

## Research paper

**CNN–LSTM–AM: A power prediction model for offshore wind turbines**Yu Sun<sup>a</sup>, Qibo Zhou<sup>b</sup>, Li Sun<sup>b</sup>, Liping Sun<sup>a</sup>, Jichuan Kang<sup>a,\*</sup>, He Li<sup>c,d,\*\*</sup><sup>a</sup> College of Shipbuilding Engineering, Harbin Engineering University, Harbin, Heilongjiang, 150001, PR China<sup>b</sup> Marine Design and Research Institute of China, Shanghai, 200001, PR China<sup>c</sup> School of Engineering, Liverpool John Moores University, Liverpool, 3 Byrom Street, L3 3AF, UK<sup>d</sup> Centre for Marine Technology and Ocean Engineering (CENTEC), Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

## ARTICLE INFO

**Keywords:**  
 Offshore wind turbine  
 Power prediction  
 Multi-sensor fusion  
 Deep learning  
 Sensor sensitivity

## ABSTRACT

This study introduces a power forecasting model, the convolutional neural network (CNN)-long short-term memory (LSTM)-attention mechanism (AM) algorithm (CNN–LSTM–AM), designed to predict the power of offshore wind turbines based on data collected by a SCADA system. The model employs a timestep parameterisation approach for offshore wind turbine prediction, facilitating automatic partitioning of the training dataset and simplifying the training process. A CNN–LSTM–AM network was presented to predict the power of offshore wind turbines using signals from multiple sensors. A variable-control comparison was conducted to complete the sensitivity analysis of the sensors, which determined the most suitable sensor group for power prediction. The model achieved a maximum improvement of 13.77% in power prediction compared to existing deep learning algorithms. The results indicate that the hub and rear-end temperatures of the high-speed shaft of the gearboxes are crucial for offshore wind power prediction. Overall, the findings of this study contribute to the operation and maintenance of offshore wind turbines and the management of offshore wind farms.

**1. Introduction**

Wind energy is anticipated to produce one-fifth of the world's electricity by 2030, with projections indicating a two-third increase by 2050 (Council GWE, 2023). In contrast to its onshore counterpart, offshore wind offers several advantages, including reduced environmental impact (Li et al., 2021a), more adaptable wind farm configurations (Kang et al., 2019), superior wind profiles (Tawn and Browell, 2022), and fewer blade size limitations (Foley et al., 2012).

However, the expansion of offshore wind power has encountered several challenges related to environmental, power grid, and market factors. Specifically, (i) the intermittent, volatile, and unpredictable nature of wind power impacts grid dispatching and operation (Qian et al., 2019); (ii) the limited adjustability of power grid resources influences the consumption of electricity, resulting in wind curtailment (McDonagh et al., 2020; Bird et al., 2016); and (iii) fluctuations in power generation hinder offshore wind turbine (OWT) operators from devising viable power production and maintenance plans to satisfy market demand (Li et al., 2022a), which in turn escalates operating costs and diminishes market competitiveness.

In light of this, accurately predicting wind power generation is of paramount importance. A mere 10% enhancement in prediction accuracy translates to an approximately 30% improvement in wind power generation capacity (Ackermann and Söder, 2000). To reduce operational costs and enhance OWT safety, precise power prediction is imperative to anticipate future generations and make informed decisions in advance (Costa et al., 2008). Various methods for offshore wind power prediction include physical, statistical, intelligent, and hybrid models (Hanifi et al., 2020). Physical-based models, such as Numerical Weather Prediction (NWP) (Wang et al., 2019; Donadio et al., 2021) and Weather Research and Forecasting (WRF) (Qureshi et al., 2017), require meteorological information and substantial computational resources (Landberg et al., 2003). However, statistical-based models predicting univariate time series (Jung and Broadwater, 2014; Raza and Khosravi, 2015; Bartholomew, 1971), such as the Autoregressive Integrated Moving Average (ARIMA), are easily formulated and well-suited for short-term wind power prediction because of their low computation time cost (Yatiyana et al., 2017; Jung and Kwon, 2013). However, they do not consider the physical condition of the wind turbines and therefore may not accurately reflect the long-term

\* Corresponding author.

\*\* Corresponding author.

E-mail addresses: [kangjichuan@hrbeu.edu.cn](mailto:kangjichuan@hrbeu.edu.cn) (J. Kang), [he.li@centec.tecnico.ulisboa.pt](mailto:he.li@centec.tecnico.ulisboa.pt) (H. Li).

performance of the devices.

The advent of deep learning has introduced novel solutions for capturing nonlinear features, known as intelligent models (Manero et al., 2018; Nezhad et al., 2024). For instance, Jyothi and Rao (2016) introduced an adaptive wavelet neural network (AWNN) for short-term wind power prediction. Kisvari et al. (2021) presented a model based on a gated recurrent unit (GRU) and long short-term memory (LSTM) to construct optimised predictive models. Lin et al. (Lin and Liu, 2020) developed a deep learning model for offshore wind-power forecasting using supervisory control and data acquisition (SCADA) data. Yu et al. (2020) conducted feature extraction and wind energy prediction using the proposed superposition graph neural network method. Addressing abnormalities, multi-modality, and uncertainty in operational data, Xin et al. (Xing and He, 2023) incorporated density-based spatial clustering of applications with noise into the forecasting model, resulting in improved accuracy in power prediction for wind turbines.

The power prediction models mentioned above are typically based on data from signal sensors. However, with the expansion of the data volume, the prediction performance of these models is limited. Therefore, researchers adopted hybrid models to address this issue. For example, Neshat et al. (2021) adopted a hybrid approach combining deep learning and evolutionary methods along with a quaternion convolutional neural network (CNN) (Neshat et al., 2022) to enhance the accuracy of wind speed prediction. Zhang et al. (2022) addressed various components of the power time-series prediction problem by utilising the discrete wavelet transform and seasonal autoregressive integrated moving average. Heydari et al. (2021) proposed a hybrid forecasting model comprising empirical mode decomposition and deep learning algorithms. Hu et al. (2021) proposed a hybrid wind power forecasting method that integrates corrected numerical weather predictions and spatial correlations into a Gaussian process. Hanifi et al. (2023) proposed a combined model based on wavelet packet decomposition, an optimised LSTM, and a CNN to predict the power of wind turbines.

Numerous methods have been proposed for predicting the power of wind turbines using meteorological data (temperature, wind speed, pressure, etc.), operational data (motor speed, pitch angle, wind power, etc.), or both. However, the meteorological sensors of wind turbines, such as ultrasonic anemometers, have been proven to fail (more than 15 times per year according to the LGS-Offshore dataset) and consequently restrict the performance of power prediction models (Li and Guedes Soares, 2022; Li et al., 2021b; Sun et al., 2023). Hence, the sector is seeking new power-prediction models that do not require meteorological sensors to secure a prediction system and provide robust results. Meanwhile, the understanding of the relationships between sensors and the power of wind turbines is limited. Understanding the contribution of each sensor to the power prediction performance of wind turbines is crucial.

Motivated by the above, a multi-sensor fusion model for wind power forecasting without meteorological data is presented. Furthermore, a sensitivity analysis of multi-sensors is carried out. The novel contributions of this study are as follows.

- (i) A CNN-LSTM-AM methodology was proposed to accurately predict the power of OWTs by integrating sensor data from multiple components.
- (ii) The timestep of the power prediction of OWTs was parameterised to update the prediction results with time.
- (iii) The sensitivity of the sensors was analysed to reflect their impact on power predictions.

The remainder of this paper is organised as follows: The methodologies are presented in Sections 2 and 3. The results, comparisons, and discussion are presented in Section 4. Finally, the conclusions are presented in Section 5.

## 2. Methodology

This study constructs a power prediction model for OWTs based on the CNN-LSTM-AM algorithm, in which the CNN extracts features from sensor data, the LSTM processes the signal with temporal features, and the attention mechanism (AM) captures critical information in a temporal sequence.

### 2.1. Convolutional neural network (CNN)

CNNs are composed of convolutional, activation, pooling, fully connected, and output layers (LeCun et al., 1998). A conventional convolutional network structure is illustrated in Fig. 1 (Zare and Ayati, 2021).

#### (1) Convolutional Layers

The convolutional layer extracts information from the input data. The feature extraction inside the convolution kernel is represented by

$$C_j^n = \sigma(w_j^n \otimes C_i^{n-1} + b_j^n), \quad (1)$$

where  $C_j^n$  is the  $j$  th feature of the  $n$  th convolutional layer;  $\sigma(\cdot)$  is the non-linear activation function;  $\otimes$  denotes the convolution calculation;  $w_j^n$  indicates the  $j$  th weight matrix of the  $n$  th convolutional layer;  $C_i^{n-1}$  represents the  $i$  th output of the  $(n-1)$  th convolutional layer; and  $b_j^n$  is the  $j$  th bias vector of the  $n$  th convolutional layer.

#### (2) Pooling layer:

The pooling layer samples the exports from the convolutional layer. In this study, the max-pooling layer is used to connect the convolutional layers for downsampling, which is represented by

$$P_j^n = \max(C_j^{n-1}, s), \quad (2)$$

where  $P_j^n$  is the output of the pooling layer,  $\max(\cdot)$  is the down-sampling function of the maximum value,  $C_j^{n-1}$  is the feature vector of the convolutional layer, and  $s$  is the pooling size.

#### (3) Fully connected layers

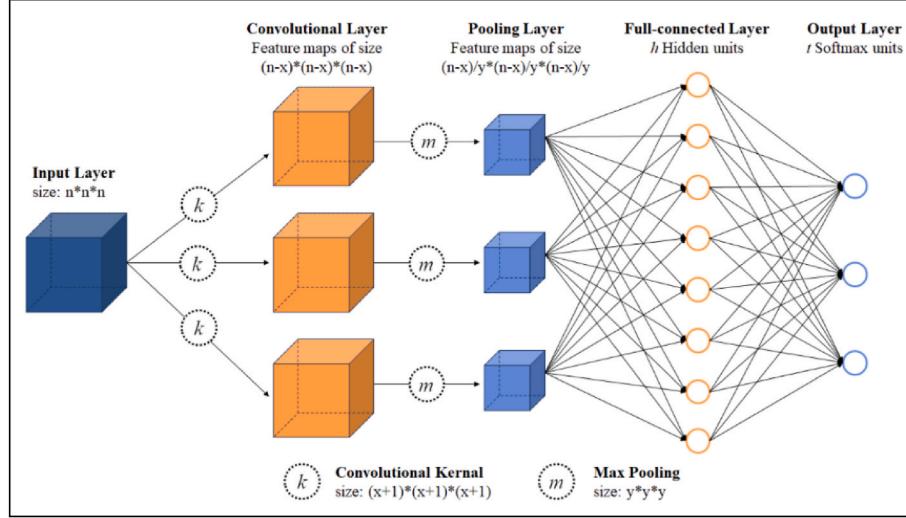
The fully connected layer achieves feature combination and transformation by connecting all neurones in the previous layer with all neurones in the current layer.

$$y^{feature} = \sigma(w_f \cdot x + b_f), \quad (3)$$

where  $y^{feature}$  indicates the feature vector,  $w_f$  denotes the weight matrix,  $b_f$  is the bias vector, and  $x$  is the input vector.  $\sigma(\cdot)$  is the non-linear activation function.

#### (4) Output layer

The output layer transmits the internal calculation results of the



**Fig. 1.** Schematic of convolutional neural networks (CNNs).

neural network to the external environment.

$$\{w, b\}^* = \operatorname{argmin}_{\{w, b\}} \frac{1}{m} \sum_{i=1}^m L(y_{true}, y_{estimated}), \quad (4)$$

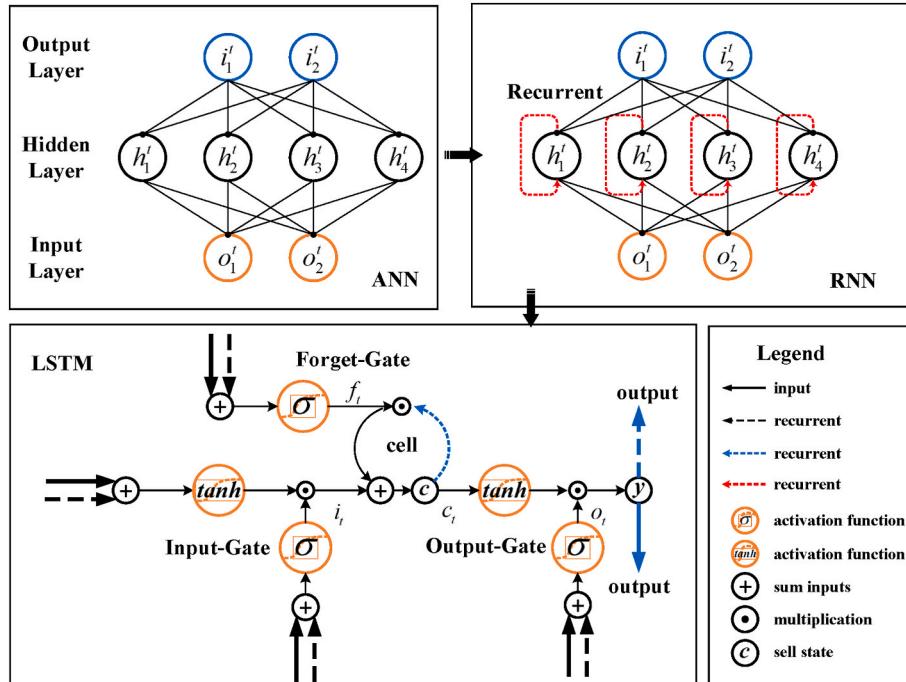
where  $m$  is the number of labelled datasets,  $L(\cdot)$  represents the loss function, and  $y_{true}$  and  $y_{estimated}$  represent the label value and estimated output of the CNN, respectively.  $\{w, b\}^*$  refers to the fine-tuned parameter weight vector  $w$  and bias  $b$ , which are obtained by minimising the loss function  $L(\cdot)$ .

The spatial features were achieved through the source data using a CNN, and the processed data were treated as inputs into the LSTM based on AM for time-series forecasting.

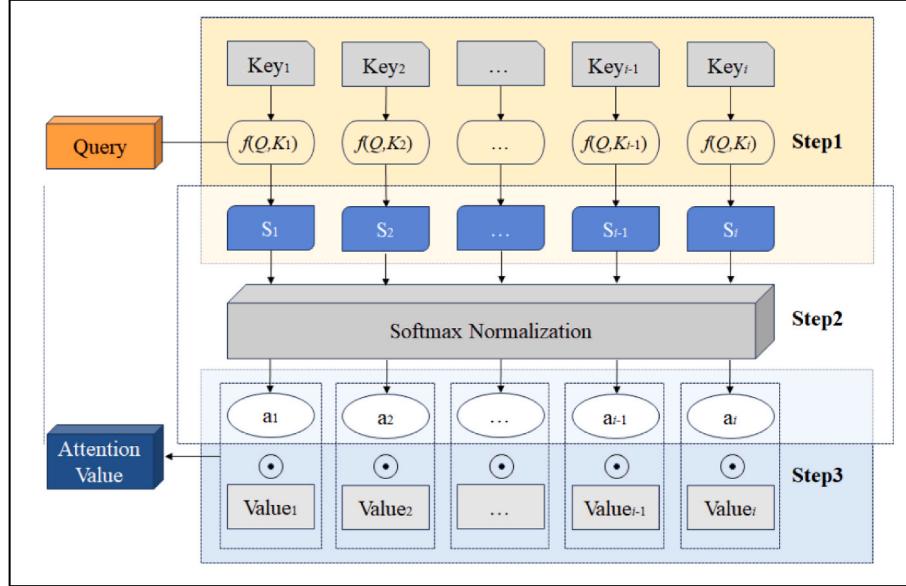
## 2.2. Long short-term memory (LSTM)

The LSTM was initially proposed by Hochreiter and Schmidhuber in 1997 (Hochreiter and Jürgen, 1997). It evolved from the recurrent neural network (RNN) architecture and developed multiple variations (Sun et al., 2022). Its superiority in classifying, processing, and predicting time series allows LSTM to outperform alternative RNNs and other time-sequenced learning methods when handling long-duration data (Graves and Schmidhuber, 2005; Greff et al., 2015). Schematics of the artificial neural network (ANN), RNN, and LSTM architectures are shown in Fig. 2.

Given the input  $x = \{x_1, x_2, \dots, x_n\}$  and the recurrent input  $h$  of the LSTM, the numerical update of the  $c$  state is performed under the control



**Fig. 2.** Schematic of the artificial neural network (ANN), recurrent neural network (RNN), and (long short-term memory) LSTM architectures.



**Fig. 3.** Schematic of the attention mechanism (AM).

of the gate. The forget gate  $f_t$  determines how much of the cell state  $c_{t-1}$  at the previous moment remains in the cell state  $c_t$  at the current moment. The output value is:

$$f_t = \text{sigmoid}(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (5)$$

where sigmoid is the activation function, which can be described in Eq. (6).  $W_f$  is the forgetting weight matrix, and  $b_f$  is the forgetting bias.

$$\text{sigmoid}(z) = \frac{1}{1 + e^{-z}}, \quad (6)$$

where  $i_t$  is an input gate that determines the number of initial cell states, and the  $\tilde{c}_t$  corresponding to the current input  $x_t$  is saved to the cell state  $c_t$  at the current moment. The output value of the gate is computed as

$$i_t = \text{sigmoid}(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (7)$$

where  $W_i$  is the input weight matrix and  $b_i$  is the input bias. The current cell state is updated as follows:

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_{t-1}, \quad (8)$$

where  $o_t$  denotes the output gate that controls the number of cell states  $c_t$ . The value of the output gate is calculated as

$$o_t = \text{sigmoid}(W_o \cdot [h_{t-1}, x_t] + b_o). \quad (9)$$

The value of the recurrent input and  $y$  is

$$y_t = o_t * \tanh(C_t), \quad (10)$$

where  $\tanh$  is the activation function, which can be described by Eq. (11).

$$\tanh(z) = (e^z - e^{-z}) / (e^z + e^{-z}) \quad (11)$$

### 2.3. Attention mechanism (AM)

The AM can be described as the mapping of a query and a set of key-value pairs to an output, where the query, key, value, and output are vectors (Mnih et al., 2014; Vaswani et al., 2017). The compatibility function, calculated in the form of a weighted sum, is the output of the weight assigned to each value (Lu et al., 2021; Xiang et al., 2021). As shown in Fig. 3, the calculation process for the AM is generally divided into three steps.

**Step #1.** Given that the information is composed of some columns of {Key, Value}, the similarity or correlation between the query (output feature) and key (input feature) is calculated as

$$s_t = f(Q, K_t) = \tanh(W_h h_t + b_h), \quad (12)$$

where  $W_h$  is the weight of the AM,  $b_h$  is the bias of the AM, and  $h_t$  is the input vector.

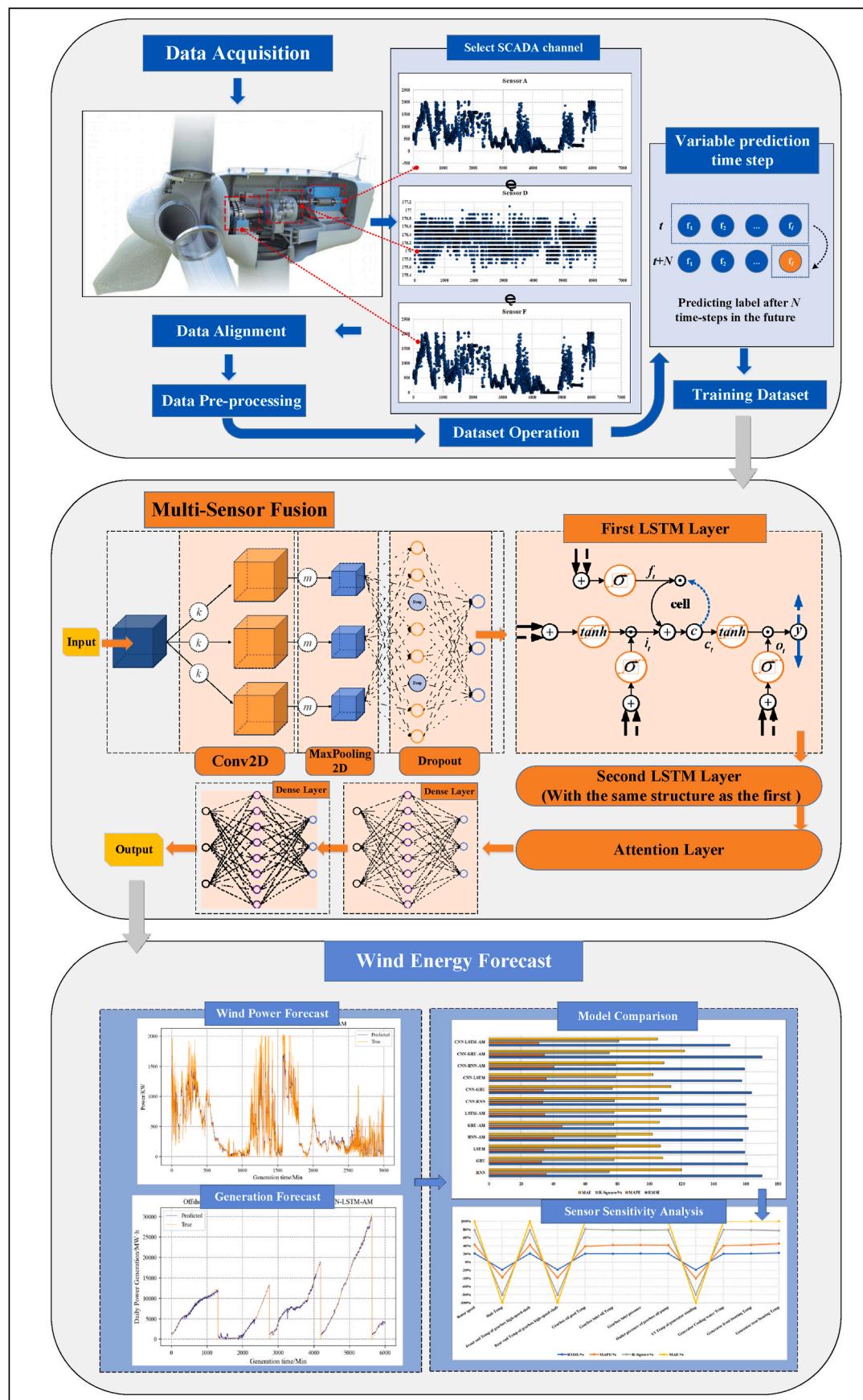
**Step #2.** The score of the first stage is normalised, and the softmax function is used to convert the attention score, which can be described as

$$a_t = \text{soft max}(s_t^T v) = \left( \exp(s_t^T v) \right) / \left( \sum_t \exp(s_t^T v) \right), \quad (13)$$

where  $v$  is the attention value.

**Step #3.** Weights are computed to obtain the final attention using

$$s = \sum_t a_t h_t. \quad (14)$$



**Fig. 4.** The overall framework for power prediction of OWTs.

**Table 1**  
CNN–LSTM–AM algorithm.

Algorithm. The CNN–LSTM–AM.	
1	Input: the prediction timestep $N$ , the training set;
2	Data Preprocessing: data recombination, data normalisation;
3	Build and initialise: CNN–LSTM–AM;
4	For epoch = 1, 2, ..., epoch do:
5	# Train the CNN–LSTM–AM
6	Data processing by CNN layer;
7	Data processing by LSTM layer;
8	Data processing by AM layer;
9	Data processing by Dense;
10	Obtaining the output of CNN–LSTM–AM;
11	# Calculate loss and gradient
12	Calculate the loss function, Calculate gradient;
13	# Adam algorithm is applied to optimise parameters;
14	End
15	Data Postprocessing: data denormalisation;
16	Output: Power prediction in N minutes

**Table 2**  
Training configuration parameters of CNN–LSTM–AM.

Parameters	Value
Iteration number	50
Batch size	512
Loss function	Mean-square Error
Optimiser	Adam

**Table 3**  
Architecture parameters of CNN–LSTM–AM.

Layer	Hyperparameters
Convolutional layer	filters = 32, kernel_size = 3, activation = ReLU
Maximum pooling layer	pool_size = 2
LSTM layer 1	10, return_sequences = True
LSTM layer 2	20, return_sequences = True
Attention	Attention(10)
Dense1	10, activation = ReLU
Dense2	1

Filter: The number of convolutional kernels; kernel\_size: The size of the convolutional kernel; activation: the activation function.

#### 2.4. The proposed CNN–LSTM–AM model

The framework and proposed power prediction model for OWTs are illustrated in Fig. 4. The following aspects were clarified: (i) Data acquisition. SCADA channels, encompassing 12 sensor types from OWTs and 13 sensor types from onshore wind turbines, were employed. (ii) Multi-sensor fusion. A CNN was used to extract spatial features from multi-sensor data. The first and second layers of the LSTM were responsible for extracting temporal features. The AM was employed to capture dependencies within long time sequences. (iii) Power prediction. The trained CNN–LSTM–AM model was used to predict the power of OWTs according to indicators, including the root-mean-square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), and R-square. The CNN–LSTM–AM algorithm is presented in Table 1.

The parameters used to train the CNN–LSTM–AM network are listed in Table 2. The mean-square error was employed as a loss function, and

**Table 4**  
Sensors of wind turbines (onshore and offshore).

Offshore		Onshore	
Components	Sensors	Components	Sensors
Rotor	Rotor speed	Main bearing	Front bearing Temp
Hub	Hub Temp		Rear bearing Temp
Gearbox	Front end Temp of HSS	Shaft brake	Hydraulic pressure
	Rear end Temp of HSS	Generator	Generator speed
	Oil pool Temp		Max Temp of GW
	Inlet oil Temp		U1 Temp of GW
	Inlet pressure		Generator Cooling water Temp
		Out pressure of GOP	Generator front bearing Temp
Generator	U1 Temp of GW		Generator rear bearing Temp
		Cooling water Temp	Generator slip ring chamber Temp
		Front bearing Temp	Generator torque
		Rear bearing Temp	Temp inside the converter
			DC voltage inside the converter

HSS: High-speed shaft; GW: Generator winding; GOP: Gearbox oil pump; Temp: Temperature.

Adam served as an optimiser to optimise the network's weights. The architectural parameters of the model are listed in Table 3.

To evaluate the performance of the proposed model, several metrics, including the RMSE (Eq. (15)), MAPE (Eq. (16)), MAE (Eq. (17)), and R-square ( $R^2$ , Eq. (18)) were computed. To be specific:

Knowing the true value  $y_i$  and the predicted value  $\hat{y}_i$  in  $n$  samples, the above criteria can be calculated as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (15)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (16)$$

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (17)$$

$$R^2 = 1 - \sum_{i=1}^n (y_i - \hat{y}_i)^2 / \left( \sum_{i=1}^n (y_i - \bar{y})^2 \right) \quad (18)$$

#### 3. Results, comparisons, and discussions

This section presents an evaluation of the CNN–LSTM–AM model. The sensitivities of the sensors used for wind energy prediction were analysed. The performance of the forecasting model is discussed by comparing it with 11 deep learning methods.

##### 3.1. Data source

In this study, SCADA data from seven components, namely, the rotor, hub, main bearing, shaft brake, gearbox, generator, and converter, were

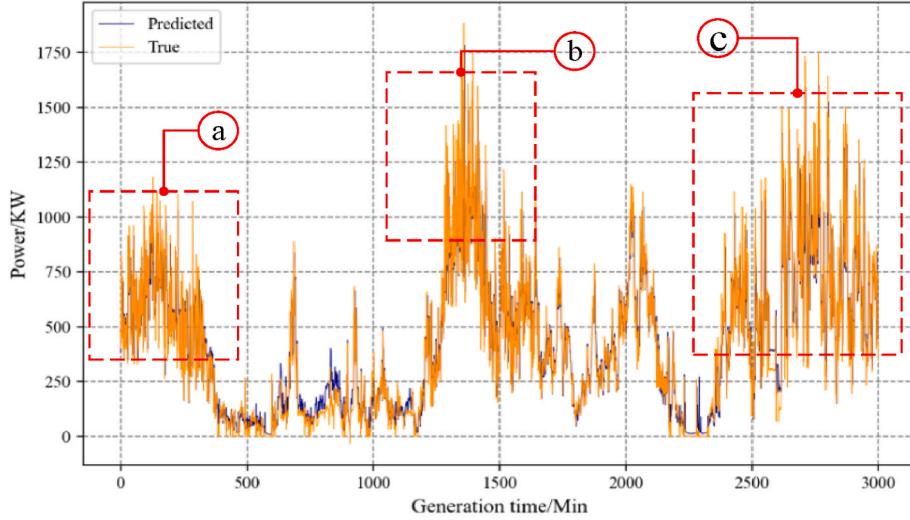


Fig. 5. Power prediction of WT1.

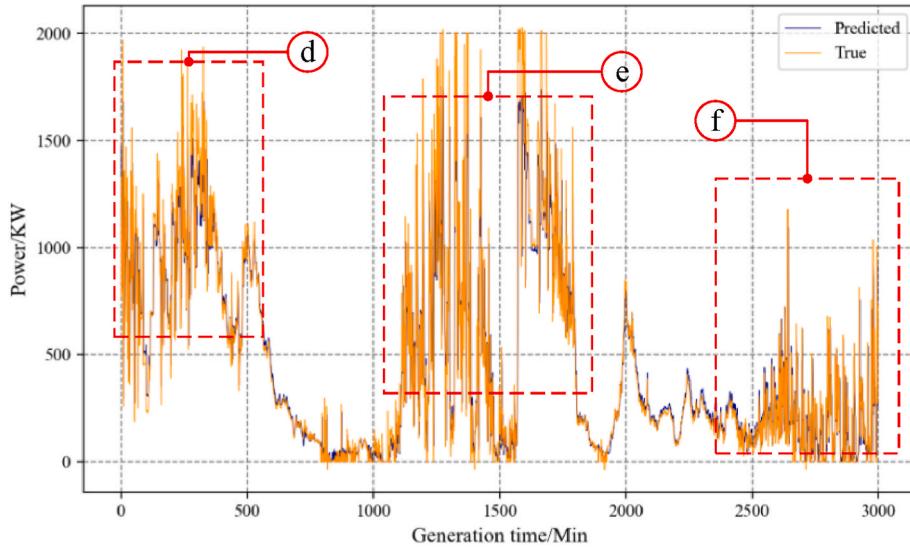


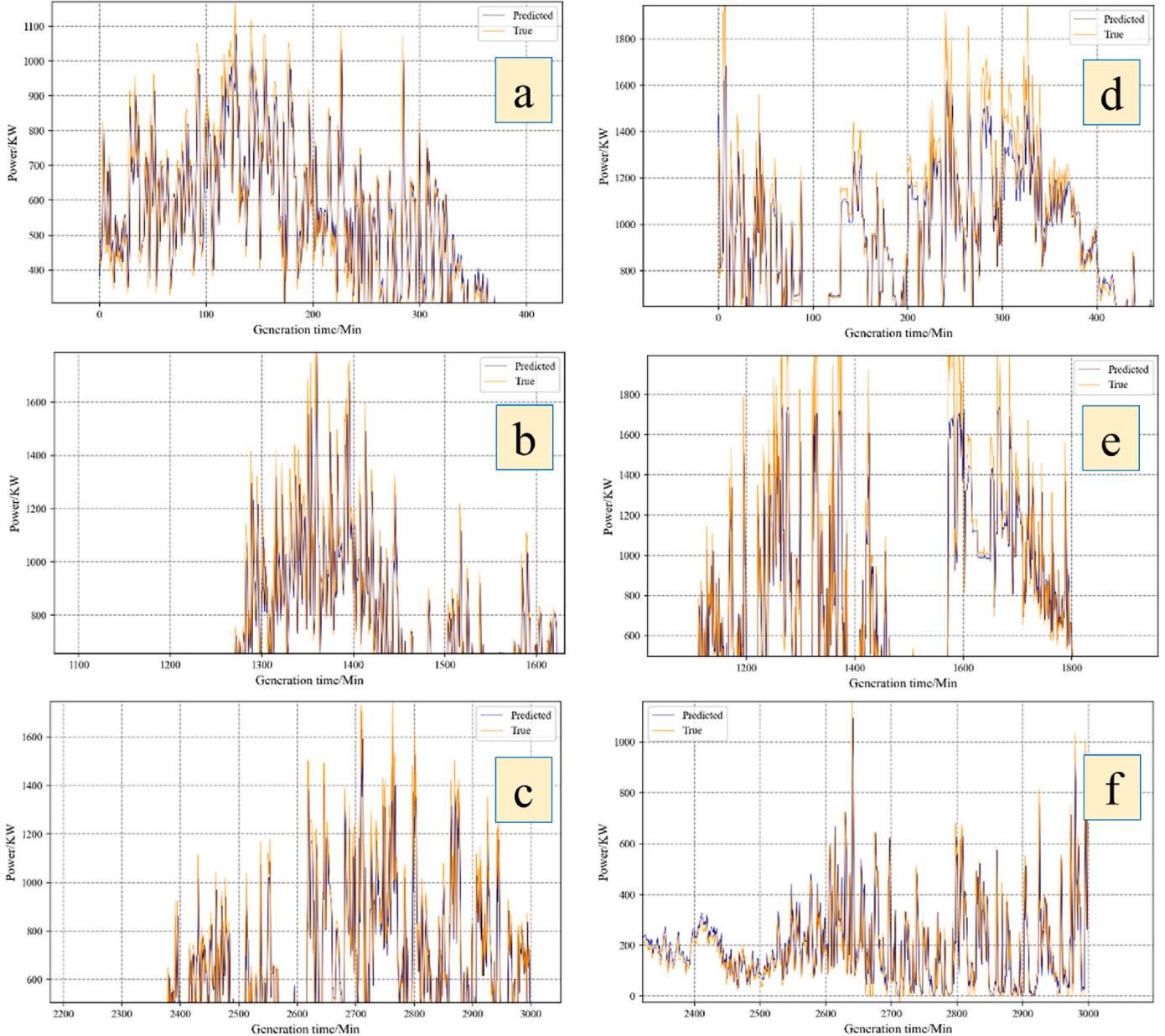
Fig. 6. Power prediction of WT2.

utilised. These components form the core of the drive chain, which is responsible for converting wind energy into kinetic energy (Li et al., 2020, 2022b, 2022c). The data employed in this study were sourced from the LGS-Offshore dataset and primarily encompassed two segments, as shown in Table 4: (i) SCADA data from eight OWTs equipped with 12 sensors, and (ii) SCADA data from two onshore wind turbines equipped with 13 sensors.

### 3.2. Results

To demonstrate the performance of the proposed method, two cases were implemented: (i) Case 1: Power prediction of onshore wind turbines with 13 sensors; and (ii) Case 2: Power prediction of OWTs with 12

sensors. Figs. 5 and 6 show the power forecasting results for onshore wind turbines 1 (WT1) and 2 (WT2) based on the proposed method. In Fig. 5, the predictions for 3000 timesteps are presented. Notably, in the timestep range of 500–1000, the predicted power exceeded the true values, similar to that in the range of 2300–2400, but remained below 500 kW. This discrepancy is attributed to factors such as low wind speeds, generator faults, and gearbox failures. In addition, as shown in Fig. 7, the true power in areas a, b, and c surpasses the predicted values. Notably, the wind power increased to a peak and then rapidly decreased within a duration of 8–16 h, influenced by meteorological conditions. Specifically, changes in wind speed directly result in power fluctuations because wind speed is the most critical factor determining wind power. Frequent shifts in the wind direction affect the rotor-swept area facing



**Fig. 7.** Detailed predictions of onshore wind turbines/WT1: (a)–(c); WT2: (d)–(f).

the wind, thereby influencing power generation.

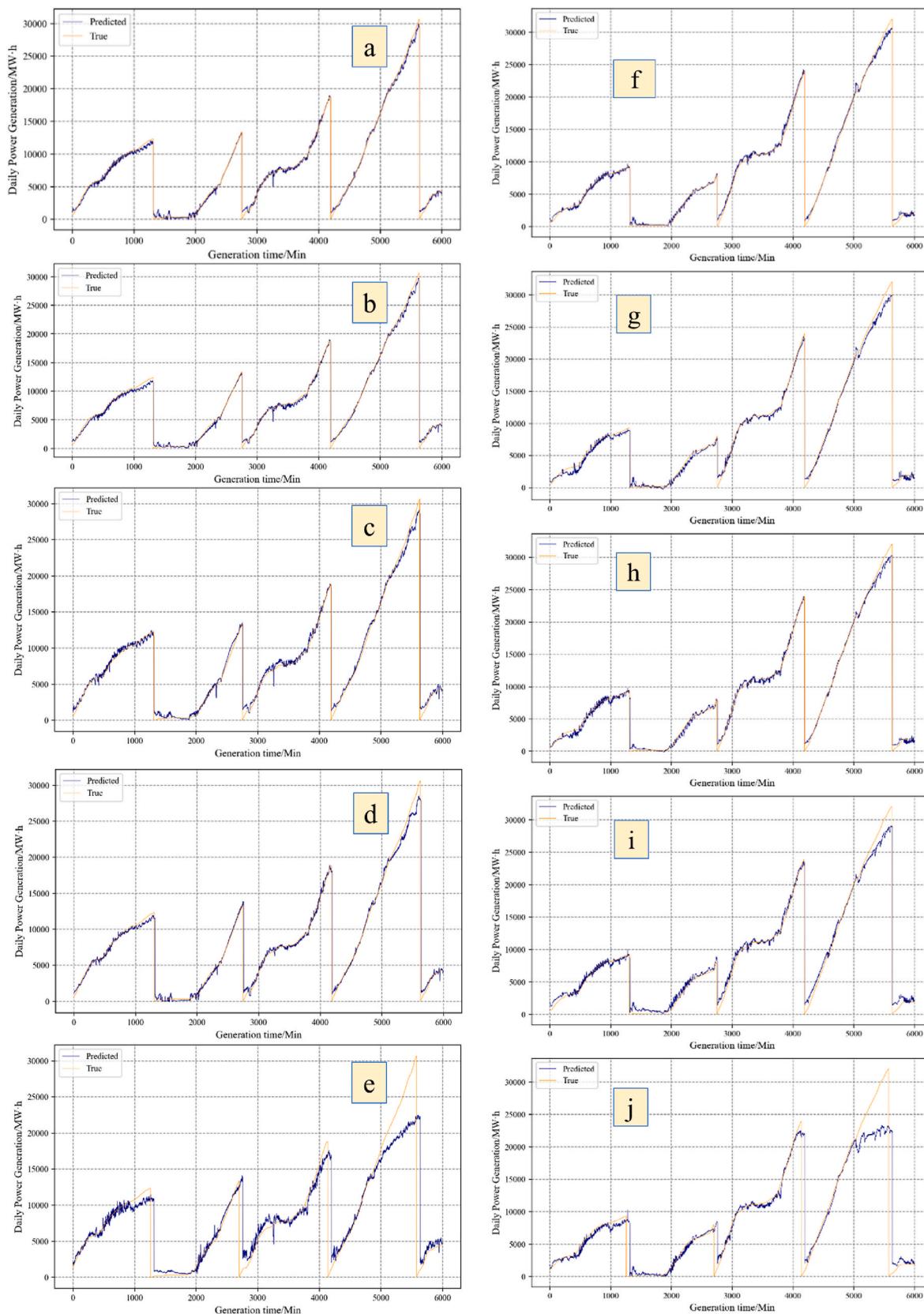
The RMSE, MAE, and R-square values for the power prediction of WT1 were 150.24, 0.81, and 105.12, respectively. These values were compared to those of WT2, which were 149.56, 30.24, and 101.36, respectively. Another notable observation is that the peak power of WT2 reaches its rated power in multiple timesteps. However, similar to WT1, the power curve of WT2 exhibits periodic fluctuations with intervals of approximately 500 timesteps.

**Fig. 8** shows the power predictions for multiple timesteps of offshore wind turbines OWT1 and OWT 8 at time points of 1, 2, 5, 10, and 60 min. The evaluation indicators for different prediction timesteps are listed in

**Table 5.** Multiple timestep predictions were achieved on the same dataset using timestep parameterisation for offshore wind-turbine prediction. It can be concluded that the accuracy of the first four types of timesteps was relatively high; however, the accuracy error started to decrease significantly after 60 min of prediction, as shown in Figs. 9 and 10. Notably, models with an R-square greater than 0.5 are considered to have good performance.

### 3.3. Comparison

To validate the accuracy of the CNN–LSTM–AM algorithm, 11 deep



**Fig. 8.** Detailed predictions of offshore wind turbines (OWTs)/OWT1: (a)–(e); OWT8: (f)–(j).

**Table 5**

Evaluation indicator of different prediction timesteps.

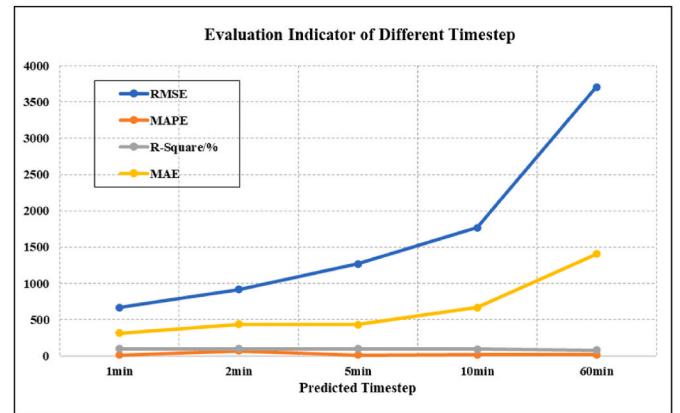
Offshore Wind Turbine 1				
Timestep	RMSE	MAPE	R-Square/%	MAE
1 min	572.33	12.167	99.2	267.696
2 min	815.29	15.35	98.6	334.48
5 min	1216.02	13.55	96	439.82
10 min	1613.39	30.35	94	456.17
60 min	3571	18	72.7	1524.42
Offshore Wind Turbine 8				
Timestep	RMSE	MAPE	R-Square/%	MAE
1 min	667.5	12.74	99.3	314.77
2 min	918.33	70.49	98.7	435.91
5 min	1272.45	13.78	97	432.87
10 min	1770.42	17.79	95	666.43
60 min	3708.94	19.79	79	1408.37

learning models for time-series prediction, including RNN, GRU, LSTM, RNN-AM, GRU-AM, LSTM-AM, CNN-RNN, CNN-GRU, CNN-LSTM, CNN-RNN-AM, and CNN-GRU-AM, were selected for comparison, and the two cases were calculated.

Table 6 provides an overview of the four power-prediction errors, clearly demonstrating that the CNN-LSTM-AM model surpasses the others in terms of performance. This consistency is further affirmed in Table 7, which presents the accuracy metrics. Compared with CNN-GRU-AM, a model with a similar structure, CNN-LSTM-AM, showed notable improvements. In Case 1, the accuracy of CNN-LSTM-AM outperforms CNN-GRU-AM by 11.7%, 9.46%, 7.4%, and 13.77%, as measured by the four indicators. Similarly, in Case 2, CNN-LSTM-AM demonstrated improvements of 9.12%, 8.81%, 11.41%, and 2.74%.

Figs. 11 and 12 show the distributions of the four errors in the two cases. Generally, incorporating CNN and AM into a time-sequence prediction model leads to reduced RMSE, MAPE, and MAE values. This indicates that using a single model for feature extraction is not as effective as using a hybrid model.

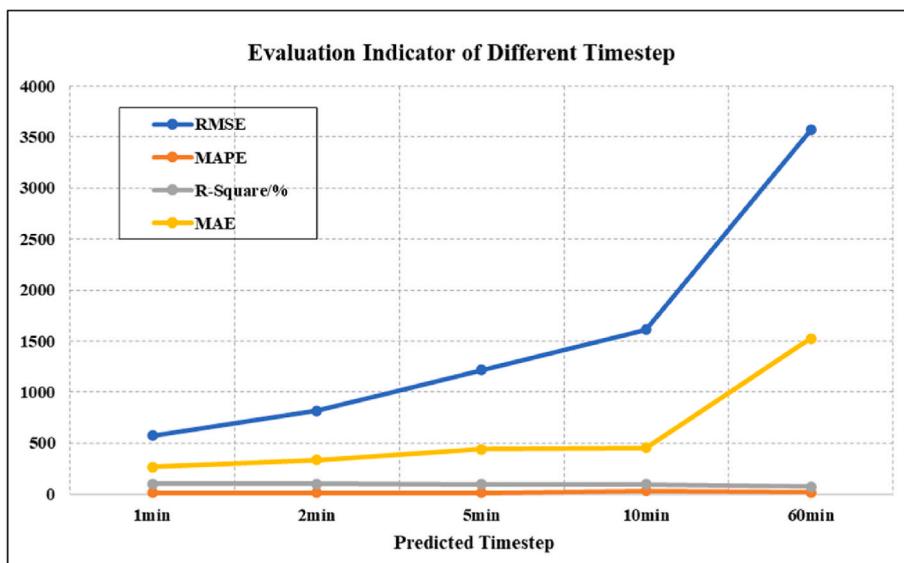
Achieving further improvements in prediction accuracy is challenging, even with an increase in data volume or hyperparameter

**Fig. 10.** Evaluation indicator of prediction timesteps (OWT8).**Table 6**

Error Comparison of deep learning models in Case 1.

Model	RMSE	MAPE	R-Square/%	MAE
RNN	170.29	35.83	75	119.98
GRU	161.34	32.82	78	108.45
LSTM	159.41	34.15	78	106.98
RNN-AM	158.12	40.41	79	102.04
GRU-AM	161.77	45.59	77.8	106.27
LSTM-AM	160.92	34.81	78	107.21
CNN-RNN	160.23	33.42	78.2	105.76
CNN-GRU	163.85	33.96	77	113.53
CNN-LSTM	157.67	35.56	79	102.25
CNN-RNN-AM	159.28	40.54	79	109.12
CNN-GRU-AM	170.17	34.54	75	121.91
CNN-LSTM-AM	150.24	31.27	81	105.12

adjustments. This limitation may stem from the absence of crucial meteorological data. In particular, this study considers the mechanical energy transmission process in the drive chain and its conversion into electrical energy. Therefore, data from the wind energy capture and initial conversion stages were not integrated into the prediction model.

**Fig. 9.** Evaluation indicator of prediction timesteps (OWT1).

**Table 7**  
Error comparison of deep learning models in Case 2.

Model	RMSE	MAPE	R-Square/%	MAE
RNN	171.46	35.66	76	117.61
GRU	160.33	32.43	78	108.08
LSTM	159.06	34.11	79	106.42
RNN-AM	159.32	40.54	78.7	105.69
GRU-AM	161.47	45.31	77.6	106.72
LSTM-AM	160.51	34.58	78.4	107.38
CNN-RNN	160.27	33.4	78.2	103.12
CNN-GRU	162.41	33.94	77.9	110.14
CNN-LSTM	157.64	35.46	79.4	106.54
CNN-RNN-AM	159.05	32.71	79.6	107.64
CNN-GRU-AM	164.56	33.16	75.4	104.22
CNN-LSTM-AM	149.56	30.24	84	101.36

### 3.4. Sensor sensitivity analysis

Sensor sensitivity analysis is crucial for guiding dataset construction and improving the accuracy of the prediction models. The sensor weights in the power prediction are measured by the error change rate, which is defined as

$$\text{Error change rate}_x = \frac{(\text{Prediction error}_{All} - \text{Prediction error}_{All/\{x\}})}{\text{Prediction error}_{All}}, \quad (19)$$

where Error change rate<sub>x</sub> represents the error change rate of sensor x, Prediction error<sub>All</sub> represents the prediction error of all sensors, and Prediction error<sub>All/\{x\}</sub> represents the prediction error after shielding a single sensor x.

When the error change rate is greater than 0, it indicates a decrease in the error without considering the sensor; the higher the absolute value, the higher the negative sensitivity. Conversely, when the error change rate is less than 0, it indicates an increase in the error without

considering the sensor; the higher the absolute value, the higher the positive sensitivity. Table 8 presents the error change rates of the 13 sensors for the onshore wind turbines. Fig. 13 illustrates the comparative results of the error change rates among the sensors of the onshore wind turbines.

Similarly, Table 9 presents the error change rates of the 12 sensors in the OWTs. Fig. 14 provides an explanatory visualisation, illustrating the comparative results of error change rates among the sensors of the OWTs.

Notably, for onshore wind turbines, the MAPE of the generator torque displayed a trend opposite to that of the other 12 sensor channels, with an exceptionally significant error change rate of  $-270.004\%$ . This indicates that removing the generator torque from the original sensor group significantly reduced the accuracy of the wind energy prediction model. In contrast, the error change rates for the other 12 sensors were all above zero, each exceeding 5%. This suggests that excluding these sensors from the wind energy forecasting process would enhance the accuracy of the predictions because data from these channels tend to contaminate the training dataset. The MAE change rate for the maximum temperature of the generator winding decreased slightly, although it was below zero. The temperature of the generator cooling water stood out significantly, with both MAE and RMSE showing sharp declines, making it the second-most critical sensor in the prediction model. Finally, the modest trends in the RMSE and R-square in the line chart highlight the relative insensitivity of the remaining onshore sensors.

When analysing Table 9 and Fig. 14 collectively, it becomes apparent that the error change rates for the RMSE and MAPE are remarkably similar. Notably, both the hub and rear-end temperatures of the high-speed shaft of the gearbox exhibited substantial decreases. This observation strongly suggests that the two sensor channels play a critical role in ensuring accurate offshore wind energy predictions.

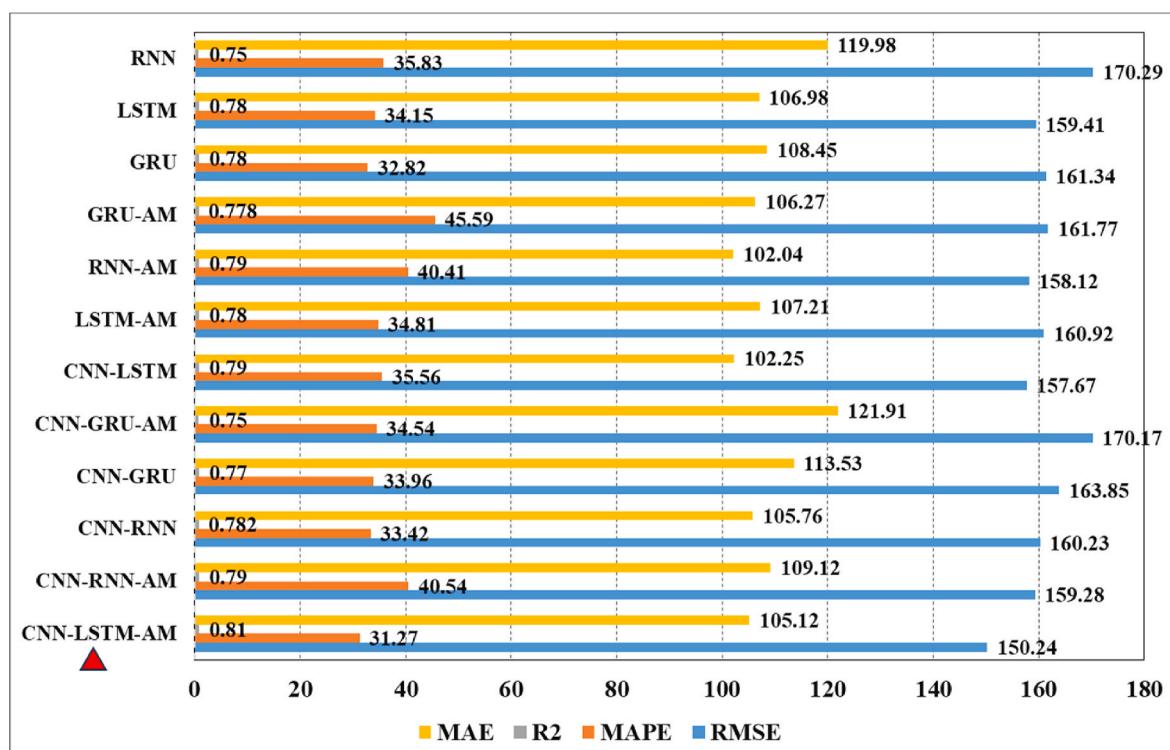


Fig. 11. Error comparison of deep learning models: Case 1/CNN-LSTM-AM: the proposed model.

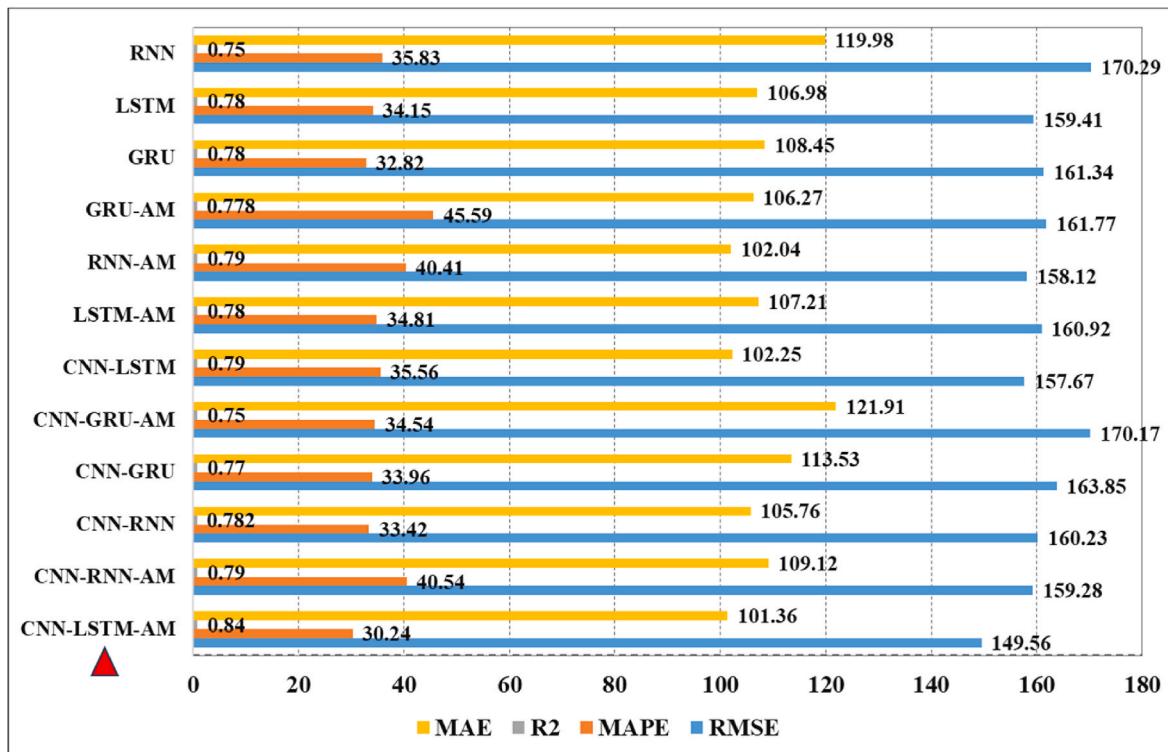


Fig. 12. Error comparison of deep learning models: Case 2/CNN-LSTM-AM: the proposed model.

Table 8

Sensitivity of onshore wind turbine sensors.

Sensor	RMSE/%	MAPE/%	R-Square/%	MAE/%
Generator speed	-0.934	-16.749	0.353	2.311
Max Temp of generator winding	-0.687	-2.519	0.235	-0.982
U1 Temp of generator winding	-0.139	-4.015	0.118	3.904
Generator Cooling water Temp	-2.754	-14.089	1.058	-2.277
Generator front bearing Temp	-1.089	-9.313	0.470	0.170
Generator rear bearing Temp	-0.225	-3.135	0.118	3.833
Generator slip ring chamber Temp	-0.363	-5.701	0.118	0.148
Generator torque	-1.611	-270.004	0.588	-0.215
Front bearing Temp of Main bearing	-0.139	-1.510	0.118	2.784
Rear bearing Temp of Main bearing	-0.604	-6.344	0.235	4.397
Hydraulic pressure of shaft brake	-0.442	-1.556	0.000	3.649
Temp inside the converter	-1.067	-7.164	0.470	5.300
DC voltage inside the converter	0.142	-8.699	0.000	4.093

### 3.5. Discussions

This case study illustrates that by amalgamating sensor data from various mechanical components and electromechanical devices, the multi-sensor fusion model addresses the issue of suboptimal wind power prediction performance owing to meteorological sensor malfunctions or data loss. Furthermore, the CNN-LSTM-AM algorithm demonstrated its capability for accurate wind power generation prediction. The comparative results show that CNN-LSTM-AM outperforms other deep-learning methodologies and their combinations. The sensitivity of multiple sensors in predicting the wind power generation capacity was determined, and the influence of different sensor types on the prediction accuracy was quantified. This allows for the rapid provision of a sensor-type selection scheme for new prediction scenarios.

However, several underlying issues require further investigation in future research. Currently, forecast results for wind power generation are exclusively employed for the sensitivity analysis of sensors. This implies that these predictions have not yet been reintegrated into model optimisation endeavours. The parameters of the model and its predictive outputs can function as inputs for parameter optimisation models. By conducting successive rounds of optimisation, the parameter combination can be identified, leading to the lowest prediction error and thereby enhancing the model's forecasting accuracy. Consequently, reoptimisation of the model has emerged as a crucial area for subsequent studies.

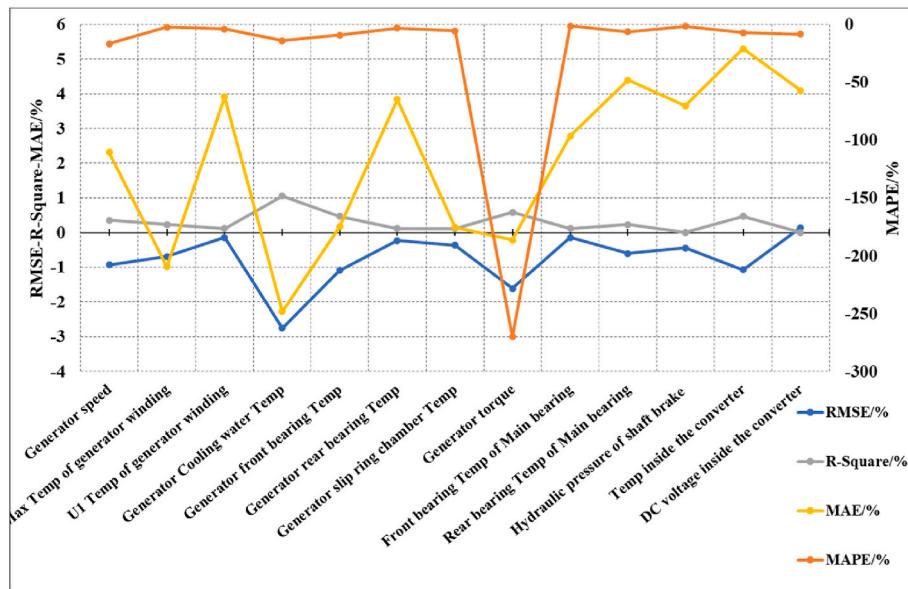


Fig. 13. Error change rate of onshore wind turbine sensors.

**Table 9**  
Sensitivity of OWT sensors.

Sensor	RMSE/%	MAPE/%	R-Square/%	MAE/%
Rotor speed	33.553	33.673	57.786	33.551
Hub Temp	-25.125	-25.000	-58.522	-24.948
Front end Temp of gearbox high-speed shaft	33.786	33.673	58.105	33.753
Rear end Temp of gearbox high-speed shaft	-17.806	-18.367	-40.126	-17.977
Gearbox oil pool Temp	4.119	3.571	8.350	3.839
Gearbox inlet oil Temp	25.541	26.020	46.106	25.932
Gearbox inlet pressure	30.846	30.612	53.988	30.870
Outlet pressure of gearbox oil pump	27.034	26.531	48.382	26.817
U1 Temp of generator winding	-2.294	-2.551	-4.801	-2.401
Generator Cooling water Temp	15.052	14.796	28.806	15.001
Generator front bearing Temp	29.840	30.102	52.538	29.885
Generator rear bearing Temp	61.555	62.245	88.178	62.214

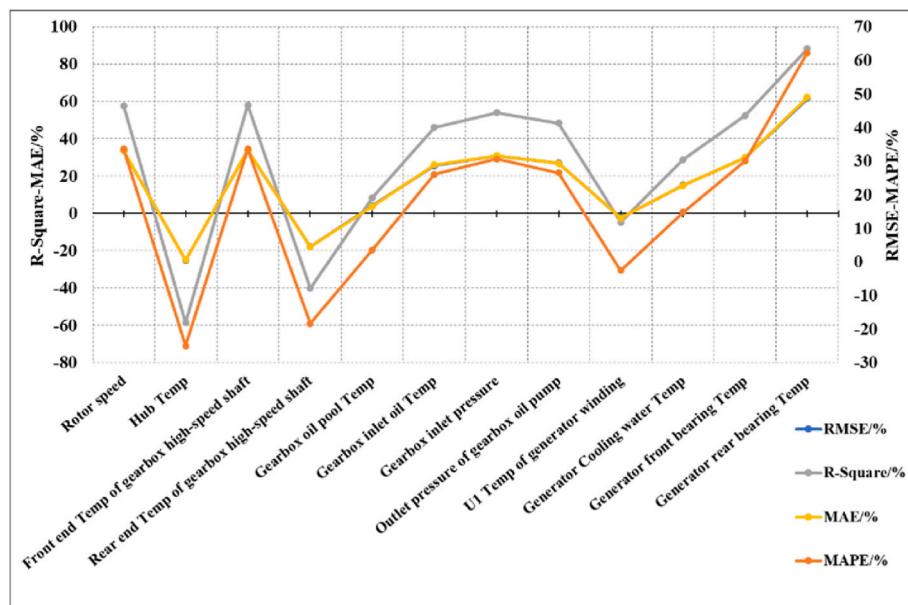


Fig. 14. Error change rate of onshore wind turbine sensors.

#### 4. Conclusion

This study introduces a wind energy forecasting framework based on multi-sensor fusion and the CNN-LSTM-AM algorithm. By employing multi-sensor fusion with the CNN-LSTM-AM algorithm, an alternative method for wind energy prediction was achieved by utilising multiple electromechanical equipment sensors in the absence of meteorological data. Automatic partitioning of the training dataset, supported by timestep parameterisation, allowed multiple timestep forecasting to run on the same dataset through only parameter adjustments. The sensor sensitivity was quantified by calculating the rate of error change of the four evaluation indicators. The study concludes that (i) the forecast results for wind-generated power indicate that the multi-sensor fusion model performs better and with higher accuracy in predicting intervals of low generated power. (ii) Multiple timestep predictions were successfully made on the same dataset through variable prediction timestep parameterisation. The accuracy in the first four types of timesteps is relatively high and starts to decrease significantly after 60 min of prediction. (iii) The accuracy of CNN-LSTM-AM improved by 11.7%, 9.46%, 7.4%, and 13.77% in Case 1 and by 9.12%, 8.81%, 11.41%, and 2.74% in Case 2, as measured by the four evaluation indicators, compared with CNN-GRU-AM. (iv) The sensor sensitivity analysis results indicate that generator torque is the most sensitive sensor in the wind energy prediction model, followed by generator cooling water temperature as the second most crucial sensor in the prediction model.

#### CRediT authorship contribution statement

**Yu Sun:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Qibo Zhou:** Writing – original draft, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis. **Li Sun:** Writing – original draft, Resources, Project administration, Investigation, Funding acquisition, Formal analysis. **Liping Sun:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Funding acquisition. **Jichuan Kang:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis. **He Li:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

#### 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.

#### Data availability

The authors do not have permission to share data.

#### Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant No. 52101305 and 72301299) and the Marie Skłodowska-Curie Postdoctoral Fellowship (No. EPSRC EP/Z001501/1).

#### References

- Ackermann, T., Söder, L., 2000. Wind energy technology and current status: a review. *Renew. Sustain. Energy Rev.* 4 (4), 315–374.
- Bartholomew, David J., 1971. Time series analysis forecasting and control. *Oper. Res. Q.* 199–201.
- Bird, L., Lew, D., Milligan, M., et al., 2016. Wind and solar energy curtailment: a review of international experience. *Renew. Sustain. Energy Rev.* 65, 577–586.
- Costa, A., Crespo, A., Navarro, J., et al., 2008. A review on the young history of the wind power short-term prediction. *Renew. Sustain. Energy Rev.* 12 (6), 1725–1744.
- Council GWE, 2023. Global wind report 2023. Available at: [https://gwec.net/wp-content/uploads/2023/04/GWEC-2023\\_interactive.pdf](https://gwec.net/wp-content/uploads/2023/04/GWEC-2023_interactive.pdf).
- Donadio, L., Fang, J., Porté-Agel, F., 2021. Numerical weather prediction and artificial neural network coupling for wind energy forecast. *Energies* 14 (2), 338.
- Foley, A.M., Leahy, P.G., Marvuglia, A., et al., 2012. Current methods and advances in forecasting of wind power generation capacity. *Renew. Energy* 37 (1), 1–8.
- Graves, A., Schmidhuber, J., 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Network*. 18 (5–6), 602–610.
- Greff, K., Srivastava, R.K., Koutník, J., et al., 2015. LSTM: a search space odyssey. *Neural Networks and Learning Systems* 10, 8.
- Hanifi, S., Liu, X., Lin, Z., et al., 2020. A critical review of wind power forecasting methods—past, present and future. *Energies* 13 (15), 3764.
- Hanifi, S., Zare-Behtash, H., Cammarano, A., et al., 2023. Offshore wind power forecasting based on WPD and optimised deep learning methods. *Renew. Energy* 218, 119241.
- Heydari, A., Majidi Nezhad, M., Neshat, M., et al., 2021. A combined fuzzy GMDH neural network and grey wolf optimization application for wind turbine power production forecasting considering SCADA data. *Energies* 14 (12), 3459.
- Hochreiter, S., Jürgen, S., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780.
- Hu, S., Xiang, Y., Zhang, H., et al., 2021. Hybrid forecasting method for wind power integrating spatial correlation and corrected numerical weather prediction. *Appl. Energy* 293, 116951.
- Jung, J., Broadwater, R.P., 2014. Current status and future advances for wind speed and power forecasting. *Renew. Sustain. Energy Rev.* 31, 762–777.
- Jung, S., Kwon, S.D., 2013. Weighted error functions in artificial neural networks for improved wind energy potential estimation. *Appl. Energy* 111, 778–790.
- Jyothi, M.N., Rao, P.V.R., 2016. Very-short term wind power forecasting through Adaptive wavelet neural network. In: Proceedings of the 2016-Biennial International Conference on Power and Energy Systems: towards Sustainable Energy, PESTSE 2016, Bengaluru, India. IEEE, Piscataway, NJ, USA, pp. 21–23.
- Kang, J., Sun, L., Guedes Soares, C., 2019. Fault Tree Analysis of floating offshore wind turbines. *Renew. Energy* 133, 1455e1467.
- Kisvari, A., Lin, Z., Liu, X., 2021. Wind power forecasting—A data-driven method along with gated recurrent neural network. *Renew. Energy* 163, 1895–1909.
- Landberg, L., Myllerup, L., Rathmann, O., et al., 2003. Wind resource estimation—an overview. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology* 6 (3), 261–271.
- LeCun, Y., Bottou, L., Bengio, Y., et al., 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86 (11), 2278–2324.
- Li, H., Guedes Soares, C., 2022. Assessment of failure rates and reliability of floating offshore wind turbines. *Reliab. Eng. Syst. Saf.* 228, 108777.
- Li, H., Guedes Soares, C., Huang, Hong-Zhong, 2020. Reliability analysis of a floating offshore wind turbine using Bayesian Networks. *Ocean. Eng.* 217, 107827.
- Li, H., Díaz, H., Guedes Soares, C., 2021a. A developed failure mode and effect analysis for floating offshore wind turbine support structures. *Renew. Energy* 164, 133–145.
- Li, H., Díaz, H., Guedes Soares, C., 2021b. A failure analysis of floating offshore wind turbines using AHP-FMEA methodology. *Ocean. Eng.* 234, 109261.
- Li, M., Jiang, X., Carroll, J., Negenborn, R.R., 2022a. A multi-objective maintenance strategy optimization framework for offshore wind farms considering uncertainty. *Appl. Energy* 321, 119284.
- Li, H., Huang, Cheng-Geng, Guedes Soares, C., 2022b. A real-time inspection and opportunistic maintenance strategies for floating offshore wind turbines. *Ocean. Eng.* 256, 111433.
- Li, H., Peng, W., Huang, C.G., Guedes Soares, C., 2022c. Failure rate assessment for onshore and floating offshore wind turbines. *J. Mar. Sci. Eng.* 10 (12), 1965.
- Lin, Z., Liu, X., 2020. Wind power forecasting of an offshore wind turbine based on high-frequency SCADA data and deep learning neural network. *Energy* 201, 117693.
- Lu, W., Li, J., Wang, J., et al., 2021. A CNN-BiLSTM-AM method for stock price prediction. *Neural Comput. Appl.* 33, 4741–4753.
- Manero, J., Béjar, J., Cortés, U., 2018. Wind energy forecasting with neural networks: a literature review. *Comput. Sist.* 22 (4), 1085–1098.
- McDonagh, S., Ahmed, S., Desmond, C., et al., 2020. Hydrogen from offshore wind: investor perspective on the profitability of a hybrid system including for curtailment. *Appl. Energy* 265, 114732.
- Mnih, V., Heess, N., Graves, A., 2014. Recurrent models of visual attention. *Adv. Neural Inf. Process. Syst.* 27.
- Neshat, M., Nezhad, M.M., Abbasnejad, E., et al., 2021. A deep learning-based evolutionary model for short-term wind speed forecasting: a case study of the Lillgrund offshore wind farm. *Energy Convers. Manag.* 236, 114002.
- Neshat, M., Nezhad, M.M., Mirjalili, S., et al., 2022. Quaternion convolutional long short-term memory neural model with an adaptive decomposition method for wind speed forecasting: north aegean islands case studies. *Energy Convers. Manag.* 259, 115590.
- Nezhad, M.M., Neshat, M., Sylaios, G., et al., 2024. Marine energy digitalization digital twin's approaches. *Renew. Sustain. Energy Rev.* 191, 114065.
- Qian, Z., Pei, Y., Zareipour, H., et al., 2019. A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. *Appl. Energy* 235, 939–953.
- Qureshi, A.S., Khan, A., Zameer, A., et al., 2017. Wind power prediction using deep neural network based meta regression and transfer learning. *Appl. Soft Comput.* 58, 742–755.
- Raza, M.Q., Hosravi, A., 2015. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renew. Sustain. Energy Rev.* 50, 1352–1372.

- Sun, Y., Kang, J., Sun, L., et al., 2022. Condition-based maintenance for the offshore wind turbine based on long short-term memory network. *Proc. Inst. Mech. Eng. O J. Risk Reliab.* 236 (4), 542–553.
- Sun, Y., Li, H., Sun, L., et al., 2023. Failure analysis of floating offshore wind turbines with correlated failures. *Reliab. Eng. Syst. Saf.* 238, 109485.
- Tawn, R., Browell, J., 2022. A review of very short-term wind and solar power forecasting. *Renew. Sustain. Energy Rev.* 153, 111758.
- Vaswani, A., Shazeer, N., Parmar, N., et al., 2017. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30.
- Wang, H., Han, S., Liu, Y., et al., 2019. Sequence transfer correction algorithm for numerical weather prediction wind speed and its application in a wind power forecasting system. *Appl. Energy* 237, 1–10.
- Xiang, L., Wang, P., Yang, X., et al., 2021. Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism. *Measurement* 175, 109094.
- Xing, Z., He, Y., 2023. Multi-modal multi-step wind power forecasting based on stacking deep learning model. *Renew. Energy* 215, 118991.
- Yatiyana, E., Rajakaruna, S., Ghosh, A., 2017. Wind speed and direction forecasting for wind power generation capacity using ARIMA mode. 2017 Australasian Universities Power Engineering Conference (AUPEC). IEEE, pp. 1–6.
- Yu, M., Zhang, Z., Li, X., et al., 2020. Superposition graph neural network for offshore wind power prediction. *Future Generat. Comput. Syst.* 113, 145–157.
- Zare, S., Ayati, M., 2021. Simultaneous fault diagnosis of wind turbine using multichannel convolutional neural networks. *ISA Trans.* 108, 230–239.
- Zhang, W., Lin, Z., Liu, X., 2022. Short-term offshore wind power forecasting-A hybrid model based on discrete wavelet transform (DWT), seasonal autoregressive integrated moving average (SARIMA), and deep-learning-based long short-term memory (LSTM). *Renew. Energy* 185, 611–628.