediction. A graph was used to represent wind data, facilitating processing by the GNN. The spatial-temporal GNN utilized multi-dimensional data for predicting the wind speed[6]. A study employs a GNN to capture uncertainty in wind power. The graph was constructed by representing wind farms and local meteorological factors. The proposed framework integrates graph convolutional ANN, Bootstrap, and bi-directional LSTM to forecast wind power. The results showed that the wind power prediction surpass those of baseline approaches[47]. An ANN based on the transformer architecture incorporating the multi-head attention mechanism is introduced for wind power prediction. The findings indicate the superiority of the proposal over classical methods[9]. These studies neglected the importance of rigorous feature engineering in forecasting tasks, despite its significance[48,49]. Additionally, feature engineering becomes crucial, particularly in resource-constrained situations[23]. However, GNN face challenges, as they demand extensive memory space and consume substantial computational resources[50].

A research endeavor enhanced the sparrow algorithm by incorporating chaotic sequences and Gaussian mutation for LSTM hyper-parameters optimization to forecast wind power. The optimized LSTM was then employed for the prediction of wind power, with the results demonstrating improved accuracy[51]. A study utilized logistic chaos atom search optimization to fine-tune ANN. The study employed the widely-used pre-processing technique, GB/T 18710-002, to enhance data quality. GA and PSO were independently applied to optimize ANN for wind power prediction, facilitating a performance comparison with the proposed logistic chaos atom search optimized ANN. The findings indicates that the proposal surpassed the performance of the compared algorithms. Nevertheless, the study acknowledged challenges, uncertainties, and limitations associated with the commonly used data pre-processing technique, suggesting the exploration of feature selection methods in future study[8]. Xiao et al.[52] suggest utilizing principal component analysis for feature extraction and a modified gated recurrent neural network with hyper-parameters optimized by PSO to forecast wind power. The hybrid algorithm proposed by Xiao et al. demonstrated superior performance compared to conventional approaches. Similarly, Alshammari[53] employed a Mayfly-optimized DRNN and stacked CNN for wind speed prediction. The results indicate an enhancement in prediction performance with the proposed Mayfly-optimized DRNN. Ak et al.[54] employed NSGA-II to train an ANN for the prediction of both wind power and load. However, some researchers argue that employing global optimization algorithms is not imperative for optimizing an ANN to mitigate local minima. Additionally, exerting excessive efforts to seek a global optimum may overly strain the entire network, potentially diminishing flexibility and increasing the likelihood of overfitting the training data in an ANN[55].

A fusion strategy is used to identify imbalances in wind turbine rotors based on a CNN operating in both the time and frequency domains. The imbalance is addressed using the Synthetic Minority Oversampling Technique. The proposed framework effectively detects imbalances[56]. A framework embedded with a stacked CNN, combining bidirectional LSTM and data assimilation, is proposed. The framework was pre-trained with data from multiple sources before being applied to predict wind speed for optimizing wind energy production. It was found to be reliable and outperformed baseline approaches[57]. A hybrid approach integrating multiple machine learning algorithms is used for wind speed forecasting. Comparative analysis indicates that the proposed integrated machine learning algorithm outperforms the baseline algorithms[58]. A piezoelectric wind energy harvester is proposed for forecasting wind energy. The harvester is validated using CNN for wind monitoring and sensing[59]. A multi-graph neural network embedded with a transformer and attention mechanisms is used to predict wind power at both the combined turbine and constituent turbine levels[60]. No transformation. A study combined large eddy simulation and CNN for the prediction of wind power[61]. A CNN embedded with Attention-Bidirectional LSTM and a CNN combined with Transformers and Multi-Layer Perceptrons were deployed for the prediction of solar and wind power. The results showed that the hybrid models outperformed the constituent algorithms[62]. LSTM combined with seasonal mean imputation is deployed to predict wind power. The results shows improvement over conventional algorithms[63]. A framework based on gradient-boosting-LightGBM is propose for assessing the power of wind turbine. The propose framework is found to improve accuracy and enhance power speed and wind load for the wind farm[64]. However, these studies ignored to incorporate data transformation and rigorous feature selection despite the significance of nature-inspired algorithms feature selection[16] and data transformation in data preprocessing[11].

Fundamentals of the framework algorithms

The fundamental concepts of the algorithms constituting various modules in the proposed framework is presented in this section. It is designed to assist readers, particularly newcomers or novice researchers, in comprehending how these algorithms operate to attain their intended objectives. Providing these basic operations renders the paper self-contained, enabling readers to grasp fundamental concepts before delving into the main technical aspects of the study.

Deep recurrent neural network

Before delving into the details of the DRNN, let's explore the fundamental workings of the RNN. An RNN is a type of ANN incorporating feedback connections to handle time series data. At each iteration step, the network receives new inputs and produces outputs, with both the output and new inputs fed back into the network. This inherent feedback connection endows the RNN with the capability to memorize and recall previous data while processing subsequent outputs[65]. In essence, the RNN takes input data, processes it, and generates output. The recurrent weights of the RNN are learned during the training stage, allowing the network to selectively retain or discard information for the progression of subsequent steps[66]. The DRNN extends the capabilities of traditional RNNs by incorporating multiple hidden layers, allowing for the extraction of hierarchical features from sequential input data. Each layer in the multiple hidden layers of the DRNN captures different levels of abstraction and complexity in the sequential data[67].

The DRNN layers are defined by L and the number of neurons in each of the later is defined by N. Assuming the times series data is represented by $\boldsymbol{S}(t)$ with a dimension of N_{in} and $y^{*}(t)$ is the target value in the time series data. Let us take $\bar{x}= [x; 1]$ to avoid writing of bias term explicitly. The hidden state of the DRNN is denoted by layer *i-th* with $a_i(t)$, as such, the equation for the update is expressed as[68]:

$$a_i\,(t) = \tanh(\boldsymbol{W_i}\,a_i\,(t-1) + Z_i\bar{a}_{i-1}(t)\;if\;i>1.$$

$$a_i\,(t) = \tanh(\boldsymbol{W_i}\,a_i\,(t-1) + Z_i\bar{s}(t))\;if\;i=1.$$

In the equations the recurrent connections and the connectivity from the lower layer or the input of the time series data are represented by $\boldsymbol{W}$ and Z respectively. The output $y(t)$ of the DRNN is expressed as:

$$y\,(t) = softmax\left(\sum{}_{i=1}^{L} U_i\;\bar{a}\;(t)\right)$$

Where the U_i is the corresponding weights of the output for the *i-th* network layer[68].

Long short term memory

The LSTM architecture is for overcoming the vanishing problem of gradient associated with the RNNs. It is comprising of three layers including the recurrent hidden layer— LSTM incorporates memory blocks, each featuring memory cells with a shared input gate and output gate. The gates are used for regulating error flow and managing conflicts of the weight within the memory cell, as detailed in[57]. A memory unit consists of a self-connected constant error carousel (CEC), where the activation functions of the CEC mirror the current state of the cell. The incorporation of multiplicative gates, such as input and output gates, improves the learning mechanism by regulating the error flow, effectively overcoming the issue of vanishing gradients as discussed by[69]. In order to mitigate unchecked growing values of the internal cell, in the context of continuous time series data segmentation, a forget gate is incorporated into the memory block. This feature enables automatic resetting of the memory block if the flow of information resulted to being obsolete, with the CEC weight replaced by the activation from the forget gate[70]. Assuming the sequence for the input is given as $x = (x_1,\ x_{2,}\ ...\ ,x_r)$ corresponding to the output given as $y = (y_1,\ y_2,\ ...\ ,y_r)$, The LSTM performs computational iteration given as: $t = 1\ to\ T$[70]:

$$i_t = \sigma\left(W_{ix}x_t + W_{im}m_{t-1} + W_{ic}c_{t-1} + b_i\right)$$

$$f_t = \sigma\left(W_{fx}x_t + W_{fm}m_{t-1} + W_{fc}c_{t-1} + b_f\right)$$

$$c_t = f_t\ \odot\ c_{t-1} + i_t \odot\ g(W_{cx}x_t + W_{cm}m_{t-1} + b_c)$$

$$o_t = \sigma\left(W_{ox}x_t + W_{om}m_{t-1} + W_{oc}c_t + b_o\right)$$

$$m_t = o_t\ \odot\ h\left(c_t\right)$$

$$y_t = \phi\left(W_{ym}m_t + b_y\right)$$

The ⊙ is the scalar product for the two vectors. The weight matrices is Ws whereas b is the bias, σ representing the activation function: sigmoid, φ is the activation function for the output layer while *i* is the input gate and *o* is the output gate. The *f* denotes the forget gate and *c* representing the cell activation vector. Lastly, cell size is represented with *m*[60,71]. The LSTM includes memory cells and gating mechanisms to handle long-term dependencies in sequences[72].

Multi-objective non-dominated sorting genetic algorithm III

The commencement of operations for NSGA-III involves utilizing a set of reference points along with their corresponding descriptions. Offspring for the population are generated through the application of genetic operators on the current parent population at a specific generation. Subsequently, the merged population is sorted based on the levels of non-domination. The members that belong to the population are preserved because they cannot be put in the last front, while excluding the other members. Subsequently select members for next generation while those that remains without being selected remains for the last front. The process of the selection is performed by analyzing the members. Normalization is applied at the beginning of the process in order to make the normalized vector the reference point to ensure a range of objective that is consistent[73].

Wind energy datasets and pre-processing

The wind energy datasets collection process, equipment used, location and period for the data collection are described in this section including the pre-processing stages involved.

Location and data collection

Location - King Abdullah City of atomic and renewable energy (KACARE)

The KACARE was founded in 2010 with the intention of developing Kingdome of Saudi Arabia (KSA) renewable energy sector. The renewable resources atlas of the KSA that KACARE is building, is one of the most significant projects in the energy sector. Nine sites are providing data on wind resources while over 32 monitoring stations across the nation are for data on solar resources. Researchers and power project developers can use the data to realize the 2032 and 2040 renewable energy objectives of the KSA[74].

Data collection

The wind data were gathered by the wind resource monitoring site at various heights above the ground in the designated locations. Multiple features, including wind speed, relative humidity, and wind direction, were collected at each site. The assessment of wind resources in the chosen area was conducted at a height of 100 m, which is the closest height to the specified wind turbines. The data were collected on an hourly basis over 12 months in Northern Saudi Arabia. The wind farm consists of five turbines, each with a production capacity of 1.5 MW, totaling 7.5 MW. Additionally, the setup includes 200 batteries with a capacity of 1 kW each, grid power production measured in kW, and a converter with a size of 250 kW.

Data pre-processing

Evaluating the wind speed potential is a crucial consideration before establishing a wind farm power plant. To evaluate the wind speed potential and anticipated power generation, a system comprising Wind Turbines and Electric Load is essential. This study utilized primary data processed through the HOMER software, developed by the National Renewable Energy Laboratory. In selected locations, a grid-connected wind energy system with 5 wind turbines (each with a capacity of 1.5 MW) was linked to an electric load for the data collection. The HOMER software imported time-series data, including load, weather data, temperature, and other relevant features, collected from a file. The data spanned one year, with a scaled annual average wind speed of approximately 7.07 m/s. The data underwent processing in the HOMER software, and the computed features are detailed in Table 1. The wind turbines' expected power output over the project's lifetime relies on factors such as wind speed potential and other features presented in Table 1, which are further analyzed in this study. The target variable is the Wind Turbines MW Power Output, while the input features comprise the 25 features outlined in Table 1. HOMER determines the wind turbine power output by assessing the wind speed at the turbine hub height in three steps. It calculates the power generated by the turbine at that wind speed and standard air density, subsequently adjusting the power output to reflect the actual air density. The description of the features in Table 1 were extracted from the HOMER's website[75].

	Feature	Description
1	Wind speed	The power generated by the wind turbine at a specific wind speed, considering standard air density, was calculated. Subsequently, HOMER adjusted the power output value to account for the real air density
2	Wind turbines MW operating status	The study employed five wind turbines, with each turbine having a production capacity of 1.5 MW. Additionally, the analysis considered a 200 kWh battery bank and a 250 kW converter system.
3	AC primary load	An integral aspect of the simulation involves load estimation. Two load categories were considered to complete the system analysis. The primary and initial hypothetical AC load is a daily community load of approximately 150 MWh. This load is expected to be integrated into the main grid, drawing power from both the grid and a renewable energy source. The second load is a DC load connected to a monitoring system linked to the DC bus, equipped with a battery bank as a backup system.
4	AC primary load served	
5	DC primary load	
6	DC primary load served	
7	Grid power price	In 2017, the Saudi Electricity Company increased the rates for both commercial and residential electricity. In the kingdom, there are two consumption categories specifically tied to residential loads. The first group applies when consumption exceeds 6000 kWh, with a fee rate of 0.080 USD/kWh. The second category applies a rate fee of 0.048 USD/kWh for usage between 1 and 6000 kWh. These updated rates were implemented to structure fixed prices at different intervals throughout the day.
8	Grid sellback rate	
9	Total electrical load served	The aggregate energy consumed to fulfill both the immediate and deferred loads over the entire year, in addition to the energy sold to the grid, constitutes the total electrical load supplied.
10	Renewable penetration	If the system incorporates renewable components, the Simulation Results window provides access to the Renewable Penetration tab. This tab offers various metrics to assess the extent to which renewable energy sources penetrate the system. The significance of these metrics depends on the specific characteristics and usage of your system. The calculated results for these metrics may vary significantly based on factors such as surplus electricity, unmet load, and other elements.
11	Unmet electrical load	Unmet load is electrical load that the power system is unable to serve
12	Inverter power input	An inverter converts DC electricity to AC electricity. The Inverter Input section contains the following inputs: **Lifetime** - The expected lifetime of the inverter, in years. Efficiency- The efficiency with which the inverter converts DC electricity to AC electricity, in %.
13	Rectifier power input	A rectifier converts AC electricity to DC electricity. The Rectifier Input box contains the following inputs. Relative Capacity: The rated capacity of the rectifier relative to that of the inverter, in %. Efficiency: The efficiency with which the rectifier converts AC electricity to DC electricity, in %.
14	Generic 1 kWh lead acid maximum charge power	The maximum charging power fluctuates from one time step to another based on its state of charge and recent charging and discharging history.
15	Generic 1 kWh lead acid maximum discharge power	The maximum discharge power undergoes fluctuations between consecutive time steps, influenced by both its state of charge and recent charging and discharging history, as dictated by the Kinetic Storage model.
16	Generic 1 kWh lead acid charge power	In every time step, HOMER calculated the maximum power absorption capacity of the storage bank.
17	Generic 1 kWh lead acid discharge power	In every time step, HOMER calculates the maximum power discharge capacity of the storage bank.
18	Generic 1 kWh lead acid input power	The input power refers to the electrical energy supplied by the converter for charging the battery.
19	Generic 1 kWh lead acid energy content	The total energy currently stored in the battery, in kWh
20	Generic 1 kWh lead acid state of charge	The minimum state of charge for the battery is the relative state of charge below which the storage bank is never utilized, expressed as a percentage of the total capacity.
21	Generic 1 kWh lead acid energy cost	At any given time step, the storage energy cost represents the average cost of the energy that the system has supplied to the storage bank up to that specific time step.
22	AC required operating capacity	The necessary operating capacity is calculated in each time step by combining the required operating reserve with the electric load.
23	DC required operating capacity	
24	AC operating capacity	Operating capacity refers to the total electrical generation capacity that is currently active and ready to produce power at any given time. It signifies the maximum electrical demand that the system can immediately accommodate.
25	DC operating capacity	

The proposed integrated framework

The illustrated framework in Fig. 1 provides a clear representation of the collaborative operations conducted by the modules, working cooperatively to predict wind power. The major components include dataset, feature selection, data transformation, hybrid DRNN-LSTM training and benchmarking.

Feature selection

The multi-objective NSGA-III, employing a wrapper-based approach, serves as the feature selection algorithm for selecting wind power features from the wind datasets outlined in section "Wind energy datasets and pre-processing". Before executing the wrapper-based NSGA-III, it is essential to set NSGA-III with hyper-parameters. In this study, the following hyper-parameter settings were adopted: a crossover probability of 0.9, a mutation probability of 0.2, a population size of 200, and 200 generations. These specific hyper-parameter values were adopted based on their promising performance exhibited in a study conducted by Almutairi et al.[38].

The multi-objective NSGA-III with wrapper method is employed to select the optimal subset of features from the wind power dataset. The evaluation process involves assessing the group of wind power features presented in Table 1 using wrapper evaluation measures. The NSGA-III, utilizing the wrapper approach, identifies and selects significant features essential for contributing significantly to wind power forecasting while eliminating irrelevant and redundant features. The wrapper method is then applied to rank the features, taking into account their contribution to wind power prediction performance. Subsequently, learning algorithms are adopted for assessing values of discriminant. The wrapper-based NSGA-III selects diverse groups of features with varying accuracy levels.

Data transformation

The group of the subset features demonstrating the highest accuracy is then forwarded to the transformation module. In this module, the features undergo normalization, a process carried out by: Identifying the actual value (Vr) to be normalized. Identifying the minimum value (Vmin) in each column. Identifying the maximum value (Vmax) in each column. Computing the difference (d) between Vmax and Vmin. At the final stage, data is normalized (V_N) for each column using the expression: VN = (Vr - Vmin)/d.

Hybrid DRNN-LSTM

The training of the DRNN typically encounter gradient vanishing and exploding problems (Li et al., 2018). However, LSTM excels at addressing long-term dependencies and overcoming the issue of vanishing gradients[76]. Thus, hybridizing the LSTMs with the DRNN can contributes to more stable training, making it feasible for the study to train the hybrid DRNN-LSTM without encountering the same level of difficulty like the conventional DRNN. The layers of the LSTM were incorporated into the DRNN architecture in which the architecture was experiments with varying number of layers to design the optimal DRNN-LSTM architecture. The long term dependencies in the sequence was handle by the memory and gates.

DRNN-LSTM training

The performance of DRNN-LSTM is contingent upon the settings of hyper-parameters, and currently, there exists no automatic or systematic approach to obtain the optimal configurations. Therefore, conducted preliminary experiment using a sample from the wind power dataset to train the DRNN-LSTM with varying hyper-parameter values for wind power prediction. The study encompassed a series of experiments exploring different combinations of hyper-parameters, including activation functions, number of neurons across layers, batch size, epochs, the number of steps, and dropout regularization. Results from the preliminary experiment revealed the optimal settings for these hyper-parameters Table 2. Subsequently, the DRNN-LSTM, configured

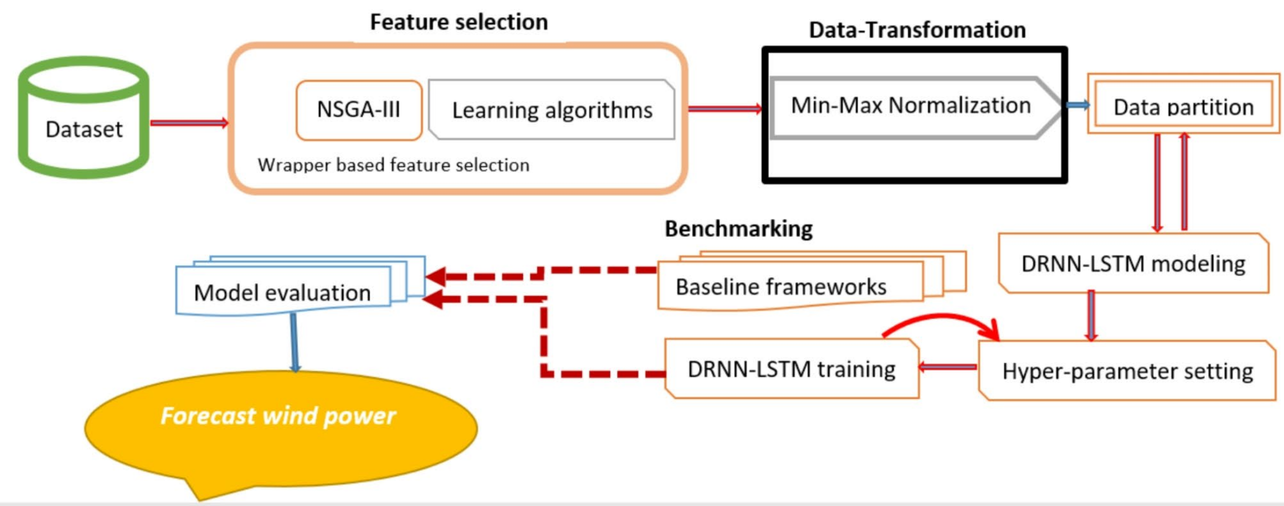

Fig. 1. The proposed enhanced framework for the forecasting of wind power.

Algorithms	Hyper-parameters	Settings
LSTM, DRNN, and DRNN-LSTM	Activation function	ReLu and Sigmoid
	Hidden layers	3
	Neurons in each hidden layer	100
	Batch size	100
	Steps	100
	Dropout regularization	0
	Epoch	100
GA	Mutation rate	0.02
	Population	100,000
	Generation	10,000
	Crossover rate	0.5
ANN	Activation function	ReLu and Sigmoid
	Hidden layers	1
	Neurons in each hidden layer	100
	Batch size	100
	Step	100
	Dropout regularization	0
	Epoch	100

with these optimal hyper-parameters identified during the preliminary stage, was deployed for a full-scale experiment using the complete dataset. The transformed data were then input into the DRNN-LSTM for training, employing multiple data partition ratios to predict wind power.

Benchmarking
To conduct a benchmark of the proposed framework against baseline frameworks, a selection of algorithms was made for the evaluation of DRNN-LSTM wind power forecasting. The choice of algorithms aimed to cover a range of approaches for comprehensive evaluation. An ANN was included due to its well-established reputation in wind energy prediction, as evident from the literature surveyed. Both DRNN and LSTM were selected as they constitute integral components of the hybridized algorithm, complementing each other to enhance the overall architecture. As previous studies suggest that hybrid algorithms outperform their single counterparts. Additionally, the GA-ANN algorithm was selected, considering the prevailing recommendation to avoid training ANN with meta-heuristic algorithms. In this case, GA was employed to train an ANN, leveraging its established status as an evolutionary algorithm. Two scenarios for comparative analysis were investigated. **Scenario_1**: The baseline algorithms and the proposed framework were applied to predict wind power using a subset of selected features and undergoing data transformation. **Scenario_2**: The baseline algorithms and the proposed framework were applied to predict wind power using features without selection and data transformation. **Scenario_2**: The baseline algorithms and the proposed framework were applied to predict wind power without features selection and no data transformation. These scenarios were examined and observed to provide a comprehensive evaluation of the proposed framework in comparison to baseline frameworks (Table 2).

Results and discussion
This segment of the paper provides a presentation of the outcomes derived from the conducted experiments, which are documented in Tables 3, 4, 5, 6 and 7.

Hardware and software
Hardware and software details used to implement the study proposed framework include the use of an HP EliteBook 8570w with Windows 10, equipped with 8GB RAM and a 500GB Hard Disk. The processor is an Intel(R) Core(TM) i7-3520 M CPU @ 2.90 GHz, and the system maintains a robust and stable internet connection. The implementation platform utilized was Google Colab, employing Python libraries such as Keras, TensorFlow, NumPy, and Pandas.

Preliminary experiment
The preliminary experiments was conducted with varying combinations of hyper-parameters and small amount of datasets to obtain the optimal settings for deploying in the large scale experiment. Several experiments were conducted using varying values of hyper-parameters. For instance, the number of hidden neurons and layers were increasing during the experiments until the performance of the model start diminishing before it was stopped. Varying combinations of activation functions, batch size, step size and dropout regularization were tried several times to obtain the optimal values. The optimal hyper-parameter settings were obtained and reported in Table 2. It was clear that the model performance highly depends on the parameter settings as altering the parameter settings typically reflect on the performance of the model.

Hyper-parameter settings
The optimal hyper-parameter configurations selected for the algorithms, following the initial experiments, are outlined in Table 2. These settings were consistently applied throughout the main study.

Feature selection
Table 3 displays the optimal subset of features chosen by the NSGA-III after undergoing the feature selection process. Upon close observation of Table 3, it is evident that different combinations of subset features were selected using varying learning algorithms, resulting in distinct accuracies across these algorithms. Nevertheless, the most effective subset feature combination is identified with the ANN, which minimizes the subset features while maximizing accuracy. The best subset features, contributing significantly to wind power forecasting, have been singled out and highlighted in bold. NSGA-III achieves this by minimizing the number of features, selecting 12 out of the 25 available features as the optimal subset for predicting wind power.

The likely reason for selecting certain features over other features is because the selected features contains significant information relevance to contribute to wind energy prediction. On the other hand, irrelevant and redundant features as well as features with insignificant information were discarded. This process underscores the importance of quality data in improving algorithmic solutions. The likely reason why the NSGA-III selected features perform better than other algorithms is because it maximizes features diversity and minimizes features count and model error.

Discussion
Tables 4, 5, 6 and 7 comprise six columns. The initial column, positioned from the left, showcases both the proposed hybrid DRNN-LSTM + NSGA-III + Transformed data framework and the baseline framework algorithms (e.g. LSTM and ANN). The second column specifies the quantity of features (e.g. 25 means 25 features and 12 means 12 features selected by the NSGA-III as shown in Table 3) utilized for the wind power forecasting by each algorithm in the framework. The third column reports the computational time including the time spent at each step. The fourth and fifth columns illustrate the performance metrics in terms of Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), respectively. Lastly, the sixth column contains

Learning algorithm	Selected features (accuracy)	Accuracy on all features (%)
SVM	Wind Speed, Wind Turbins MW Operating Status, AC Primary Load, AC Primary Load Served, DC Primary Load, DC Primary Load Served, Grid Power Price, Grid Sellback Rate, Renewable Penetration, Rectifier Power Input, Generic 1 kWh Lead Acid Maximum Discharge Power, Generic 1 kWh Lead Acid Discharge Power, Generic 1 kWh Lead Acid Input Power, Generic 1 kWh Lead Acid Energy Content, Generic 1 kWh Lead Acid State of Charge, AC Required Operating Capacity, DC Required Operating Capacity, AC Operating Capacity, DC Operating Capacity	84.44
K-nearest neighbor	Wind Turbins MW Operating Status, AC Primary Load, AC Primary Load Served, DC Primary Load, DC Primary Load Served, Grid Power Price, Grid Sellback Rate, Renewable Penetration, Unmet Electrical Load, Rectifier Power Input, Generic 1 kWh Lead Acid Maximum Discharge Power, Generic 1 kWh Lead Acid Discharge Power, Generic 1 kWh Lead Acid Input Power, Generic 1 kWh Lead Acid Energy Content, Generic 1 kWh Lead Acid State of Charge, AC Required Operating Capacity, DC Required Operating Capacity, AC Operating Capacity, DC Operating Capacity	89.40
GNB	Wind Turbins MW Operating Status, AC Primary Load, AC Primary Load Served, DC Primary Load, DC Primary Load Served, Grid Power Price, Grid Sellback Rate, Renewable Penetration, Unmet Electrical Load, Rectifier Power Input, Generic 1 kWh Lead Acid Maximum Discharge Power, Generic 1 kWh Lead Acid Discharge Power, Generic 1 kWh Lead Acid Input Power, Generic 1 kWh Lead Acid Energy Content, Generic 1 kWh Lead Acid State of Charge, AC Operating Capacity, DC Operating Capacity	83.04
Random forest	Wind Speed, AC Primary Load, AC Primary Load Served, DC Primary Load, DC Primary Load Served, Grid Power Price, Grid Sellback Rate, Renewable Penetration, Unmet Electrical Load, Rectifier Power Input, Generic 1 kWh Lead Acid Discharge Power, Generic 1 kWh Lead Acid Input Power, Content, AC Required Operating Capacity, DC Required Operating Capacity, AC Operating Capacity, DC Operating Capacity	88.98
ANN	Wind Speed, AC Primary Load, AC Primary Load Served, Grid Power Price, Grid Sellback Rate, Renewable Penetration, Unmet Electrical Load, Rectifier Power Input, Generic 1 kWh Lead Acid Discharge Power, Generic 1 kWh Lead Acid Input Power, AC Required Operating Capacity, AC Operating Capacity (90.62)	**89.76**

Model	Features	AES (seconds)	MSE	RMSE	Remarks
LSTM	25	73 s 70 ms/step	9.9424e−04	0.0315	Data were transformed without undergoing feature selection
LSTM	25	20 s 19 ms/step	0.2440	0.4939	No data transformation and feature selection
LSTM	12	29 s 21 ms/step	0.2441	0.4940	Features were selected without undergoing any data transformation
DRNN	25	30 s 29 ms/step	0.0062	0.0787	Data were transformed without undergoing feature selection
DRNN	25	58 s 56 ms/step	0.2441	0.4940	No data transformation and feature selection
DRNN	12	31 s 29 ms/step	0.2445	0.4944	Features were selected without undergoing any data transformation
GANN	25	0 s 3 ms/step	0.2319	0.4815	Data were transformed without undergoing feature selection
GANN	25	2 s 2ms/step	0.2439	0.4938	No data transformation and feature selection
GANN	12	2 s 2 ms/step	0.2439	0.4938	Features were selected without undergoing any data transformation
ANN	25	2 s 2 ms/step	0.0056	0.0748	Data were transformed without undergoing feature selection
ANN	25	3 s 3 ms/step	0.5777	0.7600	No data transformation and feature selection
ANN	12	3 s 3 ms/step	0.5777	0.7600	Features were selected without undergoing any data transformation
DRNN-LSTM	25	18 s 17 ms/step	0.0010	0.0316	Data were transformed without undergoing feature selection
DRNN-LSTM + NSGA-III + Transformed data	**12**	**40 s 38 ms/step**	**1.6078e−04**	**0.0126**	NSGA-III was employed for features selection, and the data underwent a transformation process.

Algorithms	No. of inputs	AES (seconds)	MSE	RMSE	Remarks
LSTM	25	70 s 57 ms/step	7.7289e−07	8.791e−04	Data were transformed without undergoing feature selection
LSTM	25	68 s 56 ms/step	0.2441	0.4940	No data transformation and feature selection
LSTM	12	24 s 20 ms/step	0.2442	0.4941	Features were selected without undergoing any data transformation
DRNN	25	36 s 29 ms/step	0.0072	0.0848	Data were transformed without undergoing feature selection
DRNN	25	24 s 20 ms/step	0.2442	0.4941	No data transformation and feature selection
DRNN	12	36 s 29 ms/step	0.2446	0.4945	Features were selected without undergoing any data transformation
GANN	25	0 s 3 ms/step	0.2364	0.4862	Data were transformed without undergoing feature selection
GANN	25	2 s 2 ms/step	0.2440	0.4939	No data transformation and feature selection
GANN	12	2 s 2 ms/step	0.2440	0.4939	Features were selected without undergoing any data transformation
ANN	25	3 s 2 ms/step	0.0049	0.0700	Data were transformed without undergoing feature selection
ANN	25	4 s 3 ms/step	0.4231	0.6504	No data transformation and feature selection
ANN	12	4 s 4 ms/step	0.5769	0.7595	Features were selected without undergoing any data transformation
DRNN-LSTM	25	27 s 22 ms/step	8.8814e−07	9.424e−04	Data were transformed without undergoing feature selection
DRNN-LSTM + NSGA-III + Data Transformation	**12**	**25 s 21 ms/step**	**2.6593e−10**	**1.630e−05**	NSGA-III was employed for features selection, and the data underwent a transformation process.

Model	No. of inputs	AES (seconds)	MSE	RMSE	Remark
LSTM	25	89 s 64 ms/step	2.3920e−06	0.0015	Data were transformed without undergoing feature selection
LSTM	25	85 s 60 ms/step	0.2442	0.4941	No data transformation and feature selection
LSTM	12	28 s 20 ms/step	0.2442	0.4941	Features were selected without undergoing any data transformation
DRNN	25	28 s 20 ms/step	0.2441	0.4940	No data transformation and feature selection
DRNN	25	39 s 28 ms/step	0.0051	0.0714	Data were transformed without undergoing feature selection
DRNN	12	34 s 24 ms/step	0.2447	0.4946	Features were selected without undergoing any data transformation
DRNN-LSTM	25	62 s 45 ms/step	0.0025	0.050	Data were transformed without undergoing feature selection
GANN	25	0 s 4 ms/step	0.2233	0.4725	Data were transformed without undergoing feature selection
GANN	25	2 s 2 ms/step	0.2439	0.4938	No data transformation and feature selection
GANN	12	2 s 2 ms/step	0.2439	0.4938	Features were selected without undergoing any data transformation
ANN	25	2 s 2 ms/step	0.4233	0.6506	Data were transformed without undergoing feature selection
ANN	25	4 s 3 ms/step	0.4233	0.6506	No data transformation and feature selection
ANN	12	4 s 3 ms/step	0.4233	0.6506	Features were selected without undergoing any data transformation
DRNN-LSTM + NSGA-III + Data Transformation	**12**	**51 36 ms/step**	**7.8286e−06**	**0.0027**	NSGA-III was employed for features selection, and the data underwent a transformation process.

Algorithm	Features	AES (seconds)	MSE	RMSE	Remarks
LSTM	25	60 s 68 ms/step	0.0069	0.0830	Data were transformed without undergoing feature selection
LSTM	25	51 s 59 ms/step	0.2438	0.4937	No data transformation and feature selection
LSTM	12	17 s 19 ms/step	0.2438	0.4937	Features were selected without undergoing any data transformation
DRNN	25	27 s 31 ms/step	0.0062	0.0787	Data were transformed without undergoing feature selection
DRNN	25	18 s 21 ms/step	0.2439	0.4938	No data transformation and feature selection
DRNN	12	24 s 28 ms/step	0.2443	0.4942	Features were selected without undergoing any data transformation
GANN	25	0 s 3 ms/step	0.2351	0.4848	Data were transformed without undergoing feature selection
GANN	25	2 s 2 ms/step	0.2435	0.4934	Data were transformed without undergoing feature selection
GANN	12	2 s 2 ms/step	0.2435	0.4934	Features were selected without undergoing any data transformation
ANN	25	2 s 2 ms/step	3.2134e-05	0.0056	Data were transformed without undergoing feature selection
ANN	25	2 s 3 ms/step	0.5782	0.7603	No data transformation and feature selection
ANN	12	2 s 3 ms/step	0.4218	0.6494	Features were selected without undergoing any data transformation
DRNN-LSTM	25	17 s 19 ms/step	0.0018	0.0424	Data were transformed without undergoing feature selection
DRNN-LSTM + NSGA-III + Data Transformation	**12**	**21 s 24 ms/step**	**1.0509e−09**	**3.241e−05**	NSGA-III was employed for features selection, and the data underwent a transformation process.

remarks, offering additional insights on data transformation and feature selection specific to each framework. The optimal solution for the wind power forecasting is provided.

Achieving optimality is not consistently attainable. The pursuit of optimality is made more intricate by the persistent presence of uncertainty in real-world systems. Consequently, the objective extends beyond attaining optimality to also encompass robustness. Practicality dictates that optimal solutions must exhibit a level of robustness, as those lacking in resilience are not viable in real-world applications[77]. Therefore, the study decided to measure the robustness of the proposed framework. Considering that data partition ratio significantly influences the performance of training algorithms, as indicated by Yu et al.[78], and given the absence of a standardized data partition ratio within the research community, this study opted to employ multiple data partition ratios. This approach aims to evaluate the robustness of the proposed framework across various data partition ratios, following a methodology akin to the one employed by Chiroma et al.[79]. Tables 3, 4, 5, 6 and 7 present the results corresponding to the different data partition ratios: 60%:40%, 70%:30%, 80%:20%, and, finally, 50%:50%.

As evident in Tables 4, 5, 6 and 7, the proposed DRNN-LSTM + NSGA-III + transformed data, surpasses the compared frameworks by reducing the error between the actual and forecasted wind power to minimal level. The likely reason why the proposal performs better is the possibility of LSTM stabilizing the DRNN-LSTM training.

The examination of results also highlights the beneficial impact of feature selection in improving the accuracy compared to scenarios where no feature selection is applied. Furthermore, it is observed that data transformation contributes to increase accuracy compared to feature selection without data transformation (see the remark columns). The performance of the baseline algorithms—LSTM, DRNN, ANN, and GANN—when evaluated among themselves, indicates that the use of transformed data enhances the accuracy of these baseline algorithms compared to scenarios where no data transformation is applied.

Regarding computational time, the proposed DRNN-LSTM + NSGA-III + transformed data framework exhibited prolonged training durations, particularly when training the hybrid DRNN-LSTM algorithm for wind power prediction, with a few instances surpassing the time taken by LSTM and DRNN alone. This outcome aligns with the common understanding that deep learning architectures typically necessitate extended training time compared to shallow algorithms, as noted in the literature (e.g., Ref[80]). Given that the current study involves training a hybrid DRNN-LSTM to forecast wind power, a higher computational time was anticipated, acknowledging this as a known limitation of deep learning architectures. Examination of Tables 4, 5, 6 and 7 consistently shows that shallow algorithms consistently converge more rapidly in comparison to deep learning architectures.

The performance patterns of the proposed DRNN-LSTM + NSGA-III + transformed data, as evident in Tables 4, 5, 6 and 7, suggest that the volume of training data is directly proportional to the computational time. Specifically, as the amount of wind power training data expands and test data diminishes, the computational time needed for wind power forecasting also increases. Conversely, when the training data decreases and test data increases, the computational time decreases. The results clearly indicate that the accuracy of the wind power model is influenced by the chosen data partition ratio, as evidenced by varying accuracies across different ratios. Notably, the optimal accuracy for the proposed framework was achieved at the 70%:30% partition ratio during the modeling of the algorithm for wind power forecasting.

Typically, the training partition ratio allocates a substantial portion of the data to the training phase as practice in the literature, resulting in a model biased towards that training portion. In this study, to ensure a fair evaluation and assess the robustness of the proposed framework under equal data partition, the data were evenly divided: 50%: 50%. The corresponding results are presented in Table 7. Surprisingly, the accuracy lags only behind the best accuracy observed with a data partition ratio of 70%:30% with a very little merging. Furthermore, upon closer inspection, it is noted that the computational time is reduced by 4 s compared to the computational time recorded with the 70%:30% partition ratio.

The study outcomes, as showcased in Tables 4, 5, 6 and 7, lead to the conclusion that the proposed framework consistently demonstrates improved performance in terms of accuracy and robustness across various data partition ratios. However, it is noted that the LSTM shows competitive performance at the 80:20 partition but weaken in the other partition ratios, this indicate lack of robustness and consistent performance of the LSTM in this domain. The likely reason why the proposed framework maintained better robustness and consistent performance across different conditions compared to the standard frameworks is because of the integration of transformation mechanism that normalized the wind energy features leading to balance foundation for the DRNN-LSTM + NSGA-III to ensure uniform feature contributions without bias towards higher values. The integration of evolutionary feature selection algorithm and data transformation significantly enhances the wind power forecasting.

This research contributes to the expanding field of wind energy forecasting literature by extending the baseline framework through the incorporation of evolutionary feature selection algorithm and mechanism for data transformation. Consequently, it enhances the performance of previously discussed frameworks documented in the literature.

Implication for theory and practice

The optimal features identified by the proposed framework can be employed by the research community for future wind energy data collection, eliminating the necessity to gather redundant features. This approach helps minimize computational costs, enhance productivity, reduce data collection time, and improve overall accuracy. The proposed framework exhibits the potential for real-world integration into power systems due to its robustness in possibility to address real world uncertainties. The accurate prediction of wind power, facilitated by a reliable forecasting system, can offer computed values essential for future planning. Decision-makers can rely on the robust computational wind energy values to formulate policies, especially in the context of future renewable energy, specifically wind energy. The study encourages the research community to adopt evolutionary feature selection algorithms and implement data transformation in future wind energy forecasting systems.

Limitations

The study used only numerical values datasets, thus, limit the capacity of the wind energy forecasting model in handling other modalities such as text and image. Therefore, wind energy information in text or image cannot be handled by the proposed model. In addition, external variables like the extreme weather conditions, industrial operations and weather forecast were not incorporated in the proposed wind energy forecasting model, thus, limit the robustness of the proposed model. During the experiment, we expect that, as the size of the training data increases while the test data decreases, the accuracy increases. However, we encountered a different scenario where 70:30 data partition ratio performed better than the 80:20 in terms of MSE and RSME.

Conclusions and suggestion for future research

The research introduces an innovative framework aimed at enhancing wind power prediction. This proposed framework integrates a data transformation mechanism with a multi-objective NSGA-III feature selection algorithm, coupled with a hybrid DRNN-LSTM. The DRNN was combined with LSTM and embedded with the multi-objective NSGA-III, incorporating a data transformation mechanism. The multi-objective NSGA-III selected optimal wind power features, which were then normalized in the data transformation module before being inputted into DRNN-LSTM for wind power forecasting. Results from the study demonstrated that the multi-objective NSGA-III's feature selection enhanced performance, and the additional data transformation further improved overall performance. It is evident that the hybrid DRNN-LSTM surpasses individual algorithms and genetically optimized ANN. It is worthy to note that LSTM shows competitive performance at the 80:20 data partition ratio, but, exhibit lack of robustness and consistent performance in wind energy prediction. Comparative analysis highlights the superior performance of the proposed wind power modeling framework compared to baseline algorithms. The study advocates for the adoption of evolutionary feature selection algorithms and the integration of data transformation in future wind energy forecasting systems to be proposed by the research community. It will be interesting in the future to adopt multimodal transformer for the forecasting of wind energy incorporating multimodalities (e.g., combination of text, numerical values and image) with capacity to handle varying datasets for improving the robustness and performance of the model in real world environment. In addition, weather forecast can be incorporated in the model to boost the model performance. The paper suggested future work to focus on reducing training time of deep learning architecture to predict wind power.

Data availability

The data can be requested by contacting professor Yahya Z. Alharthi via yalharthi@uhb.edu.sa.

Received: 8 February 2024; Accepted: 10 April 2025
Published online: 08 May 2025

References

1. Li, M. et al. A wind speed forecasting model based on multi-objective algorithm and interpretability learning. *Energy* **269**, 126778 (2023).
2. Wekesa, D. W., Wang, C., Wei, Y., Kamau, J. N. & Danao, L. A. M. A numerical analysis of unsteady inflow wind for site specific vertical axis wind turbine: A case study for Marsabit and Garissa in Kenya. *Renew. Energy* **76**, 648–661 (2015).
3. Global wind Report. Global wind energy Council (2021). (https://gwec.net/global-wind-report-2021/).
4. Verma, M., Ghritlahre, H. K. & Chandrakar, G. Wind speed prediction of central region of Chhattisgarh (India) using artificial neural network and multiple linear regression technique: A comparative study. *Ann. Data Sci.* **10**(4), 851–873 (2023).
5. Peiris, A. T., Jayasinghe, J. & Rathnayake, U. Forecasting wind power generation using artificial neural network:"Pawan Danawi"—A case study from Sri Lanka. *J. Electr. Comput. Eng.* **2021**, 1–10 (2021).
6. Wu, Q., Zheng, H., Guo, X. & Liu, G. Promoting wind energy for sustainable development by precise wind speed prediction based on graph neural networks. *Renew. Energy* **199**, 977–992 (2022).
7. Donadio, L., Fang, J. & Porté-Agel, F. Numerical weather prediction and artificial neural network coupling for wind energy forecast. *Energies* **14**(2), 338 (2021).
8. Zhang, Y., Li, P., Li, H., Zu, W. & Zhang, H. Short-term power prediction of wind power generation system based on logistic chaos atom search optimization BP neural network. *Int. Trans. Electr. Energy Syst.* **2023**(1), 6328119 (2023).
9. Sun, S., Liu, Y., Li, Q., Wang, T. & Chu, F. Short-term multi-step wind power forecasting based on spatio-temporal correlations and transformer neural networks. *Energy Convers. Manag.* **283**, 116916 (2023).
10. Qureshi, A. S., Khan, A., Zameer, A. & Usman, A. Wind power prediction using deep neural network based meta regression and transfer learning. *Appl. Soft Comput.* **58**, 742–755 (2017).
11. Sola, J. & Sevilla, J. Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Trans. Nucl. Sci.* **44**(3), 1464–1468 (1997).
12. Bhanja, S. & Das, A. Impact of data normalization on deep neural network for time series forecasting. *arXiv preprint* arXiv:1812.05519. (2018).
13. Passalis, N., Tefas, A., Kanniainen, J., Gabbouj, M. & Iosifidis, A. Deep adaptive input normalization for time series forecasting. *IEEE Trans. Neural Netw. Learn. Syst.* **31**(9), 3760–3765 (2019).
14. Chakraborty, G. & Chakraborty, B. A novel normalization technique for unsupervised learning in ANN. *IEEE Trans. Neural Netw.* **11**(1), 253–257 (2000).
15. Beale, M. H., Hagan, M. T. & Demuth, H. B. Neural network toolbox. *User's Guide, MathWorks* **2**, 77–81 (2010).
16. Singh, U. & Singh, S. N. Optimal feature selection via NSGA-II for power quality disturbances classification. *IEEE Trans. Industr. Inf.* **14**(7), 2994–3002 (2017).
17. Xue, Y., Tang, Y., Xu, X., Liang, J. & Neri, F. Multi-objective feature selection with missing data in classification. *IEEE Trans. Emerg. Top. Comput. Intell.* **6**(2), 355–364 (2021).

18. Zhang, M. et al. Multi-objective optimization algorithm based on clustering guided binary equilibrium optimizer and NSGA-III to solve high-dimensional feature selection problem. *Inf. Sci.* **648**, 119638 (2023).
19. Marugán, A. P., Márquez, F. P. G., Perez, J. M. P. & Ruiz-Hernández, D. A survey of artificial neural network in wind energy systems. *Appl. Energy* **228**, 1822–1836 (2018).
20. Nazir, M. S. et al. Wind generation forecasting methods and proliferation of artificial neural network: A review of five years research trend. *Sustainability* **12**(9), 3778. https://doi.org/10.3390/su12093778 (2020).
21. Ferrero Bermejo, J., Gómez Fernández, J. F., Olivencia Polo, F. & Crespo Márquez, A. A review of the use of artificial neural network models for energy and reliability prediction. A study of the solar PV, hydraulic and wind energy sources. *Appl. Sci.* **9**(9), 1844 (2019).
22. Wang, Y., Zou, R., Liu, F., Zhang, L. & Liu, Q. A review of wind speed and wind power forecasting with deep neural networks. *Appl. Energy* **304**, 117766 (2021).
23. Yin, T. et al. A robust multilabel feature selection approach based on graph structure considering fuzzy dependency and feature interaction. *IEEE Trans. Fuzzy Syst.* **31**(12), 4516–4528. https://doi.org/10.1109/TFUZZ.2023.3287193 (2023).
24. Liu, X., Cao, Z. & Zhang, Z. Short-term predictions of multiple wind turbine power outputs based on deep neural networks with transfer learning. *Energy* **217**, 119356 (2021).
25. Zhao, Z. et al. Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features. *Eng. Appl. Artif. Intell.* **121**, 105982 (2023).
26. Mei, Y. et al. Superposition graph neural network for offshore wind power prediction. *Fut. Gener. Comput. Syst.* **113**, 145–157. https://doi.org/10.1016/j.future.2020.06.024 (2020).
27. Khalid, S., Khalil, T. & Nasreen, S. A survey of feature selection and feature extraction techniques in machine learning. In *2014 Science and Information Conference* 372–378 (IEEE, 2014).
28. Puri, V. & Kumar, N. Wind energy forecasting using artificial neural network in Himalayan region. *Model. Earth Syst. Environ.* **8**(1), 59–68 (2022).
29. Pasari, S., Shah, A. & Sirpurkar, U. Wind energy prediction using artificial neural networks. In *Enhancing Future Skills and Entrepreneurship: 3rd Indo-German Conference on Sustainability in Engineering* 101–107 (Springer International Publishing, 2020).
30. Ti, Z., Deng, X. W. & Zhang, M. Artificial Neural Networks based wake model for power prediction of wind farm. *Renewable Energy* **172**, 618–631 (2021).
31. Khosravi, A. K. R. M. L. P. J., Koury, R. N. N., Machado, L. & Pabon, J. J. G. Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system. *Sustain. Energy Technol. Assess.* **25**, 146–160 (2018).
32. Medina, S. V. & Ajenjo, U. P. Performance improvement of artificial neural network model in short-term forecasting of wind farm power output. *J. Modern Power Syst. Clean Energy* **8**(3), 484–490 (2020).
33. Agrawal, P., Abutarboush, H. F., Ganesh, T. & Mohamed, A. W. Metaheuristic algorithms on feature selection: A survey of one decade of research (2009–2019). *IEEE Access* **9**, 26766–26791 (2021).
34. Kisvari, A., Lin, Z. & Liu, X. Wind power forecasting–A data-driven method along with gated recurrent neural network. *Renew. Energy* **163**, 1895–1909 (2021).
35. Lin, Z. & Liu, X. Wind power forecasting of an offshore wind turbine based on high-frequency SCADA data and deep learning neural network. *Energy* **201**, 117693 (2020).
36. Wei, J., Wu, X., Yang, T. & Jiao, R. Ultra-short-term forecasting of wind power based on multi-task learning and LSTM. *Int. J. Electr. Power Energy Syst.* **149**, 109073 (2023).
37. University of Pritoria. Feature selection problems. Chapter 3. Ph.D. Thesis, University of Pritoria, Pretoria, South Africa, 2011. (2022). Available online: https://repository.up.ac.za/bitstream/handle/2263/29486/02chapter3-.pdf?sequence=3&isAllowed=y (Accessed on 14 Dec 2023)
38. Almutairi, M. S., Almutairi, K. & Chiroma, H. Selecting features that influence vehicle collisions in the internet of vehicles based on a multi-objective hybrid bi-directional NSGA-III. *Appl. Sci.* **13**(4), 2064 (2023).
39. Zhou, P., Li, P., Zhao, S. & Wu, X. Feature interaction for streaming feature selection. *IEEE Trans. Neural Netw. Learn. Syst.* **32**(10), 4691–4702 (2020).
40. Liyanage, Y. W., Zois, D. S. & Chelmis, C. Dynamic instance-wise joint feature selection and classification. *IEEE Trans. Artif. Intell.* **2**(2), 169–184 (2021).
41. Nielson, J., Bhaganagar, K., Meka, R. & Alaeddini, A. Using atmospheric inputs for Artificial Neural Networks to improve wind turbine power prediction. *Energy* **190**, 116273 (2020).
42. Zafirakis, D., Tzanes, G. & Kaldellis, J. K. Forecasting of wind power generation with the use of artificial neural networks and support vector regression models. *Energy Proced.* **159**, 509–514 (2019).
43. Shabbir, N., Kütt, L., Jawad, M., Amadiahanger, R., Iqbal, M. N. & Rosin, A. Wind energy forecasting using recurrent neural networks. In *2019 Big Data, Knowledge and Control Systems Engineering (BdKCSE)* 1–5 (IEEE, 2019).
44. Huang, B., Liang, Y. & Qiu, X. Wind power forecasting using attention-based recurrent neural networks: A comparative study. *IEEE Access* **9**, 40432–40444 (2021).
45. Karaman, Ö. A. Prediction of wind power with machine learning models. *Appl. Sci.* **13**(20), 11455 (2023).
46. Dorado-Moreno, M. et al. Multi-task learning for the prediction of wind power ramp events with deep neural networks. *Neural Netw.* **123**, 401–411 (2020).
47. Liao, W. et al. Ultra-short-term interval prediction of wind power based on graph neural network and improved bootstrap technique. *J. Modern Power Syst. Clean Energy* **11**(4), 1100–1114. https://doi.org/10.35833/MPCE.2022.000632 (2023).
48. Fan, W. et al. Interactive reinforcement learning for feature selection with decision tree in the loop. *IEEE Trans. Knowl. Data Eng.* https://doi.org/10.1109/TKDE.2021.3102120 (2021).
49. Jiao, R., Nguyen, B. H., Xue, B. & Zhang, M. A survey on evolutionary multiobjective feature selection in classification: approaches, applications, and challenges. *IEEE Trans. Evol. Comput*. (2023).
50. Hu, Z., Dong, Y., Wang, K., Chang, K. W. & Sun, Y. Gpt-gnn: Generative pre-training of graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 1857–1867 (2020).
51. Wang, J., Zhu, H., Zhang, Y., Cheng, F. & Zhou, C. A novel prediction model for wind power based on improved long short-term memory neural network. *Energy* **265**, 126283 (2023).
52. Xiao, Y., Zou, C., Chi, H. & Fang, R. Boosted GRU model for short-term forecasting of wind power with feature-weighted principal component analysis. *Energy* **267**, 126503 (2023).
53. Alshammari, A. Generation forecasting employing deep recurrent neural network with Metaheruistic feature selection methodology for renewable energy power plants. *Sustain. Energy Technol. Assess.* **55**, 102968 (2023).
54. Ak, R., Li, Y. F., Vitelli, V. & Zio, E. Adequacy assessment of a wind-integrated system using neural network-based interval predictions of wind power generation and load. *Int. J. Electr. Power Energy Syst.* **95**, 213–226 (2018).
55. Jin, N. & Liu, D. Wavelet basis function neural networks for sequential learning. *IEEE Trans. Neural Netw.* **19**(3), 523–528 (2008).
56. Aires, L. F. et al. Convolutional neural networks using the SMOTE algorithm and features fusion for wind turbine fault prediction. *IEEE Lat. Am. Trans.* **23**(3), 191–197 (2025).
57. Sankar, S. R. WindForecastX: A dynamic approach for accurate long-term wind speed prediction in wind energy applications. *Ocean Dyn.* **75**(1), 1–25 (2025).

58. Wang, J., Che, J., Li, Z., Gao, J. & Zhang, L. Hybrid wind speed optimization forecasting system based on linear and nonlinear deep neural network structure and data preprocessing fusion. *Futur. Gener. Comput. Syst.* **164**, 107565 (2025).
59. Liao, W., Su, X. & Fang, F. A centrifugal spring mechanism empowers self-adjusting in piezoelectric wind energy harvesting. *Nano Energy* **133**, 110462 (2025).
60. Hou, G., Li, Q. & Huang, C. Spatiotemporal forecasting using multi-graph neural network assisted dual domain transformer for wind power. *Energy Convers. Manag.* **325**, 119393 (2025).
61. Songyue, L. I. U., Qiusheng, L. I., Bin, L. U. & Junyi, H. E. Prediction of offshore wind turbine wake and output power using large eddy simulation and convolutional neural network. *Energy Convers. Manag.* **324**, 119326 (2025).
62. Bashir, T., Wang, H., Tahir, M. & Zhang, Y. Wind and solar power forecasting based on hybrid CNN-ABiLSTM, CNN-transformer-MLP models. *Renew. Energy* **239**, 122055 (2025).
63. Khan, S., Muhammad, Y., Jadoon, I., Awan, S. E. & Raja, M. A. Z. Leveraging LSTM-SMI and ARIMA architecture for robust wind power plant forecasting. *Appl. Soft Comput.* **170**, 112765 (2025).
64. Wang, Q. et al. Towards machine learning applications for structural load and power assessment of wind turbine: An engineering perspective. *Energy Convers. Manag.* **324**, 119275 (2025).
65. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**(6088), 533–536 (1986).
66. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997).
67. Pascanu, R., Gulcehre, C., Cho, K. & Bengio, Y. How to construct deep recurrent neural networks. *arXiv preprint* arXiv:1312.6026. (2013).
68. Hermans, M. & Schrauwen, B. Training and analysing deep recurrent neural networks. *Adv. Neural Inf. Process. Syst.* **26**, (2013).
69. Ma, X., Tao, Z., Wang, Y., Yu, H. & Wang, Y. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transp. Res. Part C Emerg. Technol.* **54**, 187–197 (2015).
70. Sak, H., Senior, A. & Beaufays, F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling has. In *Fifteenth Annual Conference of the International Speech Communication Association.* (2014).
71. Khan, M., Jan, B., Farman, H., Ahmad, J., Farman, H. & Jan, Z. Deep learning methods and applications. *Deep Learn. Converg. Big Data Anal.* 31–42 (2019).
72. Yu, Y., Si, X., Hu, C. & Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* **31**(7), 1235–1270 (2019).
73. Ishibuchi, H., Setoguchi, Y., Masuda, H. & Nojima, Y. Performance of decomposition-based many-objective algorithms strongly depends on Pareto front shapes. *IEEE Trans. Evol. Comput.* **21**(2), 169–190 (2016).
74. Saudi Vision 2023 (2022). Annual Report. Available at https://www.vision2030.gov.sa/ar/ (Accessed on 15 Dec 2022).
75. HOMER. (2023). HOMER Software documentation. Available at www.homerenergy.com (Accessed 22 Apr 2023).
76. Graves, A. & Graves, A. (2012). Long short-term memory. *Supervised sequence Labelling with Recurrent Neural Networks* 37–45.
77. Yang, X. S. & Deb, S. Cuckoo search: Recent advances and applications. *Neural Comput. Appl.* **24**, 169–174 (2014).
78. Yu, L., Wang, S. & Lai, K. K. A neural-network-based nonlinear metamodeling approach to financial time series forecasting. *Appl. Soft Comput.* **9**(2), 563–574 (2009).
79. Chiroma, H. et al. A new approach for forecasting OPEC petroleum consumption based on neural network train by using flower pollination algorithm. *Appl. Soft Comput.* **48**, 50–58 (2016).
80. Chiroma, H., Gital, A. Y. U., Rana, N., Abdulhamid, S. I. M., Muhammad, A. N., Umar, A. Y., & Abubakar, A. I. (2020). Nature inspired meta-heuristic algorithms for deep learning: recent progress and novel perspective. In *Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 1 1* (pp. 59–70). Springer International Publishing.

Acknowledgements
Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R178), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author contributions
Conceptualization, Y.Z.A. and H.C.; methodology, Y.Z.A and H.C.;formal analysis, Y.Z.A.; investigation, Y.Z.A ; data curation, Y.Z.A and H.C; writing—original draft preparation, Y.Z.A. and H.C.; writing—review and editing, H.C.; visualization, Y.Z.A. and L.A.G: Revised after revision decision for intellectual content.

Declarations

Competing interests
The authors declare no competing interests.

Declarations
The authors declare that there is no conflict of interest.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-98212-8.

Correspondence and requests for materials should be addressed to Y.Z.A., H.C. or L.A.G.