Renewable Energy 218 (2023) 119241

Contents lists available at ScienceDirect

Renewable Energy

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

Offshore wind power forecasting based on WPD and optimised deep learning methods

Shahram Hanifi [a,*], Hossein Zare-Behtash [a], Andrea Cammarano [a], Saeid Lotfian [b]

[a] James Watt School of Engineering, University of Glasgow, G12 8QQ, Glasgow, UK

[b] Department of Naval Architecture, Ocean and Marine Engineering, University of Strathclyde, G4 0LZ, Glasgow, UK

ARTICLE INFO

Keywords:
Wind power forecasting
Wavelet packet decomposition (WPD)
Convolutional neural network (CNN)
Long short-term memory (LSTM)
Optuna

ABSTRACT

Accurate wind power forecasting is vital for (i) wind power management, (ii) penetration increment of the power generated into the power grid, and (iii) making maintenance more efficient. Motivated by the recent application of wavelet transforms and advancements in deep learning methods, a hybrid forecasting method is developed based on the wavelet packet decomposition [WPD], Long Short-Term Memory Network [LSTM], and Convolutional Neural Network [CNN] to improve the accuracy of wind power forecasting. WPD is employed to decompose pre-processed wind power data into sublayers with different frequencies. Sequential Model-Based optimisation (SMBO) with the Tree Parzen Estimator (TPE) is then used to tune the hyper-parameters of LSTM and CNN, efficiently. The optimised LSTM is employed to predict the low-frequency sub-layer that has both long-term and short-term dependencies, and CNN is used to forecast the high-frequency sub-layers with short-term dependencies. To evaluate the prediction performance of the developed method, seven forecasting models, including random forest (RF), feed-forward neural network (FFNN), CNN, LSTM, WPD-FFNN, WPD-CNN, and WPD-LSTM models, are considered as comparison models. Comparing the prediction results of all involved models proves that the developed model improves the prediction accuracy by at least 77.4% compared to methods that do not use WPD. In addition, the proposed combination of optimised CNN and LSTM improves the forecasting accuracy by 26.25% compared to methods that use only one deep learning model to forecast all sub-series.

1. Introduction

Before the current gauge-political in Europe, the world was rapidly developing renewable energy production to reduce the harmful effects of burning fossil fuels, including pollution, climate change, and depletion of the ozone layer [1]. Numerous international agreements have been made between countries around the world in recent years to move away from burning fossil fuels for energy production. Paris Agreement 2015 to reduce global greenhouse gas emissions by limiting the global temperature to 2 °C above the pre-industrial level in century 21st, as well as the recent Glasgow Climate Pact (COP26) to keep 1.5 °C, can be mentioned as examples. These agreements are based on an increased focus towards renewable energy sources [2]. The war in Ukraine has disrupted the global energy supply, and the rapid 'hike' in the price of oil and gas due to the following economic sanctions (in response to the invasion) has further highlighted the need to develop renewable energy sources [1].

During the last decade, by amplifying the impact of climate and environmental changes, governments and scientists increasingly considered a variety of renewable energy resources such as wind, solar, wave, tidal, etc. They started to address the existing and rising challenges in harvesting renewable energy generation. Among all those resources, wind power is a success story and plays a critical role in replacing fossil fuels [3]. However, the high level of uncertainty due to wind speed fluctuations still acts as the main obstacle in front of wind power penetration into the power grids [1]. Accurate wind power forecasting is necessary for proportional electricity distribution planning to meet consumers (industrials and households) demands. It is also important for determination of the optimum operating conditions and maintenance planning to reduce the levelised cost of energy for the fair pricing objective.

In general, two steps need to be taken to develop an accurate wind power forecasting model: (i) prediction and (ii) optimisation. While the prediction aspect is carried out by the application of prediction methods,

* Corresponding author. R703C Level 7, School of Engineering, James Watt South Building, Glasgow, G12 8QQ, UK.
E-mail address: Shahram.hanifi@Glasgow.ac.uk (S. Hanifi).

https://doi.org/10.1016/j.renene.2023.119241
Received 17 November 2022; Received in revised form 21 August 2023; Accepted 28 August 2023
Available online 31 August 2023

the optimisation step improves the prediction performance through signal processing or parameter optimisation of the prediction models [4].

Power forecasting methods are categorised into three main methods, including physical, statistical and hybrid approaches. Physical processes utilise Numerical Weather Prediction (NWP) data, geographic descriptions of wind turbines/farms and weather information to model onsite conditions. The provided model is then used for wind speed and wind power predictions. These models are computationally complex and therefore require considerable computing resources. In addition, due to the high sensitivity to initial conditions, there is a need for synchronisation with other methods to improve prediction performance [3].

A precise mapping between input variables (e.g. NWP data, historical data, etc.) and target variables (like wind speed or wind power) is the foundation of statistical approaches [3]. Time series-based and machine learning (ML) are the two main methodologies used in these techniques [1]. The wind speed or wind power predictions in time series-based approaches are achieved by application of the historical forecasted variable itself. Time series-base methods have employed for very short-term (a few minutes to a few hours) prediction as they are able to recognise the hidden random features of wind speed. One of the most popular statistical techniques is the autoregressive integrated moving average (ARIMA) model established by Box-Jenkins [5], though it cannot handle non-linear dependencies. By understanding the correlations between input and output variables, ML techniques like neural networks (NNs) can build deductive models. These techniques are simple to develop, do not need more geographic details or specifics about wind farms or turbines, and have a more significant time horizon for prediction. However, their performance is highly dependent on the selected parameters. As a result, these methods are combined with various methods of feature selection or parameter optimisation to improve prediction performance.

Hybrid methods combine different methods to utilise their unique merits and improve the overall prediction accuracy. Table 1 presents some proposed hybrid methods in the recent years, along with their main features and performance improvements. What is clear is that although combining different methods improves overall performance, on the other hand it complicates the model and increases the required computation time. Hence, it is vital to obtain a balance between accuracy and efficiency.

Choosing the appropriate input features is also critical for the accuracy and reliability of wind power forecasting models. As shown in Table 1, wind power is used more often than other variables in wind power prediction models developed over the past few years. Nevertheless, some research indicates that additional inputs may improve wind power forecasts. For example [6], et al. in the development of a wind power forecasting model based on fandom forest (RF) showed that more accurate predictions can be made when the average wind speed and wind direction are added to the wind power as the input features. However, adding these features sometimes degrades the results significantly. In another research, Velazquez et al. [7] investigated the impact of wind speed, wind power density, and power output on the performance of ANN models. According to the findings, considering wind direction as an input can decrease the forecasting errors.

The accuracy of deep learning methods used in hybrid wind power prediction methods strongly depends on the proper selection of hyper-parameters [15]. Instead of manual tuning which can be extremely time consuming, grid and random search are used widely to set up a network of hyper-parameters and then run the train, predict and evaluation cycle automatically [16]. Nevertheless, without considering the past evaluated hyper-parameters, these tuning methods are relatively inefficient as they spend a significant amount of time evaluating improper hyper-parameters, i.e., wrong selection of activation functions of deep learning models. Bayesian model-based methods in contrast, through evaluation of hyper-parameters that appear more promising in the past results, can find better hyper-parameters in less time [17].

Table 1
Hybrid wind power prediction models.

Combined model	Year	Inputs	Accuracy improvement	Features
BPNN, RBFNN, LSSVM [8]	2017	wind speed & direction, temp.	Significant improvement in accuracy.	Pearson correlation coefficient (PCC) is used to improve the mapping accuracy
MODA, ELNN [9]	2018	Wind speed	43.96% MAPE reductions compared to comparison models	MODA application for ELNN optimisation
LSTM, GMM [10]	2019	wind speed	Up to 4.96% RMSE improvement over traditional methods	LSTM used for prediction and GMM for uncertainty description
CNN, RBFNN, DGF [11]	2019	wind power	accurate than traditional models for 24 h-ahead wind power prediction	the novel double Gaussian function (DGF) employed for RBFNN
ICEEMDAN, MOMFO, Wavelet NN [12]	2020	wind power	62.38% improvement in MAPE compared to wavelet NN	A robust hybrid method with appropriate accuracy and stability
GA, LSTM [13]	2021	wind power	Up to 30% accuracy improvement compared to existing methods	GA application for LSTM window size and neurons number optimisation
IF, GRU, LSTM [14]	2021	wind power	IF filtering improved forecasting performance by over 92%.	A robust model with less sensitivity to noise in SCADA data

Signal processing methods such as data decomposition, data denoising or data feature selection can effectively improve the accuracy of power forecasting methods. All decomposition-based forecasting models published in the literature use the same framework. In this framework, the original non-stationary time series is decomposed into stationary sub-series. Then, independent forecasting models are used to predict each sub-series. Finally, all predictions are added together to form the final forecast. Independent forecasting of each sub-series can efficiently enhance the prediction accuracy [18].

Su et al. [19] decomposed the wind speed data into four low-frequency and four high-frequency components by WPD. Then the four high-frequency components were decomposed into 60 intrinsic mode functions (IMFs) through ensemble empirical mode decomposition (EEMD). These components were then fed to individual LSTM models with yaw error and rotor speed data. The power prediction results of the proposed approach showed an improvement in accuracy. However, the effect of the direct application of the wind power dataset for prediction was not investigated. Zu et al. [20] used WPD to decompose wind power time series into three levels. The gained sub-series were fed to a gated recurrent unit (GRU), and the predictions were reconstructed to obtain the results. Experimental results showed that the proposed WPD-GRU-SELU model has a higher prediction accuracy than other Recurrent Neural Network (RNN) models. In another research, Mujeeb et al. [21] combined Wavelet Packet Transform (WPT) and Deep convolutional neural network (DCNN) to predict the day-ahead hourly wind power of ISO New England's wind farm, however, the authors did not attempt to forecast the sub-series with different independent methods. In addition to WPD, other wavelet transform methods have recently been used in the wind power prediction field. For example, Azimi et al. [22], with a combination of the K-means clustering method with discrete wavelet transform (DWT) and multilayer perceptron neural network (MLPNN), improved the wind power forecasting accuracy of the National Renewable Energy Laboratory (NREL). Shi et al. [23] employed variational mode decomposition (VMD) and LSTM to provide hourly predictions of day-ahead wind power of a Chinese wind farm. In another study, Liu et al. [24] combined empirical mode decomposition (EMD), LSTM and Elman neural network (ENN) to develop a hybrid model and obtained satisfactory results for multi-step wind speed predictions. To obtain better forecasting results, some researchers use error correction mechanisms through application of the double decomposition methods. For example, Ma et al. [25] used this decomposition approach with LSTM model and proved the better prediction performance of the proposed model than models without double decomposition.

This research proposes a novel hybrid forecasting model for 10-min-ahead wind power forecasting of an offshore wind turbine in Scotland. The proposed model is based on applying WPD, optimised CNN and LSTM models without imputing future weather forecasting data. The proposed model is applicable when a complex non-linear relationship between the variables exists in the wind power time series data. The novelty of this research lies in creating an ensemble forecasting method by intelligent combination of WPD and sequential model-based optimised (SMBO) CNN and LSTM models for short-term offshore wind power forecasting. WPD can decompose the original time series data into different sub-series with different frequencies. With access to both high-frequency and low-frequency components, patterns and trends displayed in wind power time series can be extracted and analyzed more effectively. In this way, it is possible to focus on the most important features of the data while reducing noise and irrelevant details. In this study, various mother wavelets and decomposition levels have been analyzed for the first time, to determine which would yield the best performance for WPD. Deep learning models with different performance in learning short-term and long-term dependencies can extract linear and non-linear relations from historical wind power data to make an accurate prediction. The present study is the first to employ CNN models for forecasting high-frequency components of wind power time series, and LSTM models for forecasting low-frequency components. In addition, the hyperparameters of both the LSTM and CNN models were tuned by SMBO with the Tree Parzen Estimator (TPE) that has never been used in the wind power prediction field. This method can significantly increase hyperparameter selection speed, resulting in improved prediction accuracy and efficiency.

2. Methodology

The framework of the proposed WPD-LSTM-CNN model is demonstrated in Fig. 1, and the entire process is depicted in detail in the following steps:

Step 1) The raw SCADA data of an offshore wind turbine are pre-processed by removing the negative power values as obvious outliers, imputing the missed data, and resampling. The detail of the pre-processing part is described in section 3.
Step 2) WPD decomposes the pre-processed wind power time series into several approximations and detail coefficients (sub-series). The detail of the decomposition method is described in section 2.1.
Step 3) The optimised CNN with surrogate optimisation method is used to predict the high-frequency sub-layers obtained from the WPD. Details and structure of this method are provided in section 2.2.
Step 4) The LSTM is tuned and employed for predicting the low-frequency sub-layer with details described in section 2.3.
Step 5) After the prediction of each sub-layer, the final forecasting result is generated by summing all the predictions of the sub-layers. The result is compared with models including RF, feed-forward neural network (FFNN), CNN, LSTM, WPD-LSTM and WPD-CNN to evaluate the forecasting performance of the developed method.

2.1. Wavelet packet decomposition (WPD)

As a signal processing method, WPD is an efficient mathematical solution for decomposing signals into approximation and detail components with different time frequencies [26]. WPD uses low-pass and high-pass filters to decompose the signals to the components mentioned above. The approximate coefficient, obtained by applying a low-pass filter, has the low-frequency part of the signal and represents the long-term dependencies. On the other hand, the detailed coefficients gained by applying a high-pass filter include high-frequency components and depict the short-term dependencies [27]. In contrast to the wavelet decomposition (WD) process in which only the approximation

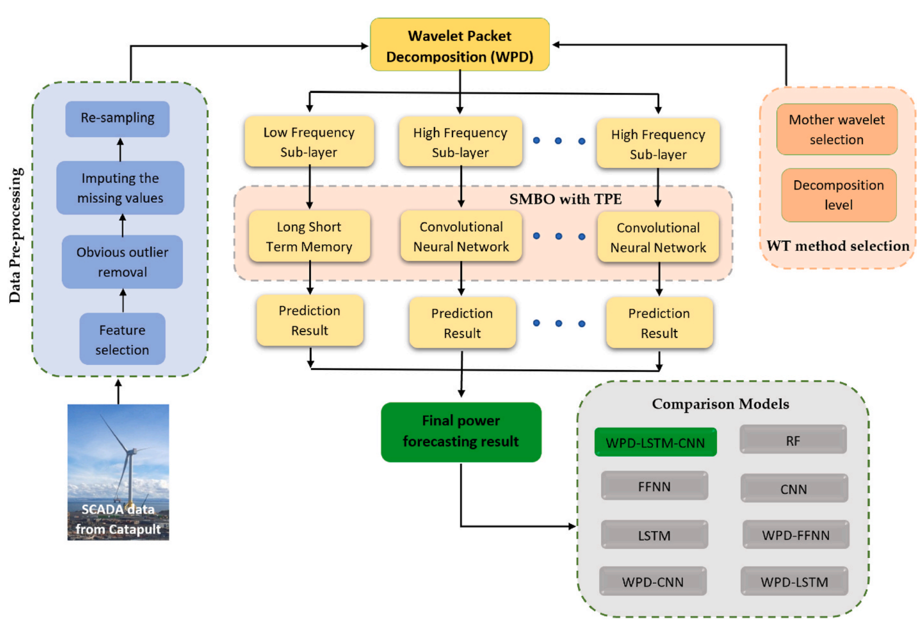

Fig. 1. Diagram of the applied methodology.

coefficients are decomposed, in WPD, the detail coefficients can also be decomposed [28]. As a result, it can contribute to higher accuracies in signal analysis than normal wavelet transforms methods. In addition, through decomposition by WPD, the high-frequency component of the signal can have a better resolution [29].

There are two types of WPDs; discrete wavelet transform and continuous wavelet transform. Continuous wavelet transform for a signal *f(t)* can be described as:

$$CWT_f(a,b) = \langle f(t), \Psi_{a,b}(t)\rangle = \int_{-\infty}^{+\infty} f(t)\Psi * ((t-b)/a) / \sqrt{a}dt \tag{1}$$

where $\Psi(t)$ denotes the selected mother wavelet function, *a* and *b* are the scale and translation coefficients, respectively, and * indicates the complex conjugate. The scale and translation coefficient in discrete wavelet transform can be explained by:

$$\begin{cases} a = 2^j \\ b = k2^j \end{cases} \tag{2}$$

where *j* and *k* are scale and translation factors, respectively, the following equations can illustrate the decomposition process of WPD:

$$\begin{cases} P_j^{2i-1}(t) = HP_{j-1}^i(t) \\ P_j^{2i}(t) = GP_{j-1}^i(t) \end{cases} \tag{3}$$

and the reconstruction process can be described as follows:

$$P_j^i(t) = H * \ P_{j+1}^{2i-1}(t) + G * \ P_{j+1}^{2i}(t) \tag{4}$$

where *t* is the time index, P_j^i represents the *i*-wavelet packet for level *j* and *H* and *G* are the low- and high-pass filters.

The performance of WPD is highly dependent on the selected mother wavelet and the chosen level of decomposition. According to the literature, the normal decomposition level is in the range of 2–4 for time series prediction models [30]. In this study, the 2-level framework of WPD is employed with the schematic diagram shown in Fig. 2.

In addition, various mother wavelets are examined due to the impact of the mother wavelet on the decomposition performance and prediction accuracy. In this study, sixteen wavelets from four widely used wavelet families in the literature (Daubechies, Haar, Sym, and Coif) were selected, and their performance in prediction improvement of forecasting models, including linear regression (LR), RF, FFNN and LSTM are assessed.

As seen in Table 2 and Fig. 3, the Daubechies wavelets of order 5 have the best performance and are therefore selected as the mother wavelet in this study.

The corresponding decomposition result of the application of WPD for the wind power time series used in this research is shown in Fig. 4. The upper graph in this Figure represents the wind power time-series before decomposition and the next four graphs show four sub-series obtained after decomposition.

In this research, to improves the decomposition which leads to more accurate predictions, the single branch reconstruction method is used. In fact, during the reconstruction of each component of the final decomposition to the original level, the values of other components of the same level were considered zero [28].

High-frequency components of wind power time series have short-term dependencies, while low-frequency components have long-term dependencies. Detection of short-term dependencies in high-frequency components is possible with fully connected layers, and calculations are performed faster than LSTM recurrent layer [31]. Therefore, this paper selects the CNN model to predict the high-frequency sublayers. On the other hand, the LSTM recurrent layer, which can better sequentially process the temporal data with long-term dependencies, is used for predicting the low-frequency sub-series.

2.2. Convolutional neural network (CNN)

CNN is an effective method of extracting hidden features by automatically creating filters [31]. Although it is more common in the field of image processing, it has shown great potential for dealing with time series, including wind speed forecasting [32], wind power prediction [11], and solar irradiance forecasting [28]. CNNs are a multilayer perceptron (MLP) version that can resist overfitting data [33]. While the full connection of neurons in each layer to neurons in other layers in MLP exposes them to overfitting, CNNs benefit from the hierarchical pattern in data and gather increasingly complex patterns using easier patterns embossed in their filters. Each convolutional layer in CNN can be represented as follows:

$$h_{ij}^k = f\left(\left(W^k * x\right)_{ij} + b_k\right) \tag{5}$$

where *f* represents the activation function and W^k represents the connected kernel weights to the *k*th feature map. In this study, the CNN models consist of two convolutional layers and a fully connected dense layer. The optimiser algorithm independently selects the channel number of each convolutional layer and the activation function for different datasets and sub-series.

2.3. Long short-term memory (LSTM)

The LSTM model with details and features explained in the [15] is employed in this study based on its great capability of processing the temporal data with long-term and short-term dependence [34]. The number of LSTM units in the hidden layer, the batch size (the number of processed samples before updating the weights), the iterations through the dataset training, the sample number in each epoch during weight updating, and the difference order that makes data stationary are some

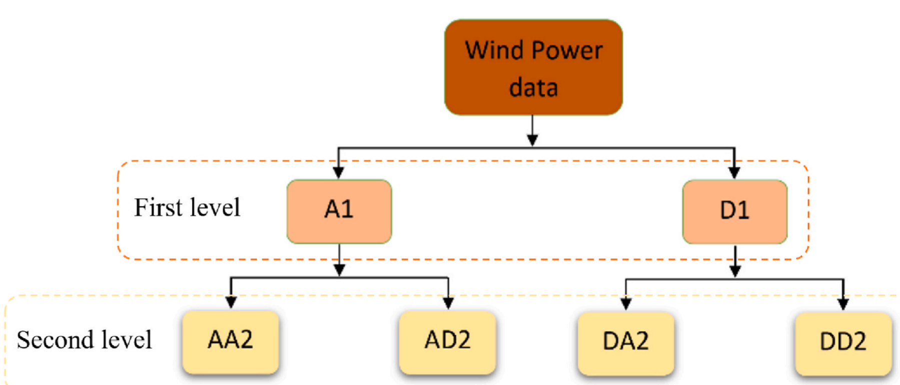

Fig. 2. Schematic diagram of WPD with two layers.

Table 2
RMSE values of wind power forecast (kW) based on the application of different mother wavelets.

Model	db1	db2	db3	db4	db5	db6	db7	db8	db9	db10	haar	sym2	sym5	sym8	coif 1	coif 5
LR	795	788	783	784	712	799	804	804	799	795	774	788	801	803	786	803
RF	799	755	735	722	687	710	714	713	711	704	722	757	730	734	729	719
FFNN	822	773	757	732	686	737	730	753	750	764	754	774	765	768	757	761
LSTM	812	764	746	722	690	723	717	733	730	728	746	765	747	751	743	740

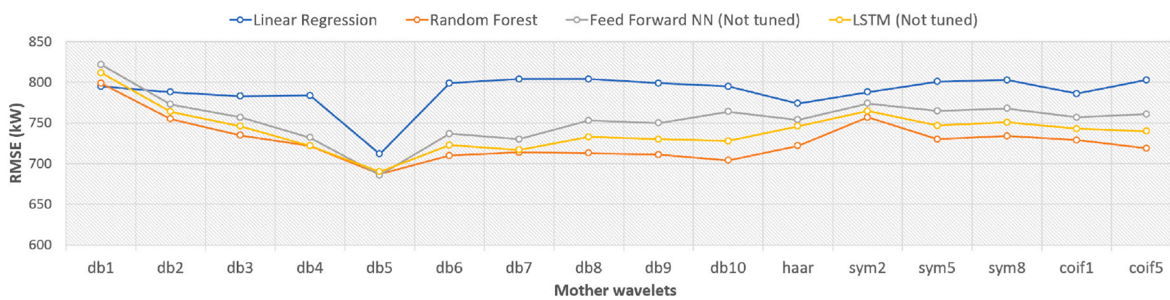

Fig. 3. Prediction performance of forecasting models based on decomposition with different mother wavelets.

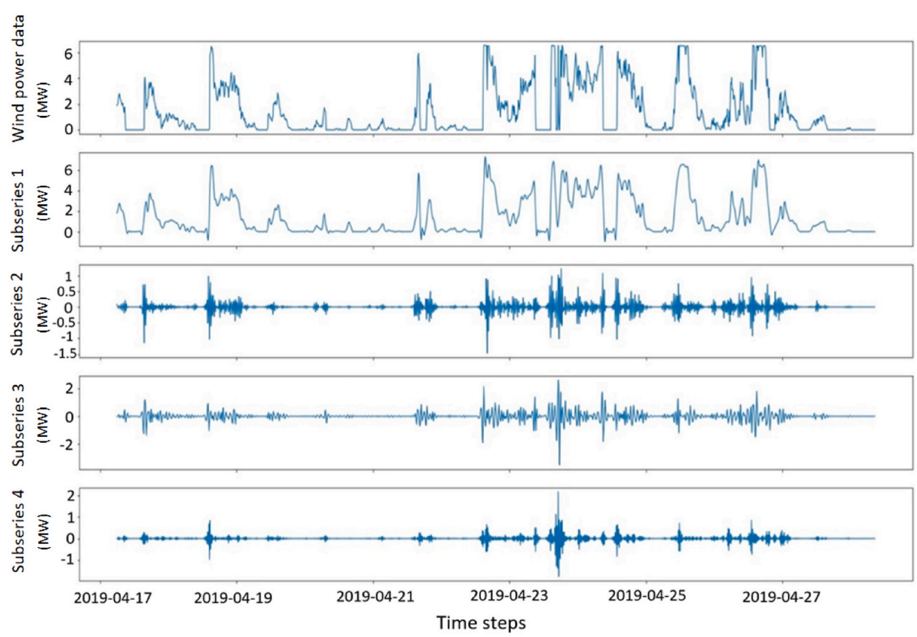

Fig. 4. Wavelet packet decomposition result of wind power time series.

of the hyper-parameters for LSTMs.

To Specify the best LSTM model for wind power forecasting, it is vital to determine the best combination of their hyper-parameters. In this way, overfitting is prevented, and the generalisation of the algorithm is improved.

2.4. Hyper-parameter optimisation

To improve both the CNN and LSTM models' prediction accuracy, hyper-parameters can be estimated through iterative trial and error. This process can be very challenging, leading to prediction errors. As a result, researchers try to find hyper-parameters through methods such as grid search or random search. The grid search process is very time consuming and needs considerable computing resources as it tries all possible combinations of hyper-parameters without considering the past evaluated hyper-parameters [35]. On the other hand, a random search algorithm looks randomly for a set of combinations rather than searching for better results.

The forecasting models in this study are tuned using the Optuna optimisation method. The open-source optimisation software Optuna [34] has a number of benefits over other optimisation frameworks. Depending on the algorithm used to choose the parameters, other optimisation techniques typically vary. Gaussian Processes are used, for instance, by GPyOpt and Spearmint [35], Random Forests are used by SMAC [36], and a Tree-structured Parzen Estimator is used by Hyperopt [37]. (TPE). Three issues with these approaches stand out. First, for large-scale experiments with numerous possible parameters, defining the parameter search space is a very challenging procedure. Furthermore, they lack a powerful pruning approach for high-performance optimisation while using constrained resources. Thirdly, they are unable to conduct extensive handling trials with a basic setup. With Optuna's define-by-run architecture, the user can dynamically create the search space. Optuna includes two strategies for finding the best hyper-parameters, a Sampling strategy for concentration on areas of hyper-parameters with better results and a Pruning strategy with constantly checking the training process to terminate combination with worse results [15]. In this study, the Tree-structured Parzen estimator algorithm is used for sampling and the median stopping rule is employed to prune trials. More details of the optimisation process of this package can be found in Ref. [36].

The first step of the optimisation algorithm is providing the search space for the dynamic generation of the hyper-parameters for each trial. The search space for the hyper-parameters of the CNN and LSTM models in this research is defined in Table 3.

2.5. Prediction performance criteria

To evaluate the performance of the developed hybrid forecasting model, four prediction performance evaluation metrics were used: the mean absolute error (MAE), the mean square error (MSE), the root mean square error (RMSE), and R-square (R^2). These metrics can be computed as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|Y_i - \widehat{Y}_i| \tag{12}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\widehat{Y}_i - Y_i)^2 \tag{13}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{Y}_i - Y_i)^2} \tag{14}$$

$$R^2 = 1 - \sum_{i=1}^{N}\frac{(Y_i - \widehat{Y}_i)^2}{(Y_i - \overline{Y})^2} \tag{15}$$

where Y_i represents the recorded offshore wind power at the time step t, $\widehat{Y}_i$ is the prediction of wind power values for the identified time step, $\overline{Y}$ is the mean power value, and N represents the number of data points. For further comparison of the forecasting performance of the proposed model with existing prediction models, the P_{MSE} as the promoting percentages of mean square error, P_{RMSE} as the promoting percentages of root mea , and P_{MAE} as the promoting percentages of mean absolute error can be calculated from Eq. (16) to Eq. (18). In these equations, index 1 is related to the existing prediction models and index 2 is related to the proposed WPD–CNN–LSTM model.

$$P_{MSE} = |(MSE_1 - MSE_2) / MSE_1| \tag{16}$$

$$P_{RMSE} = \left|\frac{RMSE_1 - RMSE_2}{RMSE_1}\right| \tag{17}$$

$$P_{MAE} = |(MAE_1 - MAE_2) / MAE_1| \tag{18}$$

Table 3
Search space for hyper-parameters of CNN and LSTM models.

LSTM Model		CNN Model	
Hyper-parameter	Search Space	Hyper-parameter	Search Space
Batch size	50, 100, 150, 200	1st convolution layer channels	12, 16, 20, 24, 28, 32, 36, 40
Epoch numbers	100, 150, 200, 250	2nd convolution layer channels	12, 16, 20, 24, 28, 32, 36, 40
LSTM units	20, 40, 60, 80, 100	Epoch numbers	100, 150, 200, 250
Neurons in dense layer	20, 40, 60, 80, 100	Activation function	Sigmoid, tanh, Relu
Activation function	Sigmoid, tanh, Relu, Softmax		

3. Case study

The Leven mouth Demonstration Turbine (LDT), an offshore wind turbine in Scotland with configuration and key parameters detailed in Refs. [15,37], provided the source SCADA data used in this investigation. The datasets were recorded for four months, from January 1, 2019 to April 30, 2019. The 574 different observations of each timestamp in this time series data include generated power, wind speed at various levels, blade pitch angle, nacelle orientation, etc. With the exception of the time stamp, wind speed, and active power, all extraneous variables were deleted to reduce the dataset's size and speed up calculations. Selection of these input features has been based on previous researches using the same wind power dataset [12,13].

An assessment of active power values revealed some obvious errors in the SCADA data, with plenty of negative values. In wind energy power prediction, negative values have no real-world applications. Shen et al. hypothesised that these numbers represent time stamps taken when the turbine's control system was in use and the blades were not rotating, which resulted in no power being produced [38]. These values were replaced with zero to keep the time continuity of the time series. In addition, to decrease the negative effect of wind turbulence on the correlation between the measured wind speed and the output power [15], the resolution of data averaged 10 min. This resolution value corresponds to the recommended average time by the international standard for power performance measurements of electricity-producing wind turbines (IEC 61400-12-1) [39]. It is also equivalent to the maximum sampling rate used to predict wind speed and wind power in the survey conducted by Hanifi et al. [1].

4. Experimental results and discussions

In order to investigate the prediction performance of the proposed method, the data processed in section 3 is divided into four equal experimental parts, each including 4200 samples, the 1st-3800th samples are used for training, and the 3801th-4200th samples are considered for testing. Fig. 5 shows these four power time series, and Table 4 provides their statistical descriptions.

For comparison, seven wind power forecasting models, including the RF, FFNN, CNN model, LSTM model, WPD-FFNN model, WPD-CNN model, WPD-LSTM model, and the proposed WPD–CNN–LSTM model were selected.

Python programming language and packages are employed to carry out all the steps of the proposed method. A PC with Intel Core™ i7–11850H 2.5 GHz CPU and 16 GB RAM (without GPU processing) is used to run the numerical experiments. To better investigate the forecasting performance of the various models, all selected models have similar parameters; for example, the selected time lag (input layer length) was set at 10 for all of them. The values of MAE, MSE, RMSE and R-square of all prediction models for the four datasets are shown in Tables 5 and 6.

Fig. 6 shows the prediction results of all forecasting models for the last day of the dataset 1, and Fig. 7 shows the same forecast for only 2 h of the last day.

As can be seen from Tables 5 and 6 and Figs. 6 and 7, when the wind power generation encounters abrupt changes, the methods that use WPD to decompose the data have a better prediction performance than the other methods. Figs. 8–11 show the forecasting results of the proposed WPD-LSTM-CNN model for datasets 1 to 4, respectively.

Based on Figs. 6–11 and Tables 5 and 6, the following conclusions can be drawn:

1) Comparing the prediction performance of the FFNN, CNN and LSTM models with WPD-FFNN, WPD-CNN and WPD-LSTM models, respectively, the significant impact of WPD on improving the prediction capability is evident;

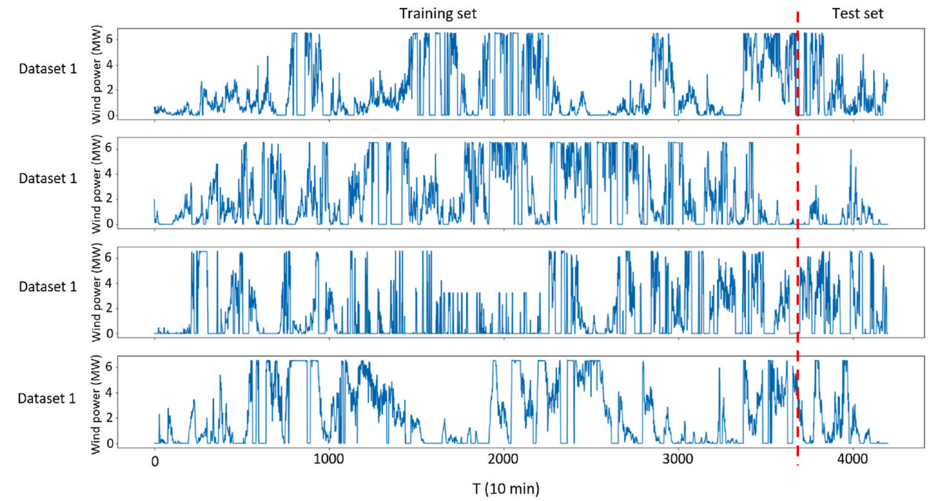

Fig. 5. Four sets of 10-min averaged wind power time series.

Table 4
Statistical descriptions of the wind power data.

Data	Mean	Min	Max	Standard Derivation
Dataset 1	1794.2	0	6566.9	2167.8
Dataset 2	2029.7	0	6567.5	2316.3
Dataset 3	1372.4	0	6569.6	2040.3
Dataset 4	2121.3	0	6551.2	2349.9

2) The developed WPD–CNN–LSTM model has the highest prediction precision among all the models which are considered here;
3) A combination of two optimised deep learning methods, CNN, and LSTM, for the prediction of different sub-series increases the forecasting accuracy compared to forecasting all sub-series with only one of them;

For further assessment of the forecasting performance of the proposed hybrid model, the P_{MSE}, P_{RMSE}, P_{MAE} of the trial tests are used to provide a clear comparative analysis between the WPD-LSTM-CNN model and other forecasting models. Table 7 provides a comparative analysis between the proposed model and other involved forecasting models for the four experimental tests, respectively.

Based on the reported promoting percentages in Table 7 and it can be recognised that:

1) The developed WPD-LSTM-CNN is the most accurate short-term forecasting model for wind power time series among all evaluated models;
2) The WPD-LSTM-CNN model outperforms all forecasting models based on the use of non-decomposed data. For example, in experimental dataset 1, the RMSE value of the WPD-LSTM-CNN model, compared to models RF, FFNN, CNN and LSTM, was reduced by 79.08%, 78.43%, 77.72% and 77.4%, respectively, and the MAE value of the proposed model for the same experimental dataset, compared to models RF, FFNN, CNN and LSTM was reduced by 77.62%, 77.41%, 76.84% and 76.49%, respectively;

Table 5
Performance comparison between WPD–CNN–LSTM and other models for datasets 1 and 2.

Comparison models	Dataset 1				Dataset 2			
	MSE	RMSE	MAE	R-square	MSE	RMSE	MAE	R-square
RF	350741.8	592.2	405.8	0.836	177039.3	420.8	190.5	0.73
FFNN	329902.1	574.3	402	0.846	140227.3	374.4	158.9	0.786
CNN	309210.7	556.1	392.1	0.856	144140.7	379.6	167.1	0.78
LSTM	300599.8	548.3	386.2	0.858	142455.1	377.4	208.1	0.783
WPD-FFNN	28240.2	168	118.5	0.987	9741.4	98.6	48.3	0.985
WPD-CNN	16890.39	129.9	99.88	0.8937	6970.04	83.5	44.66	0.891
WPD-LSTM	16844.8	129.7	95.3	0.992	6348.4	79.6	46.1	0.99
WPD–CNN–LSTM	15354.9	123.9	90.8	0.993	6336.4	79.6	40.6	0.99

Table 6
Performance comparison between WPD–CNN–LSTM and other models for the datasets 3 and 4.

Comparison models	Dataset 3				Dataset 4			
	MSE	RMSE	MAE	R-square	MSE	RMSE	MAE	R-square
RF	1303094	1141.5	636.5	0.741	287230.2	535.9	300.8	0.914
FFNN	1297934.5	1139.2	667.3	0.741	254659.1	504.6	269.8	0.923
CNN	1249351.3	1117.7	658.3	0.751	241560.7	491.5	269.1	0.928
LSTM	1181880.9	1087.1	631	0.769	235092.6	484.9	272.6	0.929
WPD-FFNN	66242.6	257.3	172.2	0.986	16465.3	128.3	72.1	0.995
WPD-CNN	44450.1	210.8	148.2	0.991	13064.04	114.3	71.39	0.8964
WPD-LSTM	46707.87	216.1	161.37	0.8919	11220.9	105.9	64.7	0.997
WPD–CNN–LSTM	42461.7	206.1	146.7	0.991	11876.4	108.9	64.9	0.996

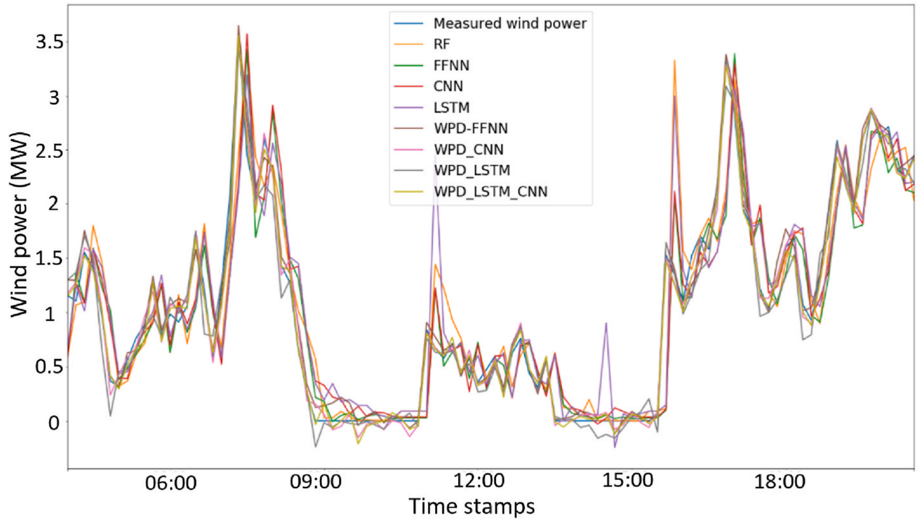

Fig. 6. Forecasting results of the involved models for dataset 1.

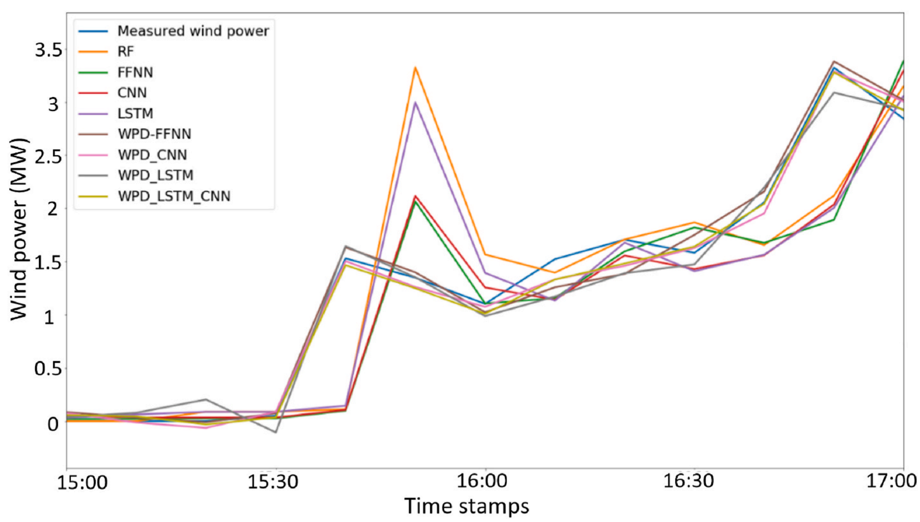

Fig. 7. Forecasting results of the involved models for 2 h of dataset 1.

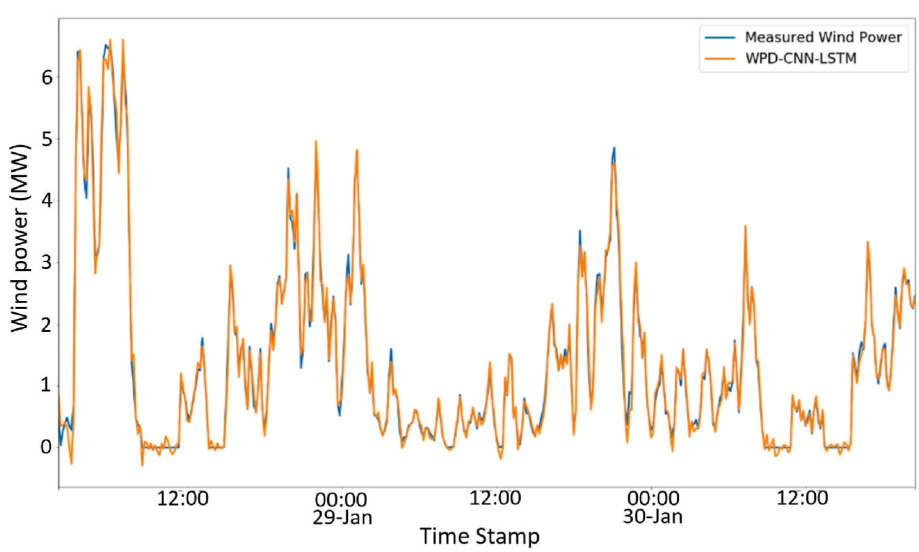

Fig. 8. Wind power forecasting with the proposed WPD-LSTM-CNN model for data set #1.

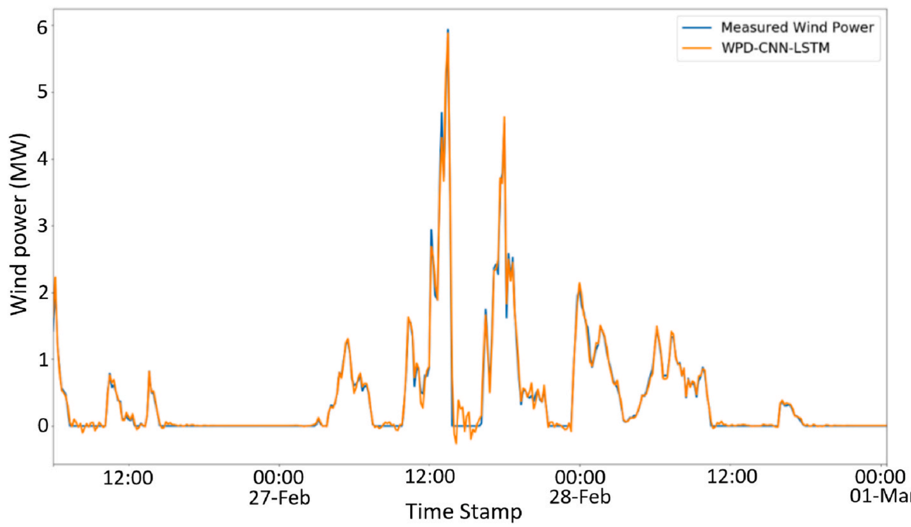

Fig. 9. Wind power forecasting with the proposed WPD-LSTM-CNN model for data set #2.

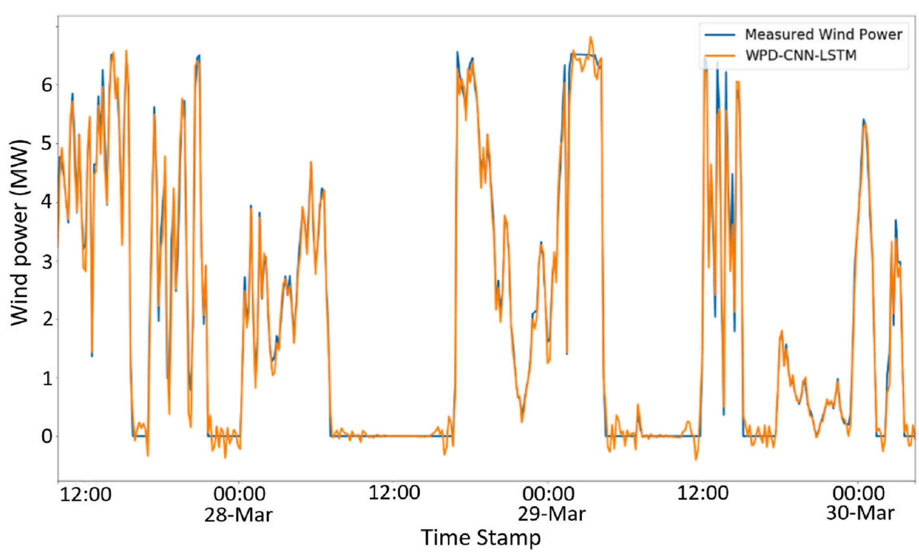

Fig. 10. Wind power forecasting with the proposed WPD-LSTM-CNN model for data set #3.

3) Comparing the proposed model with other decomposed-based models, as shown in Table 7, the WPD-LSTM-CNN model can significantly outperform the WPD-FFNN model. For example, in experimental dataset 2, the RMSE values of the developed model are reduced by 19.27%, the MSE value is reduced by 34.95%, and the MAE level is reduced by 15.94% compared to the WPD-FFNN;
4) The developed model can also outperform the WPD-CNN model. As can be seen from the evaluation criteria of the experimental dataset 3, for example, the RMSE, MSE and MAE values of the WPD-LSTM-CNN model, compared to the WPD-CNN, are reduced by 2.23%, 4.47% and 1.01%, respectively.
5) The WPD-LSTM-CNN can also outperform the WPD-LSTM model. For example, in experimental dataset 4, the RMSE, MSE and MAE values of the WPD–CNN–LSTM model, compared to the WPD-LSTM, are decreased by 2.75%, 5.52% and 0.31%, respectively.

5. Conclusions

This paper proposes a novel wind power forecasting method based on the combination of WPD, optimised LSTM and CNN models. In the developed WPD-LSTM-CNN model, first, the obvious outliers that diminish the prediction accuracy, are removed and the resolution of data averaged over 10 min in order to mitigate the influence of turbulence. After an assessment of various mother wavelets and selection of the db5 mother wavelet, resulting in the best performance, WPD is employed to decompose the pre-processed wind power time series into several sub-series with different frequencies. The appropriate decomposition of signals into several sub-series increases the stationary of data and thus makes the prediction models more efficient. Three tuned independent CNNs are employed for the prediction of the high-frequency sub-series, and one optimised LSTM model is adopted to complete the forecasting of the low-frequency sub-layer. For the optimisation of these deep learning models, the SMBO method as a formalisation of Bayesian optimisation, provided in the Optuna optimisation package, is used to reduce the dependence on computational resources.

For the prediction performance assessment of the proposed model, various forecasting models were employed, including the RF model, FFNN model, CNN model, LSTM model, WPD-FFNN model, WPD-CNN model, and WPD-LSTM model.

Based on the prediction results for four different datasets it is observed that WPD, through extracting the hidden features of the signals and reducing noise and irrelevant details can effectively improve the

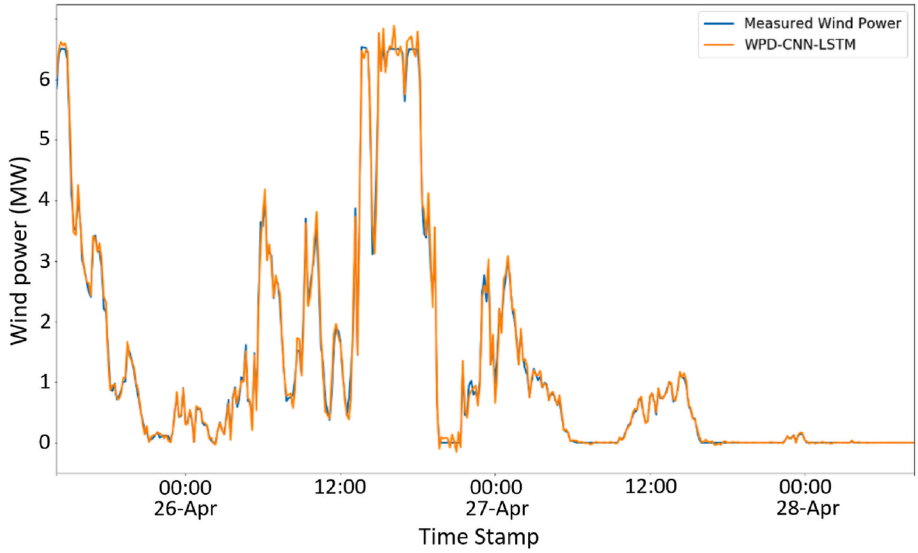

Fig. 11. Wind power forecasting with the proposed WPD-LSTM-CNN model for data set #4.

Table 7
Promoting percentages of the involved forecasting models by the WPD–CNN–LSTM model.

Promoting percentages	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Average
P_{RMSE} (%)					
RF	79.08%	81.08%	81.94%	80.24%	80.59%
FFNN	78.43%	78.74%	81.91%	79.01%	79.52%
CNN	77.72%	79.03%	81.56%	78.45%	79.19%
LSTM	77.40%	78.91%	81.04%	78.16%	78.88%
WPD-FFNN	26.25%	19.27%	19.90%	17.46%	20.72%
WPD-CNN	4.62%	4.67%	2.23%	7.35%	4.72%
WPD-LSTM	4.47%	0.09%	4.63%	2.75%	2.99%
P_{MSE} (%)					
RF	95.62%	96.42%	96.74%	96.09%	96.22%
FFNN	95.35%	95.48%	96.73%	95.59%	95.79%
CNN	95.03%	95.60%	96.60%	95.35%	95.65%
LSTM	94.89%	95.55%	96.41%	95.23%	95.52%
WPD-FFNN	45.63%	34.95%	35.90%	31.85%	37.08%
WPD-CNN	9.09%	9.09%	4.47%	14.11%	9.19%
WPD-LSTM	8.84%	0.19%	9.09%	5.52%	5.91%
P_{MAE} (%)					
RF	77.62%	78.69%	76.95%	78.49%	77.94%
FFNN	77.41%	74.45%	78.02%	76.02%	76.47%
CNN	76.84%	75.70%	77.72%	75.96%	76.55%
LSTM	76.49%	80.49%	76.75%	76.27%	77.50%
WPD-FFNN	23.38%	15.94%	14.81%	10.26%	16.10%
WPD-CNN	9.09%	9.09%	1.01%	9.38%	7.14%
WPD-LSTM	4.72%	11.93%	9.09%	0.31%	6.51%

prediction performance of the forecasting models. This improvement is more pronounced during time steps when the wind power encounters abrupt changes. Considering the four different datasets, using WPD improved the average accuracy of the FFNN, CNN and LSTM models, by 74.10%, 78.13% and 78.38%, respectively.

It is also observed that the optimised CNN and LSTM models have good performance in learning the short-term and long-term dependencies of the wind power time series. Using SMBO methods for hyper-parameter selection of these deep learning models, instead of commonly used methods such as grid and random search, increases the prediction accuracy and efficiency. Using only the CNN model increases the forecasting accuracy by 6.56% and 1.57% on average compared to the RF and FFNN models, respectively. Likewise, the application of only the LSTM model improves the prediction accuracy by an average of 8% and 3.05% compared to the RF and FFNN models, respectively.

Furthermore, the simultaneous application of the CNN and LSTM models to predict the approximation and detail components of the decomposed time series, instead of using only one of them to predict both components, was shown to improve the prediction performance. The results of the simulations have shown that this approach leads to an improvement in accuracy of up to 11.93% and 14.11%, compared to the application of only the CNN and LSTM models, respectively.

This research validated the ability of an ensemble method employing the WPD, CNN, LSTM and SMBO for short-term offshore wind power forecasting. The proposed model can potentially be adapted for onshore wind power prediction. However, some factors including geographical differences, data adaptation, and temporal and seasonal variabilities need to be considered before transition. Geographical differences between offshore and onshore wind turbines can lead to variations in wind patterns, turbulence, and other atmospheric conditions that affect generated power. In the case of data adaptation, a significant amount of onshore wind power data needs to be gathered for retraining the model. This is to ensure the required level of accuracy. The proposedforecasting method has the potential to advance wind power prediction applications. During the simulations carried out in this study, the effect of random initialisation while training the deep learning methods was observed. As a result, in future work, this issue will be assessed to provide the most robust prediction model with less sensitivity to the random initialisation. In addition, due to the lack of access to different wind power data sets, the model proposed here is limited to predicting the generated power of only one wind turbine. Future research should investigate the performance of the model for predictions based on other wind power data sets. Meanwhile, we aim to increase the forecasting horizon up to several hours, or even longer.

Data availability statement

The data presented in this study are available on request from the corresponding author. The data are not publicly available because it also forms part of an ongoing study.

CRediT authorship contribution statement

Shahram Hanifi: Conceptualization, Methodology, Software, Investigation, Validation, Formal analysis, Writing – original draft, Visualization. **Hossein Zare-Behtash:** Writing – review & editing, Methodology, Supervision. **Andrea Cammarano:** Writing – review & editing, Methodology, Supervision. **Saeid Lotfian:** Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was supported by the EPSRC Doctoral Training Partnership (EP/R513222/1). The authors thank the Offshore Renewable Energy (ORE) Catapult for provisions of the SCADA database.

Nomenclature

Latin symbols

a	Scale coefficient
b	Translation coefficient
db	Daubechies
f	Activation function
G	High pass filter
H	Low pass filter
h_t	LSTM overall output
h_{t-1}	cell state vector at time step t −1
i_t	LSTM input gate
o_t	LSTM's output gate
P_t	measured wind power at the time t
$\widehat{P}_{t+k/t}$	predicted wind power for the future time k
$\widehat{P}_{t+k/t}$	predicted wind power for the future time k
t	Time index
U_i, U_o, U_f	LSTM assigned weights
W_i, W_o, W_f	LSTM assigned weights
W^k	kernel weights
x_t	neuron input at time step t
Y_i	Measured Wind Power
$\widehat{Y}_i$	Forecasted wind power
$\overline{Y}$	Mean wind power

Greek symbols

σ_l	Activation function
σ_s	activation function
φ_t	ARIMA model coefficient
θ_t	ARIMA model coefficient
$\Psi(t)$	mother wavelet function

Abbreviation

ANN	Artificial Neural Network
ARIMA	Auto-Regressive Integrated Moving Average
CEC	Constant Error Carousel
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DBN	Deep Belief Network
DCNN	Deep convolutional neural network
DGF	Double Gaussian Function
EEMD	ensemble empirical mode decomposition
FFNN	Feed Forward Neural Network
GRU	Gated Recurrent Unit
IMFs	intrinsic mode functions
LDT	Levenmouth Demonstration Turbine
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Square Error
NN	Neural Network
NWP	Numerical Weather Prediction
ORE	Offshore Renewable Energy
PRMSE	promoting percentages of root mean square error

PMAE promoting percentages of mean absolute error
PMSE promoting percentages of mean square error
RBF Radial Basis Function
RMSE Root Mean Square Error
RF Random Forest
RNN Recurrent Neural Network
R2 R-square
SCADA Supervisory Control and Data Acquisition
SMBO Sequential Model-Based Optimisation
TPE Tree Parzen Estimator
VMD variational model decomposition
WPD Wavelet Packet Decomposition
WPT Wavelet Packet Transform

References

[1] S. Hanifi, X. Liu, Z. Lin, S. Lotfian, A critical review of wind power forecasting methods-past, Present Future Energ. 13 (2020) 1–24.
[2] K.T. Kit, Challenge of Net-Zero Carbon Construction and Measurement of Energy Consumption and Carbon Emission Reduction to Climate Change , Economy and Job Growths in Hong Kong and Australia, 2022.
[3] B. Yang, L. Zhong, J. Wang, H. Shu, X. Zhang, T. Yu, L. Sun, State-of-the-art one-stop handbook on wind forecasting technologies: an overview of classifications, methodologies, and analysis, J. Clean. Prod. 283 (2021), 124628.
[4] Y. Zhao, L. Ye, Z. Li, X. Song, Y. Lang, J. Su, A novel bidirectional mechanism based on time series model for wind power forecasting, Appl. Energy 177 (2016) 793–803.
[5] C. Yu, Y. Li, H. Xiang, M. Zhang, Data mining-assisted short-term wind speed forecasting by wavelet packet decomposition and Elman neural network, J. Wind Eng. Ind. Aerod. 175 (2018) 136–143.
[6] A. Lahouar, J. Ben Hadj Slama, Hour-ahead wind power forecast based on random forests, Renew. Energy 109 (2017) 529–541.
[7] S. Velázquez, J.A. Carta, J.M. Matías, Influence of the input layer signals of ANNs on wind power estimation for a target site: a case study, Renew. Sustain. Energy Rev. 15 (2011) 1556–1566.
[8] J. Liu, X. Wang, Y. Lu, A novel hybrid methodology for short-term wind power forecasting based on adaptive neuro-fuzzy inference system, Renew. Energy 103 (2017) 620–629.
[9] J. Wang, W. Yang, P. Du, Y. Li, Research and application of a hybrid forecasting framework based on multi-objective optimization for electrical power system, Energy 148 (2018) 59–78.
[10] J. Zhang, J. Yan, D. Infield, Y. Liu, Lien F. sang, Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model, Appl. Energy 241 (2019) 229–244.
[11] Y.Y. Hong, C.L.P.P. Rioflorido, A hybrid deep learning-based neural network for 24-h ahead wind power forecasting, Appl. Energy 250 (2019) 530–539.
[12] P. Du, J. Wang, W. Yang, T. Niu, A novel hybrid model for short-term wind power forecasting, Appl. Soft Comput. J. 80 (2019) 93–106.
[13] F. Shahid, A. Zameer, M. Muneeb, A novel genetic LSTM model for wind power forecast, Energy 223 (2021), 120069.
[14] A. Kisvari, Z. Lin, X. Liu, Wind power forecasting – a data-driven method along with gated recurrent neural network, Renew. Energy 163 (2021) 1895–1909.
[15] S. Hanifi, S. Lotfian, H. Zare-behtash, A. Cammarano, Offshore wind power forecasting—a New hyperparameter optimisation algorithm for deep learning models, Energies 15 (2022) 6919.
[16] S. Seyedzadeh, F. Pour Rahimian, P. Rastogi, I. Glesk, Tuning machine learning models for prediction of building energy loads, Sustain. Cities Soc. 47 (2019), 101484.
[17] J. Bergstra, R. Bardenet, Y. Bengio, B. Kégl, Algorithms for hyper-parameter optimization Adv, in: Neural Inf. Process. Syst. 24 25th Annu. Conf. Neural Inf. Process. Syst. 2011, NIPS, 2011, pp. 1–9, 2011.
[18] H. Liu, C. Chen, Data processing strategies in wind energy forecasting models and applications: a comprehensive review, Appl. Energy 249 (2019) 392–408.
[19] Y. Su, J. Yu, M. Tan, Z. Wu, Z. Xiao, J. Hu, A LSTM based wind power forecasting method considering wind frequency components and the wind turbine states, in: *2019 22nd Int. Conf. Electr. Mach. Syst. ICEMS 2019*, 2019, pp. 19–24.
[20] X.R. Zu, R.X. Song, Short-term wind power prediction method based on wavelet packet decomposition and improved GRU, J. Phys. Conf. Ser. 1087 (2018).
[21] S. Mujeeb, T.A. Alghamdi, S. Ullah, A. Fatima, N. Javaid, T. Saba, Exploiting deep learning for wind power forecasting based on big data analytics, Appl. Sci. 9 (2019).
[22] R. Azimi, M. Ghofrani, M. Ghayekhloo, A hybrid wind power forecasting model based on data mining and wavelets analysis *Energy Convers*, OR Manag. 127 (2016) 208–225.
[23] X. Shi, X. Lei, Q. Huang, S. Huang, K. Ren, Y. Hu, Hourly day-ahead wind power prediction using the hybrid model of variational model decomposition and long short-term memory, Energies 11 (2018) 1–20.
[24] H. Liu, Mi X. wei, Li Y. fei, Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network, Energy Convers. Manag. 156 (2018) 498–514.
[25] Z. Ma, H. Chen, J. Wang, X. Yang, R. Yan, J. Jia, W. Xu, Application of hybrid model based on double decomposition, error correction and deep learning in short-term wind speed prediction, Energy Convers. Manag. 205 (2020), 112345.
[26] P. Li, K. Zhou, X. Lu, S. Yang, A hybrid deep learning model for short-term PV power forecasting, Appl. Energy 259 (2020), 114216.
[27] A. Gupta, A. Kumar, K. Boopathi, Intraday wind power forecasting employing feedback mechanism, Elec. Power Syst. Res. 201 (2021), 107518.
[28] X. Huang, Q. Li, Y. Tai, Z. Chen, J. Zhang, J. Shi, B. Gao, W. Liu, Hybrid deep neural model for hourly solar irradiance forecasting, Renew. Energy 171 (2021) 1041–1060.
[29] J. Wang, J. Wang, A New hybrid forecasting model based on SW-LSTM and wavelet packet decomposition, A Case Study of Oil Futures Prices 2021 (2021).
[30] N. Amjady, F. Keynia, Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm, Energy 34 (2009) 46–57.
[31] S. Oehmcke, O. Zielinski, O. Kramer, Input quality aware convolutional LSTM networks for virtual marine sensors, Neurocomputing 275 (2018) 2603. –15.
[32] H. Liu, X. Mi, Y. Li, Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM, Energy Convers. Manag. 159 (2018) 54–64.
[33] S. Seyedzadeh, F.P. Rahimian, Data-Driven Modelling of Non-domestic Buildings Energy Performance: Supporting Building Retrofit Planning, Springer, 2021.
[34] Y. Li, H. Wu, H. Liu, Multi-step wind speed forecasting using EWT decomposition, LSTM principal computing, RELM subordinate computing and IEWT reconstruction, Energy Convers. Manag. 167 (2018) 203–219.
[35] S. Seyedzadeh, F. Pour Rahimian, S. Oliver, S. Rodriguez, I. Glesk, Machine learning modelling for predicting non-domestic buildings energy performance: a model to support deep energy retrofit decision-making, Appl. Energy 279 (2020), 115908.
[36] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: a next-generation hyperparameter optimization framework, in: Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min, 2019, pp. 2623–2631.
[37] N. Dethlefs, Deep Learning with Knowledge Transfer for Explainable Anomaly Prediction in Wind Turbines, 2020, pp. 1693–1710.
[38] X. Shen, X. Fu, C. Zhou, A combined algorithm for cleaning abnormal data of wind turbine power curve based on change point grouping algorithm and quartile algorithm, IEEE Trans. Sustain. Energy 10 (2019) 46–54.
[39] M. Jafarian, A.M. Ranjbar, Fuzzy modeling techniques and artificial neural networks to estimate annual energy output of a wind turbine, Renew. Energy 35 (2010) 2008–2014.