

## Article

# A Normal Behavior-Based Condition Monitoring Method for Wind Turbine Main Bearing Using Dual Attention Mechanism and Bi-LSTM

Xiaocong Xiao <sup>1,2</sup>, Jianxun Liu <sup>2,\*</sup>, Deshun Liu <sup>1</sup>, Yufei Tang <sup>3</sup>, Shigang Qin <sup>1</sup> and Fan Zhang <sup>1</sup><sup>1</sup> School of Mechanical Engineering, Hunan University of Science and Technology, Xiangtan 411201, China<sup>2</sup> School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China<sup>3</sup> Department of Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431, USA

\* Correspondence: liujx@hnust.edu.cn

**Abstract:** As clean and low-carbon energy, wind energy has attracted the attention of many countries. The main bearing in the transmission system of large-scale wind turbines (WTs) is the most important part. The research on the condition monitoring of the main bearing has received more attention from many scholars and the wind industry, and it has become a hot research topic. The existing research on the condition monitoring of the main bearing has the following drawbacks: (1) the existing research assigns the same weight to each condition parameter variable, and the model extracts features indiscriminately; (2) different historical time points of the condition parameter variable are given the same weight, and the influence degree of different historical time points on the current value is not considered; and (3) the existing literature does not consider the operating characteristics of WTs. Different operating conditions have different control strategies, which also determine which condition parameters are artificially controlled. Therefore, to solve the problems above, this paper proposes a novel method for condition monitoring of WT main bearings by applying the dual attention mechanism and Bi-LSTM, named Dual Attention-Based Bi-LSTM (DA-Bi-LSTM). Specifically, two attention calculation modules are designed to extract the important features of different input parameters and the important features of input parameter time series, respectively. Then, the two extracted features are fused, and the Bi-LSTM building block is utilized to perform pre-and post-feature extraction of the fused information. Finally, the extracted features are applied to reconstruct the input data. Extensive experiments verify the performance of the proposed method. Compared with the Bi-LSMT model without adding an attention module, the proposed model achieves 19.78%, 2.17%, and 18.92% improvement in MAE, MAPE, and RMSE, respectively. Compared with the Bi-LSTM model which only considers a single attention mechanism, the proposed model achieves the largest improvement in MAE and RMSE by 28.84% and 30.37%. Furthermore, the proposed model has better stability and better interpretability of the monitoring process.

**Keywords:** wind turbine; main bearing; condition monitoring; attention mechanism; Bi-LSTM



**Citation:** Xiao, X.; Liu, J.; Liu, D.; Tang, Y.; Qin, S.; Zhang, F. A Normal Behavior-Based Condition Monitoring Method for Wind Turbine Main Bearing Using Dual Attention Mechanism and Bi-LSTM. *Energies* **2022**, *15*, 8462. <https://doi.org/10.3390/en15228462>

Academic Editor: Davide Astolfi

Received: 16 September 2022

Accepted: 9 November 2022

Published: 12 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

At present, the new round of the world's energy pattern is undergoing in-depth adjustment and fundamental transformation, dominated by clean, low-carbon, smart, and efficient energy sources and supplemented by fossil energy. Gradually reducing the consumption of fossil energy and gradually increasing the proportion of renewable energy is becoming a new state [1,2]. Wind energy is one of the clean and low-carbon green renewable energy sources, which has naturally attracted the attention of many countries. Since the end of the 20th century, especially in the past ten years, many large-scale wind farms and WTs have been put into use. Owing to the severe operating environment and complicated working

conditions, WT are prone to failure, and the reasons for some failures are still unclear [3]. To assure the stable and dependable operation of WT and reduce operational and maintenance costs, it is crucial to carry out condition monitoring and fault diagnosis [4,5].

The main bearing is a core component of the transmission system in the wind turbine [6,7]. However, the WT main bearing is a low-speed and low-frequency large component, its operating condition is very complex, and the dynamic load and dynamic behavior during operation are, so far, not clear [8,9]. According to literature reports, alternating load and strong impact are some of the main reasons for the failure of WT main bearing, and its fault rate has reached 15% to 30% [10]. The condition monitoring and fault diagnosis of the main bearing of WT is worth studying. For WT with a 20-year service life, which are equipped with SCADA systems as standard, wind farms have generated enormous SCADA data. Many scholars have carried out some research work on the main bearing of wind turbines due to the availability of abundant data collected by the SCADA system and the necessity of not installing other hardware devices [11,12]. Natili et al. [13] studied the main bearing temperature trends of SCADA systems and built a support vector model of normal behavior to monitor the difference between measured and estimated values to identify impending failures. Beretta et al. [14] constructed a main bearing condition identification model by using some SCADA parameters, such as main bearing temperature, wind speed, rotor speed, and power. Guo et al. [15] propose a condition monitoring method for the main bearing of WT using Gaussian process regression and double sliding window residual processing based on some condition parameters such as temperature, wind speed, power, and torque. Also based on some SCADA condition parameters, the authors [16–18] take the previous values of these parameters to build machine-learning models to identify main bearing failures. The above research work mainly focuses on the main bearing temperature parameters and their related parameters. Historical information about main bearing condition parameters is rarely considered. Some studies use weekly statistical indicators for research and analysis; however, the research results are rough, and the modeling accuracy is not enough. Therefore, some modeling methods that consider more historical information of main bearing condition parameters and new deep learning models have been proposed [19–21]. Some researchers have also begun to pay attention to the study on shaft misalignment [22], main bearing degradation [23], and remaining service life [24].

The above studies mainly focus on the related parameters of the WT drive chain in the SCADA system, such as temperature, vibration, voltage, current, and some external environmental parameters. The research results deepen the understanding of the operation monitoring of the main bearing. These proposed data-driven models also improve the condition monitoring level and the ability of condition abnormal identification, especially in some deep learning model building, such as autoencoder, and LSTM. However, there are still three shortcomings in the above research: (1) The existing literature assigns the same weight to each parameter variable so that the model can extract features indiscriminately during the model learning, but in the actual process, some parameters change slowly, some parameters change rapidly, and the influence degree of each parameter varies, so it is necessary to assign different weights to these parameters, and select some parameter variables in a targeted and differentiated manner. (2) The current value of the parameter variable is influenced by its historical value, that is, the degree of influence of different historical time points on the current value is also different, and it is necessary to assign different weights to different historical time points. (3) The existing literature does not consider the operating characteristics of the WT. Different wind speed ranges have different control strategies. For example, during the maximum wind energy tracking phase, the WT adjusts the rotor speed according to the wind speed to keep the optimal tip speed ratio. The speed ratio enables the WT to obtain the maximum wind energy, and in the constant power stage, the WT maintains a constant rotational speed, so that the WT has a constant output power by adjusting the pitch angle.

To solve the above problems, taking the large-scale direct-driven WT main bearing of the transmission system as the research object, this paper put forward a condition monitoring method for the WT main bearing by using the dual attention mechanism and Bi-

LSTM, named Dual Attention-Based Bi-LSTM (DA-Bi-LSTM). Recently, some deep learning models incorporating attention mechanisms have shown good performance [25–27]. In terms of feature extraction of time series data, some of them have been applied in the field of wind power. Su et al. [28] utilize a dual attention mechanism and gated recurrent unit (GRU) to realize the condition monitoring of an offshore wind turbine gearbox based on SCADA data. Xiang et al. [29] use convolutional neural network (CNN) and LSTM with attention mechanism based on SCADA data to diagnose early failures of wind turbines. Motivated by the above research, we introduce the attention mechanism and Bi-LSTM model to the condition monitoring of the wind turbine main bearing. The main contributions of this paper are summarized in the following three points:

(1) A novel operating condition monitoring method for WT main bearing by using WT operating characteristics and SCADA dataset is proposed, named DA-Bi-LSTM, based on the dual attention mechanism and Bi-LSTM.

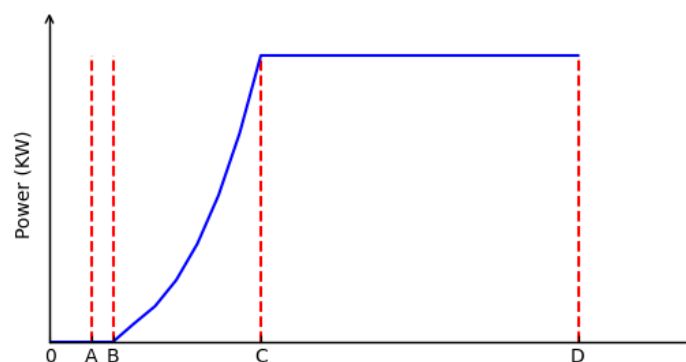
(2) Utilizing the dual attention mechanism, the proposed model can further probe the spatiotemporal relationship among the condition parameters of the main bearing itself and its related condition parameters, strengthen the key information in the input data, and weaken the secondary information. The parameter attention and temporal attention mechanism can express the interpretability of the DA-Bi-LSTM model more clearly.

(3) Extensive experiments are conducted on a real-world SCADA dataset. The experimental results indicate that the proposed model outperforms Bi-LSTM and two other Bi-LSTM models that only consider a single attention mechanism. The proposed model has better stability and better interpretability of the monitoring process.

The remainder of the paper is organized as follows. Section 2 depicts the working condition analysis and condition parameters selection of large-scale direct-driven WTs. The proposed condition monitoring model DA-Bi-LSTM is presented in Section 3, including the problem definition, the framework for the proposed model with six submodules, and the training algorithm. The experimental setup, results, performance analysis, and interpretability are presented in Section 4. Conclusions are drawn in Section 5.

## 2. Working Condition Analysis and Parameters Selection of Large-Scale Direct-Driven WT

According to the control principle of the direct-driven WT, its working conditions can be divided into four stages. Figure 1 gives a detailed illustration.



**Figure 1.** The curve of wind speed and power output of direct-driven wind turbine.

(1) Shutdown stage (seen in OA and D+): This stage includes two working conditions. Working condition 1 (OA stage) means that the wind speed is below the cut-in wind speed set on the nameplate, and the wind speed at this time is not enough to start the WT. Working condition 2 (D+ stage) means the wind speed exceeds the cut-out wind speed set on the nameplate. At this moment, the wind speed is very large. From the perspective of safety and protection of the WT, the WT is forced to stop.

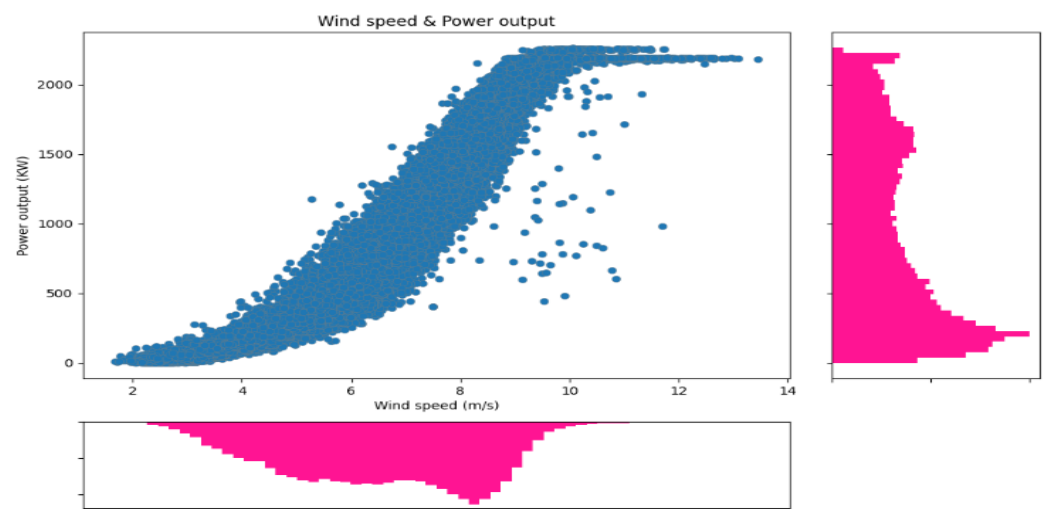
(2) Startup stage (seen in AB): The characteristics of the working condition in this stage is that the wind speed exceeds the cut-in wind speed. After the startup condition is

satisfied, the blades are turned from the stop position to the starting position (set to  $30^\circ$  in our study), but the rotational speed of the hub is very low, less than the cut-in speed (set to 3.5 r/min in our study), and the WT is pre-generating.

(3) Maximum wind energy tracking stage (seen in BC): The characteristics of working condition in this stage is that the wind speed is between the cut-in wind speed and the rated wind speed. The hub speed is greater than the cut-in speed (3.5 r/min in our study) and less than the rated speed. The blades turn to the working position and remain there at all times. At this stage, the WT control system adjusts the speed of the blades according to the wind speed, keeps the optimal tip speed ratio, and maximizes the wind power utilization coefficient.

(4) Constant power stage (seen in CD): The characteristics of the working condition in this stage is that the wind speed is between the rated wind speed and the cut-out wind speed. The pitch control system adjusts the pitch angle of the blade by driving the pitch motor, variable speed gearbox, and pitch bearing, and then controls the rotational speed of the hub. The change of the pitch angle can keep the speed of the hub around the rated speed, and the output power is around the rated output power. At this time, the wind energy utilization coefficient is not the optimal value.

From the above analysis, it can be seen that the direct-drive WTs have different control strategies at different stages. Based on the data we collected from our research object of large-scale direct-drive WTs, we derived a scatter plot of wind speed and output power, as well as a distribution map of the data in the two dimensions of wind speed and output power, which are shown in Figure 2. In Figure 2, the BC stage and the CD stage are the main operating stages of WT operation. These two stages mainly involve four core parameters, namely wind speed, power, rotational speed, and pitch angle. Specifically, the main feature of the BC stage is that the rotational speed is adjusted according to the wind speed, and the pitch angle is always kept constant. In this way, the maximum wind energy utilization coefficient value is the largest, which also ensures that the WT can obtain more wind energy and power. The main feature of the CD stage is to adjust the pitch angle based on the wind speed to keep the rotation speed almost constant, the output power is always stable near the rated power, and the WT is in a full state. Therefore, like most researchers, this paper also selects the data sets of these two segments for modeling and analysis.



**Figure 2.** The scatter and distribution of wind speed and power.

To accurately monitor and evaluate the operating condition of the main bearing, combined with our previous research results [30], there are eleven parameters that should be considered in the BC stage, namely, wind speed, ambient temperature, main bearing temperature, rotor speed, output power, generator operation frequency, generator torque, generator stator temperature, 5-s mean yaw to wind, vibration in the X direction, and

vibration in the Y direction. In the CD stage, the pitch angle parameter should be added, so there are twelve parameters in this stage.

### 3. Proposed Condition Monitoring Model DA-Bi-LSTM

In this section, we first give the problem definition of condition monitoring of the WT main bearing. Then, we present the framework for the proposed model DA-Bi-LSTM. Finally, we design the corresponding algorithm pseudocode for training the DA-Bi-LSTM model.

#### 3.1. Problem Definition

Under normal conditions, the condition data of the WT main bearing itself and its associated condition data will satisfy a dynamic and stable internal equilibrium relationship, with the data fluctuating within a certain range and maintaining some constraint characteristics between them. In the event of a failure of the main bearing, the dynamic equilibrium is broken. The WT main bearing condition monitoring problem can be defined as a combination of a nonlinear model and reconstruction error, which is described as follows:

$$\begin{cases} Z(t) = F(X(t)) \\ R_e = |Z(t) - X(t)| \end{cases} \quad (1)$$

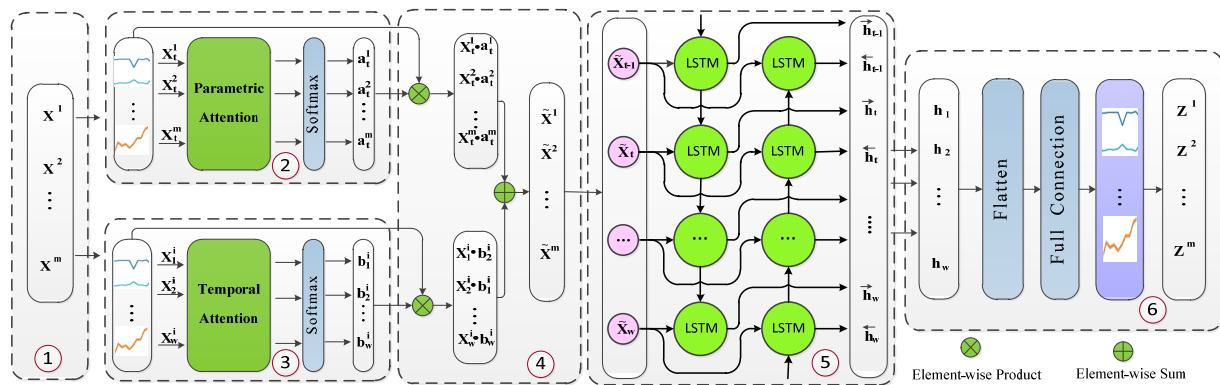
where  $t$  denotes sampling time,  $X(t)$  is the condition parameters describing the main bearing at the  $t$ th moment, which is the input vector of the function  $F(\cdot)$ , especially,  $X(t) = \{X^1(t), X^2(t), \dots, X^m(t)\} \in R^{m \times w}$ .  $Z(t)$  is the output value of the function  $F(\cdot)$ , which represents the reconstructed value of the condition parameters of the main bearing at the  $t$ th moment, especially,  $Z(t) = \{Z^1(t), Z^2(t), \dots, Z^m(t)\} \in R^{m \times w}$ .  $m$  is the number of feature parameters that characterize the operating condition of the main bearing.  $w$  represents the length of the sliding window.  $X^i(t)$  is a row vector containing the  $i$ th condition parameter with  $w$  measured values, and  $Z^i(t)$  is also a row vector of the  $i$ th condition parameter with  $w$  predicted values.  $F(\cdot)$  represents a series of nonlinear mapping functions.  $R_e$  is the reconstruction error value. Therefore, the essence of condition monitoring of the main bearing is that the reconstructed vector  $Z(t)$  is as similar as possible to the input vector  $X(t)$ , with a minimum residual between the two vectors.  $R_e$  represents the residual value. The  $R_e$  value will exceed the threshold or alarm value if there is an abnormality or failure in the main bearing. Under normal conditions, this residual value fluctuates within the specified normal range.

#### 3.2. Framework for the Proposed Model

The architecture of the proposed model for condition monitoring of the WT main bearing is shown in Figure 3. The architecture consists of six submodules, namely the input module, the temporal attention computation module, the parametric attention computation module, the dual attention merging module, the Bi-LSTM module, and the reconstruction module. The following subsections describe these modules in detail.

##### 3.2.1. Input Module

For a comprehensive description of the operating condition of the main bearing, the condition parameters of the main bearing itself and the associated condition parameters should be extracted. Note that the extracted parameters will be different for different operation phases because the control strategy is different for different phases. According to the analysis in Section 2, we obtain the input condition parameters for the maximum wind energy tracking phase as an input matrix  $X$ , which is defined in this form  $(m, w)$ , where  $m$  is the number of variables and  $w$  is the length of the window. In this study,  $m$  is set to 11, and  $w$  is set to 5, while in the constant power phase,  $m$  is set to 12, and  $w$  is also set to 5. For the convenience of subsequent calculations, we only discuss the BC phase here. Thus, the input vector at the  $i$ th moment can be described as  $X(i) = \{X_i^1, X_i^2, \dots, X_i^m\}$ , where  $i$  can take values in the range of 1, 2, 3, 4, and 5.



**Figure 3.** The framework of the proposed model DA-Bi-LSTM. ① Input module; ② parametric attention computation module; ③ temporal attention computation module; ④ dual attention merging module; ⑤ Bi-LSTM module; ⑥ reconstruction module.

### 3.2.2. Parametric Attention Computation Module

In the process of parameter feature extraction, to distinguish the importance of different condition parameters, and assign different weights to different condition parameters, the parametric attention computation module can calculate the weights of the condition parameters within each time step. For the given input vector  $X_t = \{X_t^1, X_t^2, \dots, X_t^m\}$ , we can use Equations (2) and (3) to compute parametric attention weight distribution values.

$$p_t = \sigma(W_t * X_t + d_t) \quad (2)$$

$$a_t^j = \exp(p_t^j) / \sum_{k=1}^m \exp(p_t^k). \quad (3)$$

where  $W_t$  represents the weight coefficient matrix, and  $d_t$  represents the bias term,  $\sigma$  represents the sigmoid function, and  $a_t^j$  is the normalized attention weight for the  $j$ th condition parameter, which is a calculation method of entropy weight, and the following is the same.

### 3.2.3. Temporal Attention Computation Module

The values of the condition parameters of the main bearing have a clear time series characteristic. The current value of the sequence is influenced by its historical moment value. However, the contribution or influence of different historical moment values is different. The temporal attention computation module can calculate the weights for each condition parameter over multiple consecutive time steps. For the given input vector  $X^i = \{X_1^i, X_2^i, \dots, X_w^i\}$ , we can use Equations (4) and (5) to compute their temporal attention weight distribution values.

$$e_t = \sigma(W^i * X^i + d^i) \quad (4)$$

$$b_t^i = \exp(e_t^i) / \sum_{k=1}^w \exp(e_t^k) \quad (5)$$

where  $W^i$  represents the weight coefficient matrix, and  $d^i$  represents the bias term.  $\sigma$  represents the sigmoid function, and  $b_t^i$  is the normalized attention weight for the  $i$ th condition parameter.

### 3.2.4. Dual Attention Merging Module

To merge and fuse the temporal attention feature value and the parameter attention feature value, the dual attention merging module first calculates the temporal attention and

the parametric attention by the element-wise product method separately, then, calculates the final input vector using the element-wise sum method by using Equation (6).

$$\tilde{X} = X_t * a_t^j + X^i * b_t^i \quad (6)$$

### 3.2.5. Bi-LSTM Module

The Bi-LSTM (Bi-directional Long Short-Term Memory) is a bi-directional recurrent neural network based on LSTM, which combines information from input data in both forward and backward directions, and is a variant of LSTM. The Bi-LSTM model has made remarkable achievements in speech processing, temporal data prediction, and text classification [25,31]. Figure 4 illustrates the architecture of a single LSTM cell/unit, where  $f_t$ ,  $i_t$ , and  $O_t$  are the forget, input, and output gates, respectively. The  $\tilde{X}_t$  represents input vector. The  $h_{t-1}$  represents the output of the previous unit at time  $t - 1$ , and  $h_t$  represents the output of the current time  $t$ . The  $C_{t-1}$  represents the cell state at the previous time  $t - 1$ , and  $C_t$  represents the cell state at the current time  $t$ . The  $\sigma$  and  $\tanh$  are activation functions. These parameters at the time  $t$  are updated by:

$$f_t = \sigma(W_f * h_{t-1} + W_f * \tilde{X}_t + d_f) \quad (7)$$

$$i_t = \sigma(W_i * h_{t-1} + W_i * \tilde{X}_t + d_i) \quad (8)$$

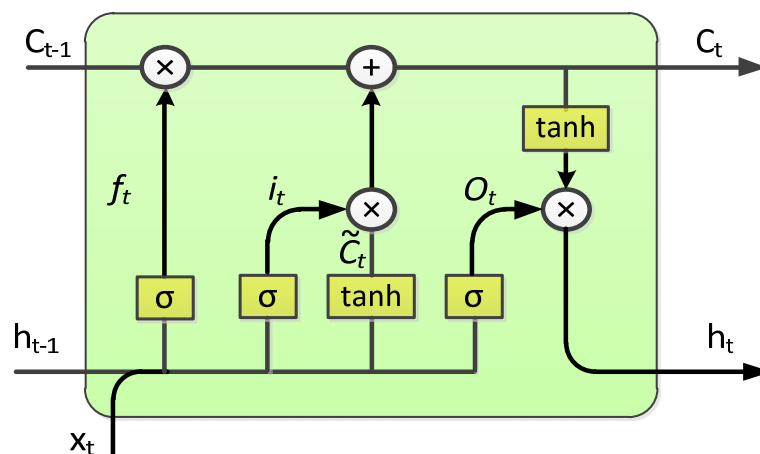
$$\tilde{C}_t = \tanh(W_c * h_{t-1} + W_c * \tilde{X}_t + d_c) \quad (9)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (10)$$

$$O_t = \sigma(W_o * h_{t-1} + W_o * \tilde{X}_t + d_o) \quad (11)$$

$$h_t = O_t * \tanh(C_t) \quad (12)$$

where  $W_f$ ,  $W_i$ ,  $W_c$ ,  $W_o$ ,  $b_f$ ,  $d_i$ ,  $d_c$ , and  $d_o$  are weight parameters and their biases, respectively. These parameter values are obtained by iterative iterations during the DA-Bi-LSTM model training.

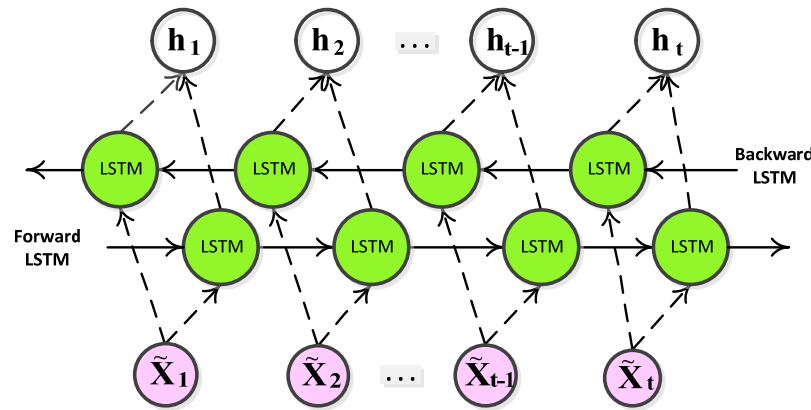


**Figure 4.** The internal architecture of a single LSTM unit/cell.

The Bi-LSTM model is created by adding an LSTM layer in the opposite direction of the information flow, which has the benefit of capturing the past and future temporal dependencies. Figure 5 shows the architecture of Bi-LSTM. For the  $k$ th time point,  $\tilde{X}_k$  is input to the forward and reverse LSTM, and then we get the forward and reverse hidden layer representations vectors  $\vec{h}_k$  and  $\overleftarrow{h}_k$ , respectively, and we use Equation (13)

to combine these two hidden layer representations to get the output  $h_k$  at  $k$ th time point. After the sequence  $\tilde{X}$  is input into the Bi-LSTM, we get the final output of the module, i.e.,  $h = \{h_1, h_2, \dots, h_w\}$ .

$$h_k = \vec{h}_k + \overleftarrow{h}_k \quad (13)$$



**Figure 5.** The architecture of Bi-LSTM.

### 3.2.6. Reconstruction Module

This module is a reconstruction of the extracted feature vectors. Specifically, the output vector of Bi-LSTM is nonlinearly transformed, and then the reconstructed vector  $Z$  of the same dimension as the input vector is obtained, which is calculated by Equation (14).

$$Z = \sigma(W_h * h + d_h) \quad (14)$$

where  $W_h$  represents the weight coefficient matrix, and  $d_h$  represents the bias term, and  $\sigma$  represents the sigmoid function.  $Z$  is the reconstruction vector.

### 3.3. Training Algorithm for the DA-Bi-LSTM Model

Based on the framework in Figure 3, an algorithm for the DA-Bi-LSTM model is proposed. The critical procedures are described as: (1) loading SCADA data in CSV files under different operating conditions in normal conditions; (2) performing data cleaning and resampling; (3) selecting the relevant condition parameter variables; (4) constructing train, verify and test datasets according to the given ratio; (5) building the proposed model DA-Bi-LSTM; (6) training the model according to the range given by the hyperparameters; (7) preserving the optimal model. A more detailed description of the training process is listed in the following pseudo-code Algorithm 1.

---

#### Algorithm 1: Pseudo-code for DA-Bi-LSTM

---

**Input:**  $D = \{(X(1), X(1)), (X(2), X(2)), \dots, (X(i), X(i)), \dots\}$ ,  $X(i) \in R^{m \times w}$

**Output:** The optimal DA-Bi-LSTM model

- 1: Read SCADA historical data in a normal state;
  - 2: Clean and resample data;
  - 3: Select condition parameter variables related to the main bearing;  
//generate training data set, test dataset, verify data set;
  - 4:  $X = \emptyset$ ,  $TRD = \emptyset$ ,  $VD = \emptyset$ ,  $TED = \emptyset$
  - 5: **while**  $i$  in  $(1, n-w)$  **do**:
  - 6:      $X = X \cup X(i)$
  - 7: **end while**
-

**Algorithm 1:** *Cont.*


---

```

8:      According to the ratio of 64%, 16%, and 20%, generate TRD, VD, TED
      //train DA-Bi-LSTM model
9:      Set the range of units  $s_l$ , hidden layers  $n_l$ , iterations  $e$ , learning rate
       $l_r$ ,  $n_e$  and batch size;
10:     Initialize parameters;
11:     while  $i \leq e$  or  $n_e$  = epoch numbers when the loss has not changed:
12:         Execute the temporal attention module;
13:         Execute the parametric attention module;
14:         Combine the values of the two attention modules;
15:         Reconstruct input data;
16:         Update parameters with the Adam algorithms to minimize
         reconstruction error in TRD;
17:         Verify the DA-Bi-LSTM model using the VD;
18:         Save the structure and parameters of the optimum model;
19:     end while
20:     Test the model using the test dataset TED;
21:     Return the optimal DA-Bi-LSTM model;

```

---

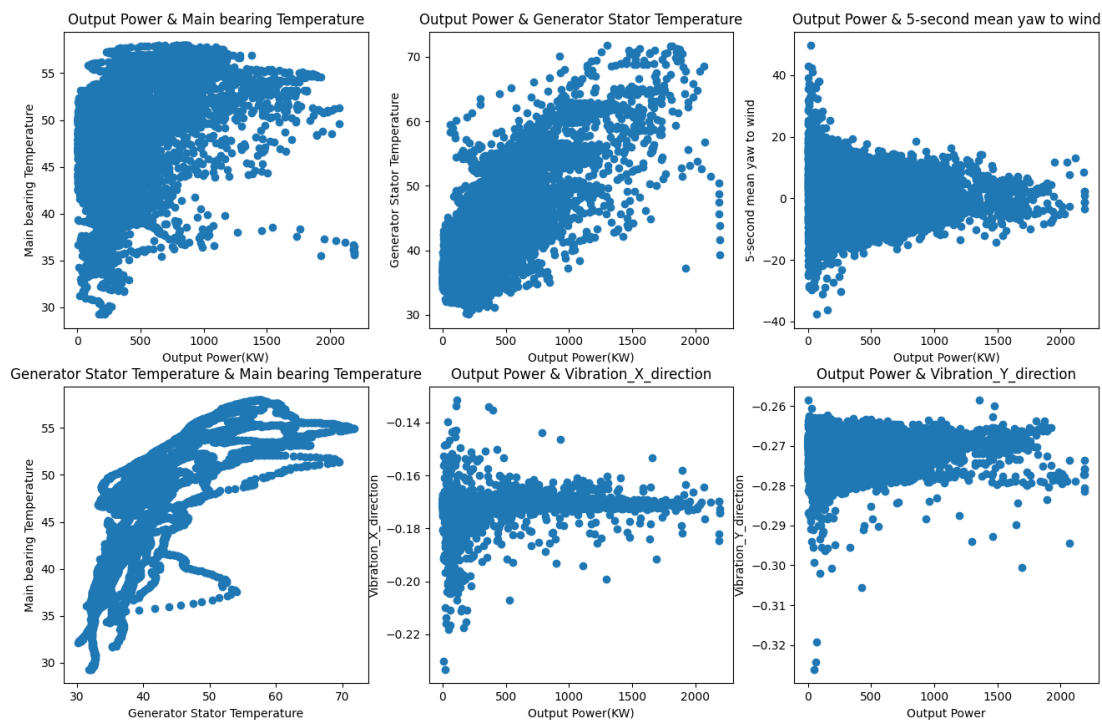
**4. Experiment Setup and Result Analysis**

In this section, we first probe the distribution characteristics of the relevant condition parameters of the WT main bearing based on the WT SCADA system. Second, we introduce methods for data cleaning, data resampling, and dataset construction. Third, we present the structure determination of the proposed model and the performance comparison with other models. Finally, we present the attention and interpretability analysis of the model. The experiment setup is as follows: Python 3.6 and the deep learning framework Keras API are used for experimental design.

**4.1. Distribution Characteristics of Partial Conditional Parameters**

In this paper, the main bearing of a 2 MW direct-drive WT is investigated. The wind turbine's cut-in, cut-out, and rated wind speeds are 3, 25, and 11 m/s, respectively. The wind turbine is equipped with a SCADA system that collects data at 1 Hz and stores it in a 10-min CSV file. We collected data for three months, from 5 September 2020, to 4 December 2020. According to the discussion in Section 2, we extracted the continuous data of the BC segment and the CD segment. After removing the outage data, we found that the amount of data in the CD segment was relatively small, which was related to the wind conditions at that time. The wind speed is not above rated wind speed most of the time. Therefore, the following experiments were carried out on the BC stage.

Figure 6 depicts a scatter plot of some parameters associated with the WT main bearing. As shown in Figure 6, there is a heavy nonlinear relationship among the parameters. Temperature, output power, and vibration are the core parameters that characterize the operating condition of the main bearing. The temperature increases with the increase of power, but in the same output power value range, the temperature varies widely. The vibration values in the two directions also keep the range reduced with the increase of the output power, but the performance is more obvious in the axial direction (*X*-axis direction), and weaker in the radial direction (*Y*-axis direction). The complex nonlinear relationship between these parameters requires powerful nonlinear fitting functions to realize. We designed such a complex function  $Z(t) = F(X(t))$  to solve this problem.



**Figure 6.** The scatter diagram of partial parameters related to WT main bearing.

#### 4.2. Data Cleansing and Resampling

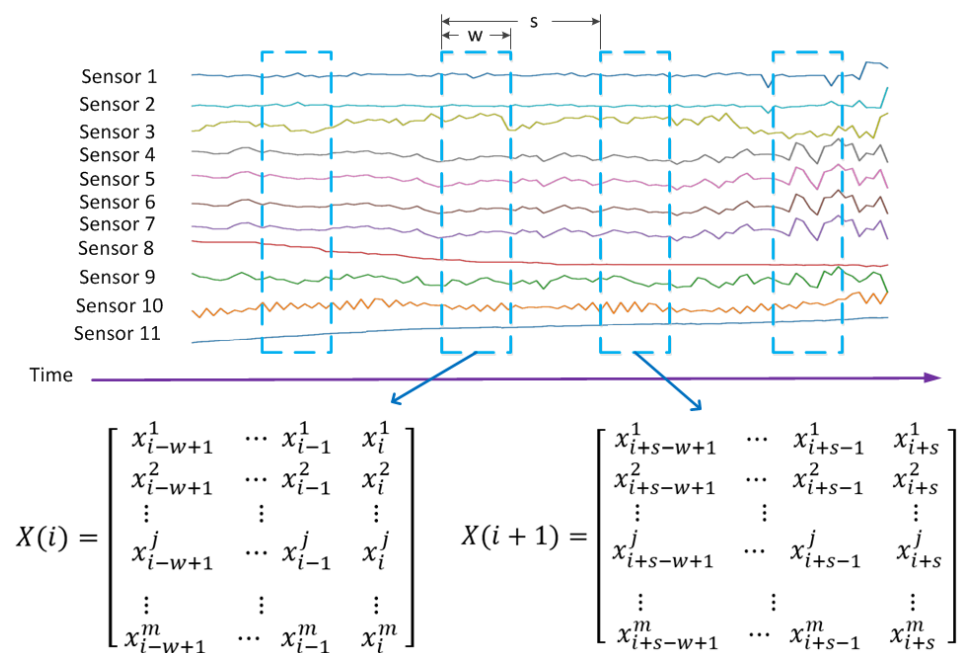
Data cleaning aims to deal with invalid data such as null values and outliers in the SCADA dataset due to transmission problems and WT downtime. These invalid data are usually empty, negative data, packet loss data, and unreasonable abnormal data. Additionally, according to existing research, data resampling techniques were used [32,33]. In this study, the resampling frequency is 1 min. The processing formula is as follows:

$$\left\{ \begin{array}{ll} \text{delete } x_i, & \text{for } x_i \in \text{hault data} \\ x_i = x_{i+1} \text{ or } x_{i-1}, & \text{for } x_i \in \text{packet loss data} \\ x_i = 0, & \text{for } x_i < 0 \text{ or } x_i \text{ is null} \\ x_i = \frac{1}{2}(x_{i-1} + x_{i+1}) & \text{for } x_i \geq 2x_{i-1} \text{ and } x_i \geq 2x_{i+1} \text{ or} \\ & \text{for } x_i \leq x_{i-1}/2 \text{ and } x_i \leq x_{i+1}/2 \\ \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, & \text{for data resampling} \end{array} \right. \quad (15)$$

where,  $x_i$  represents the data value collected by sensors, which is second-level data.  $n$  represents the number of data points.  $\bar{x}$  represents the resampled data, which is an average value.

#### 4.3. Dataset Construction

To continuously extract the observation data and generate the training samples and test samples required by the proposed model, we use the sliding window method to process the original condition parameter data. Figure 7 shows its specific construction process, where  $w$  represents the width of the sliding window,  $s$  represents the sliding step size,  $i$  represents a certain time point, and  $m$  represents the number of condition parameters representing the main bearing. For each input vector with window length  $w$  and input dimension  $m$ , the input vector is denoted as  $X$ .



**Figure 7.** Sliding window procedure for dataset construction.

#### 4.4. Evaluation Metrics

In this paper, three different evaluation metrics are chosen to evaluate the performance of the proposed model and competitor models. The evaluation metrics are the mean absolute error (MAE), the mean absolute error percentage (MAPE), and the root mean square error (RMSE). Their calculation expressions are defined as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |Z_i - X_i| \quad (16)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{Z_i - X_i}{X_i} \right| \quad (17)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Z_i - X_i)^2} \quad (18)$$

where  $Z_i$  is the reconstructed value at the  $i$ th time.  $X_i$  is the measured value at the  $i$ th time.  $N$  is the number of sample points.

In this study, we collect 495,900 sample data. After data cleaning and resampling, the total number of samples is 8265, of which 6612 samples are used for training and 1653 samples are used for testing. Partial samples are shown in Table 1.

**Table 1.** Partial samples for training and testing.

Main Bearing Temperature	Generator Stator Temperature	Wind Speed	Ambient Temperature	Rotor Speed	Power Output	Generator Operation Frequency	Generator Torque	5-s Mean Yaw to Wind	Vibration in X Direction	Vibration in Y Direction
38.6058	52.9444	8.9750	24.7250	14.8435	1954.3667	15.2940	1220.3000	11.6267	−0.1697	−0.2777
38.8133	55.0144	8.5450	24.6900	14.2665	1732.3333	14.7403	1120.2333	6.6992	−0.1698	−0.2767
39.0233	53.3061	7.2950	24.5283	12.3272	973.0333	12.7153	727.9333	8.4467	−0.1688	−0.2768
39.2017	54.2426	6.7783	24.5000	11.4243	688.2500	11.7608	555.4583	7.2292	−0.1690	−0.2778

#### 4.5. Determination of the DA-Bi-LSTM Model

Neural networks and deep learning models have certain randomness during training, such as weight initialization randomness, regularization randomness, and optimization

randomness. Therefore, for the same data given, the output of the model has a certain difference. A common approach is to run the network multiple times, then some statistical methods are used to summarize the performance of the model, and finally, the best model is selected [34,35]. During the training of DA-Bi-LSTM, many hyperparameters need to be determined. Since the structure of the model designed in this paper is relatively simple, it is an auto-encoding reconstruction model, so only the learning rate and batch size need to be determined here. This paper uses the grid search method to run the model multiple times. The experimental results of some evaluation indicator values, such as MAE, MAPE, and RMSE, are shown in Table 2. It should be pointed out that the time required for the model to run is also an important evaluation indicator, and its value is given accordingly.

**Table 2.** Performance indexes of different hyperparameters in DA-Bi-LSTM.

Batch Size	Learning Rate	MAE	MAPE	RMSE	Time (min)
4	0.001	1.304305 ± 0.836807	0.052495 ± 0.397203	3.519565 ± 2.322510	{8,9}
	0.002	0.864058 ± 0.596909	0.056605 ± 0.512678	2.319833 ± 1.678099	{7,11}
	0.003	0.856221 ± 0.557203	0.061018 ± 0.570364	2.260499 ± 1.569071	{8,14}
	0.004	0.785344 ± 0.524116	0.049320 ± 0.450459	2.084963 ± 1.470836	{8,16}
	0.005	0.899751 ± 0.576920	0.056278 ± 0.548135	2.352721 ± 1.596467	{11,17}
	0.006	0.831057 ± 0.597686	0.072474 ± 0.758543	2.177138 ± 1.677285	{9,20}
	0.007	0.967146 ± 0.620336	0.073080 ± 0.715136	2.515065 ± 1.719582	{12,27}
8	0.001	1.896889 ± 1.116898	0.061770 ± 0.409709	5.176646 ± 3.097641	{4,5}
	0.002	0.968291 ± 0.659174	0.048102 ± 0.392758	2.594862 ± 1.822798	{6,7}
	0.003	0.926366 ± 0.604838	0.047017 ± 0.399360	2.510316 ± 1.702183	{5,8}
	0.004	0.841708 ± 0.542544	0.053985 ± 0.456547	2.239096 ± 1.516001	{6,9}
	<b>0.005</b>	<b>0.823974 ± 0.543782</b>	<b>0.056058 ± 0.531651</b>	<b>2.197976 ± 1.530926</b>	<b>{7,9}</b>
	0.006	0.882462 ± 0.619273	0.060000 ± 0.582828	2.325437 ± 1.722824	{6,10}
	0.007	0.904848 ± 0.601930	0.056594 ± 0.542598	2.383833 ± 1.671474	{9,12}
16	0.001	2.556960 ± 1.320332	0.073033 ± 0.438661	6.974637 ± 3.682909	{2,3}
	0.002	1.650315 ± 1.015608	0.062074 ± 0.481981	4.478816 ± 2.802205	{3,4}
	0.003	1.148293 ± 0.758176	0.053121 ± 0.413128	3.114942 ± 2.115525	{4,5}
	0.004	0.936918 ± 0.653807	0.052956 ± 0.438176	2.519213 ± 1.825722	{4,6}
	0.005	0.892945 ± 0.605100	0.044812 ± 0.351199	2.387302 ± 1.682407	{5,12}
	0.006	0.982730 ± 0.651114	0.053701 ± 0.453724	2.619971 ± 1.799305	{5,9}
	0.007	0.975500 ± 0.634203	0.060583 ± 0.559166	2.587093 ± 1.773496	{6,8}
32	0.001	3.017067 ± 1.558755	0.127041 ± 1.102463	8.207864 ± 4.373313	{2,2}
	0.002	2.275105 ± 1.203766	0.071594 ± 0.481565	6.197045 ± 3.352271	{2,3}
	0.003	2.002779 ± 1.120925	0.076850 ± 0.608935	5.439411 ± 3.070086	{2,3}
	0.004	1.431755 ± 0.858475	0.064426 ± 0.534677	3.830619 ± 2.366009	{3,6}
	0.005	1.276032 ± 0.750272	0.061427 ± 0.506284	3.417497 ± 2.065064	{4,6}
	0.006	1.303502 ± 0.728047	0.064615 ± 0.546178	3.471839 ± 2.030046	{5,16}
	0.007	1.200432 ± 0.745935	0.086588 ± 0.844691	3.187429 ± 2.047071	{9,26}
64	0.001	3.927755 ± 2.075948	0.251386 ± 2.178752	10.464730 ± 5.834217	{1,2}
	0.002	2.925926 ± 1.526210	0.106664 ± 0.925961	7.986035 ± 4.261907	{1,2}
	0.003	2.606756 ± 1.318098	0.086379 ± 0.693012	7.130800 ± 3.717076	{2,2}
	0.004	2.284217 ± 1.219294	0.075765 ± 0.571431	6.228751 ± 3.406993	{2,3}
	0.005	2.094401 ± 1.116843	0.076662 ± 0.578012	5.678073 ± 3.089247	{2,3}
	0.006	2.180400 ± 1.137641	0.087591 ± 0.770607	5.922371 ± 3.188089	{2,4}
	0.007	1.861028 ± 0.983870	0.088477 ± 0.795060	5.015731 ± 2.743637	{4,4}

From Table 2, as the batch size increases, the metrics MAE, MAPE, and RMSE values on the test dataset tend to increase. The average proportions of batch-by-batch increase are, in terms of MAE: 11.32%, 26.21%, 36.78%, and 42.97%. In terms of MAPE: −8.96%, 4.37%, 38.04%, and 39.89%; in terms of RMSE: 12.76%, 27.04%, 36.75%, and 43.48%. The training time of the model shows a gradual downward trend. The proposed model is a reconstruction model, which is relatively simple, so it does not mean that the larger the batch size, the better, or the smaller the better. The optimal value of MAPE is reflected in

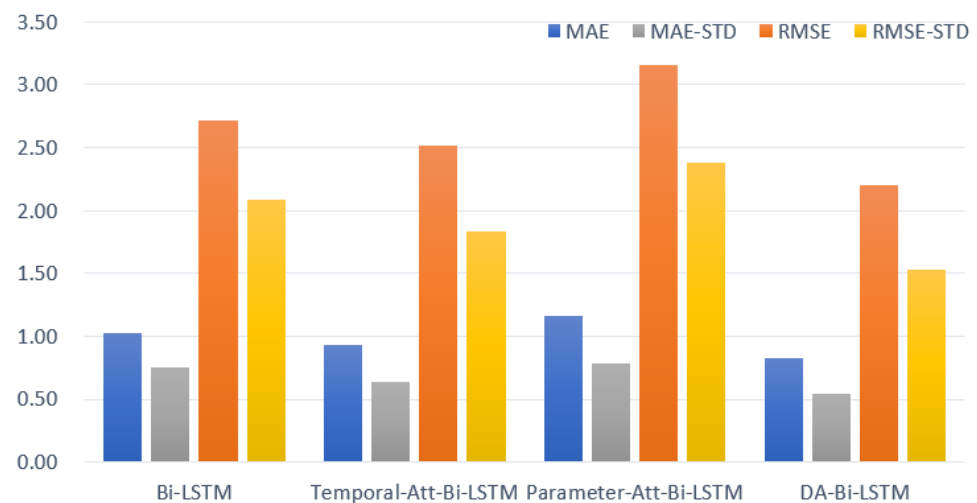
the batch size of 8, not 4. In addition, although the training time of the model shows a gradual downward trend, according to the relevant selection principles, within the time range commonly used for data collection in the industry, the smaller the three evaluation metrics values, the better [8]. Therefore, our final choice is that the batch size is set at 8, and the learning rate is set at 0.005.

#### 4.6. Performance Comparison

To compare the performance of the proposed model, three baseline models are used, such as the Bi-LSTM model, the temporal attention-based Bi-LSTM (Temporal-Att-Bi-LSTM) model, and the parameter attention-based Bi-LSTM (Parameter-Att-Bi-LSTM). All models use the same input data format, but each input data is handled differently within each model. Among them, the Bi-LSTM model is not added with an attention module, and it only considers the before and after information of the time series. The Temporal-Att-Bi-LSTM model takes into account both the temporal attention information and the before and after information of the time series. The Parameter-Att-Bi-LSTM model considers the parameter attention information and the before and after information of the time series. The comparison results are listed in Table 3 and Figure 8.

**Table 3.** Performance indexes of different models (best values displayed in boldface).

Models	MAE	MAPE	RMSE
Bi-LSTM	1.027185 $\pm$ 0.752840	0.057300 $\pm$ 0.533444	2.710831 $\pm$ 2.086212
Temporal-Att-Bi-LSTM	0.934075 $\pm$ 0.642100	<b>0.054747 <math>\pm</math> 0.508762</b>	2.513496 $\pm$ 1.836888
Parameter-Att-Bi-LSTM	1.157930 $\pm$ 0.784342	0.055885 $\pm$ 0.417240	3.156833 $\pm$ 2.374727
<b>DA-Bi-LSTM</b>	<b>0.823974 <math>\pm</math> 0.543782</b>	0.056058 $\pm$ 0.531651	<b>2.197976 <math>\pm</math> 1.530926</b>



**Figure 8.** Performance comparison of different models with standard deviation distribution.

In Table 3 and Figure 8, the proposed DA-Bi-LSTM model shows the best performance, followed by the Temporal-Att-Bi-LSTM model and Bi-LSTM, and the worst model is the Parameter-Att-Bi-LSTM. The MAE value and RMSE value of the proposed model, and their corresponding standard deviations, are the smallest. Specifically, compared with Bi-LSTM, Temporal-Att-Bi-LSTM, and Parameter-Att-Bi-LSTM, the MAE value of the proposed model decreases by about 19.78%, 11.79%, and 28.84%, respectively. The standard deviation of the MAE values of the proposed model is lower than 22.77%, 15.31%, and 30.67%, respectively. Compared with Bi-LSTM, Temporal-Att-Bi-LSTM, and Parameter-Att-Bi-LSTM, the RMSE values of the proposed model drop by about 18.92%, 12.55%, and 30.37%, respectively. The standard deviation of the RMSE value of the proposed model is lower than 26.62%, 16.66%, and 35.53%, respectively. The MAPE values of all models are around 0.055, and the

discrimination is not obvious because MAPE is the mean of the relative percentage of the reconstruction error to the measured value.

In Figure 8, the Parameter-Att-Bi-LSTM model that only considers parameter attention is worse than the Bi-LSTM model because the model assigns different weights to different parameters before extracting the information before and after the time series, ignoring weak timing information for unimportant parameters. So, its effect is worse than the Bi-LSTM model. The Temporal-Att-Bi-LSTM model works better than Bi-LSTM because our data itself is time series data, and the time series attention module further highlights the importance of the data at different times. The DA-Bi-LSTM model performs the best because it integrates time-series attention information and parameter attention information, the time-series information of important parameters, and the weak time-series information of non-important parameters, which are all included. Specifically, the DA-Bi-LSTM model emphasizes the extraction of the importance of different input parameters and the extraction of parameter time series features. Then, the extracted information from the two aspects is further fused, and the time series features of the composite information are finally extracted and learned. Therefore, its information representation and extraction ability are the strongest, and the effect is also the best.

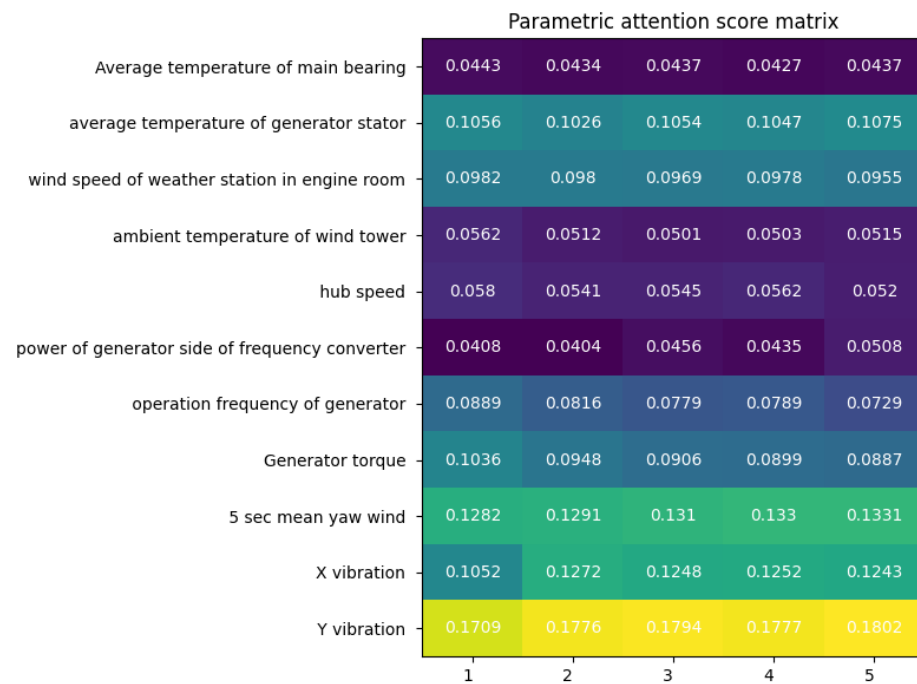
#### 4.7. Attention and Interpretability Analysis

The attention mechanism module in the proposed model can give the dynamic attention score of each parameter of each input sample to the operating condition of the main bearing in real time and can reflect the contribution of different time points and different parameter variables to the main bearing in real time. In a heatmap of attention scores, the darker the color, the higher the attention value, and vice versa.

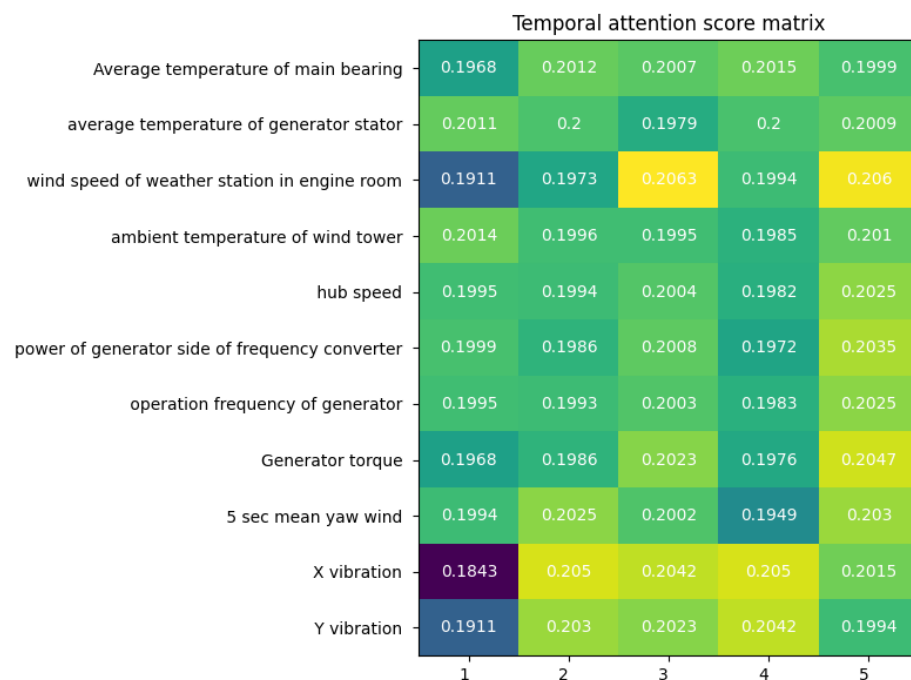
In the parametric attention score calculation module, we extract the parametric attention score value of a sample, as shown in Figure 9. Where the horizontal coordinate, X-axis, represents different time points, the vertical coordinate, Y-axis, represents 11 condition parameters, and the numbers in the figure represent the contribution of parameter attention. The attention score of each condition parameter variable is different. Vibration in two directions and “5 s mean yaw wind” are important parameters, especially in the Y direction. However, the degree of influence for some parameters such as temperature is not as obvious. Specifically, the attention score of “Y vibration value” is the largest, accounting for about 17.72% on average, and it is maintained, followed by “5 s mean yaw wind”, “X vibration”, “average temperature of generator stator”, and “wind speed”, which accounted for 13.09%, 12.13%, 10.52%, and 9.73%, respectively. The lowest attention scores are “average temperature of the main bearing” and “power output”, which are 4.36% and 4.42%, respectively. The parameter attention calculation module can select some parameter variables and can assign different attention scores to these parameters according to the speed of parameter change and the constraint relationship between parameters in the actual process of monitoring.

In the temporal attention score calculation module, we also extract the temporal attention score value of a sample, as shown in Figure 10, where the horizontal coordinate represents different time points, the vertical coordinate represents 11 condition parameters, and the numbers in the figure represent the contribution of the temporal attention score. Relative to time  $t$ , most of the temporal attention scores are highest at time  $t - 1$  and time  $t - 3$ , accounting for about 20.23% and 20.14%, respectively. The most obvious one is “wind speed”, accounting for 20.6%, followed by “generator torque”, accounting for 20.47%, followed by “power output” and “5-s mean yaw to wind”, accounting for 20.35% and 20.3%, respectively. The temperature parameter is a gradually changing quantity. However, the time-series attention score distribution of temperature parameters still has differences in the main bearing and related components parameters. The attention score distribution of the temperature of the main bearing itself is relatively balanced at about 20%, at times  $t - 2$  and  $t - 4$ , the proportion is larger, showing a high in the middle and low at both ends. However, there is no obvious regularity in the changes of the generator

stator temperature parameters and the external environment temperature parameters, and the standard deviation of its attention score is the lowest, 0.0011. The standard deviation of attention scores for the vibration parameter of “vibration in X direction” was the largest, followed by the wind speed parameter, which was 0.008 and 0.0057, respectively. This also shows that the wind speed fluctuates greatly, and the X-axis fluctuates greatly. The temporal attention calculation module can select different time points and can assign different attention scores to different historical time points, emphasizing the expression of key information at important time points.



**Figure 9.** A heatmap of parametric attention scores.



**Figure 10.** A heatmap of temporal attention scores.

## 5. Conclusions

This work presents a novel method of condition monitoring for WT main bearing by using the dual attention mechanism and Bi-LSTM. The effectiveness of the proposed method is verified by using the WT SCADA data. The research results show that the proposed model for the different working conditions can further exactly explore the spatiotemporal correlation between the condition parameters of the WT main bearing itself and its associated condition parameters, strengthen the key information in the input data, weaken the secondary information, and improve the precision of the main bearing condition monitoring and the interpretability of the monitoring model.

In the future, we will further expand the verification and application of the proposed model in the constant power stage, and also include the modeling of the model on other key components or subsystems, such as generators, hubs, pitch systems and yaw systems, etc. At the same time, we will also consider field testing and deployment of the proposed method for better monitoring and earlier warning of WT main bearing failures.

**Author Contributions:** The research in this paper was the result of the joint efforts of all authors. X.X.: methodology, software, validation, writing—original draft preparation; J.L.: conceptualization, supervision, writing—reviewing and editing, funding acquisition; D.L.: conceptualization, supervision, writing—reviewing and editing, funding acquisition; Y.T.: conceptualization, supervision, writing—reviewing; S.Q.: validation, visualization; F.Z.: validation. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Key R&D Program of China, grant number 2020YFB1-707602, the National Natural Science Foundation of China, grant number 51475160, and the Key Research and Development Project of Hunan Province, China, grant number 2018WK2022.

**Data Availability Statement:** The datasets used in the current study are available from the corresponding author upon reasonable request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Brodny, J.; Tutak, M.; Bindzár, P. Assessing the level of renewable energy development in the European union member states. A 10-year perspective. *Energies* **2021**, *14*, 3765. [\[CrossRef\]](#)
2. Li, J.; Ho, M.S.; Xie, C.; Stern, N. China's flexibility challenge in achieving carbon neutrality by 2060. *Renew. Sustain. Energy Rev.* **2022**, *158*, 112112. [\[CrossRef\]](#)
3. Yang, Q.; Liu, G.; Bao, Y.; Chen, Q. Fault Detection of Wind Turbine Generator Bearing Using Attention-Based Neural Networks and Voting-Based Strategy. *IEEE/ASME Trans. Mechatron.* **2021**, *27*, 3008–3018. [\[CrossRef\]](#)
4. Xiang, L.; Yang, X.; Hu, A.; Su, H.; Wang, P. Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks. *Appl. Energy* **2022**, *305*, 117925. [\[CrossRef\]](#)
5. Civera, M.; Surace, C. Non-Destructive Techniques for the Condition and Structural Health Monitoring of Wind Turbines: A Literature Review of the Last 20 Years. *Sensors* **2022**, *22*, 1627. [\[CrossRef\]](#)
6. Hart, E.; Clarke, B.; Nicholas, G.; Kazemi Amiri, A.; Stirling, J.; Carroll, J.; Dwyer-Joyce, R.; McDonald, A.; Long, H. A review of wind turbine main bearings: Design, operation, modelling, damage mechanisms and fault detection. *Wind Energy Sci.* **2020**, *5*, 105–124. [\[CrossRef\]](#)
7. Gbashi, S.M.; Madushele, N.; Olatunji, O.O.; Adedeii, P.A.; Jen, T.-C. Wind Turbine Main Bearing: A Mini Review of Its Failure Modes and Condition Monitoring Techniques. In Proceedings of the 2022 IEEE 13th International Conference on Mechanical and Intelligent Manufacturing Technologies (ICMIMT), Cape Town, South Africa, 25–27 May 2022; pp. 127–134.
8. Hart, E. Developing a systematic approach to the analysis of time-varying main bearing loads for wind turbines. *Wind Energy* **2020**, *23*, 2150–2165. [\[CrossRef\]](#)
9. Nejad, A.R.; Keller, J.; Guo, Y.; Sheng, S.; Polinder, H.; Watson, S.; Dong, J.; Qin, Z.; Ebrahimi, A.; Schelenz, R. Wind turbine drivetrains: State-of-the-art technologies and future development trends. *Wind Energy Sci.* **2022**, *7*, 387–411. [\[CrossRef\]](#)
10. Hart, E.; Turnbull, A.; Feuchtwang, J.; McMillan, D.; Golysheva, E.; Elliott, R. Wind turbine main-bearing loading and wind field characteristics. *Wind Energy* **2019**, *22*, 1534–1547. [\[CrossRef\]](#)
11. Tutivén, C.; Vidal, Y.; Insuasty, A.; Campoverde-Vilela, L.; Achicanoy, W. Early Fault Diagnosis Strategy for WT Main Bearings Based on SCADA Data and One-Class SVM. *Energies* **2022**, *15*, 4381. [\[CrossRef\]](#)
12. Encalada-Dávila, Á.; Puruncas, B.; Tutivén, C.; Vidal, Y. Wind Turbine Main Bearing Fault Prognosis Based Solely on SCADA Data. *Sensors* **2021**, *21*, 2228. [\[CrossRef\]](#) [\[PubMed\]](#)

13. Natili, F.; Daga, A.P.; Castellani, F.; Garibaldi, L. Multi-Scale Wind Turbine Bearings Supervision Techniques Using Industrial SCADA and Vibration Data. *Appl. Sci.* **2021**, *11*, 6785. [\[CrossRef\]](#)
14. Beretta, M.; Julian, A.; Sepulveda, J.; Cusidó, J.; Porro, O. An Ensemble Learning Solution for Predictive Maintenance of Wind Turbines Main Bearing. *Sensors* **2021**, *21*, 1512. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Guo, P.; Wang, Z. Wind turbine spindle state monitoring based on Gaussian process regression and double moving window residual processing. *Electr. Power Autom. Equip.* **2018**, *38*, 34–40. [\[CrossRef\]](#)
16. Beretta, M.; Vidal, Y.; Sepulveda, J.; Porro, O.; Cusidó, J. Improved ensemble learning for wind turbine main bearing fault diagnosis. *Appl. Sci.* **2021**, *11*, 7523. [\[CrossRef\]](#)
17. Zheng, Y.; Wei, J.; Zhu, K.; Bo, D. Fault Monitoring Method of Wind Turbine Main Bearing. *J. Vib. Meas. Diagn.* **2021**, *41*, 341–347 + 415. [\[CrossRef\]](#)
18. Liu, C.; Duan, B.; Xiaodan, Z. An abnormal identification method for the main bearing of wind turbines based on BPNN-NCT. *Power Syst. Prot. Control* **2022**, *50*, 114–122. [\[CrossRef\]](#)
19. Zhang, Y.; Zheng, H.; Liu, J.; Zhao, J.; Sun, P. An anomaly identification model for wind turbine state parameters. *J. Clean. Prod.* **2018**, *195*, 1214–1227. [\[CrossRef\]](#)
20. Xu, Z.; Yang, P.; Zhao, Z.; Lai, L.L.; Wang, X. Fault diagnosis approach of main drive chain in wind turbine based on data fusion. *Appl. Sci.* **2021**, *11*, 5804. [\[CrossRef\]](#)
21. Tutiv'en, C.; Benalcázar-Parra, C.; D'ávila Escuela, A.E.; Vidal, Y.; Puruncaias, B.; Fajardo, M. Wind Turbine Main Bearing Condition Monitoring via Convolutional Autoencoder Neural Networks. In Proceedings of the 2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Mauritius, 7–8 October 2021; pp. 1–6.
22. Tonks, O.; Wang, Q. The detection of wind turbine shaft misalignment using temperature monitoring. *J. Manuf. Sci. Technol.* **2017**, *17*, 71–79. [\[CrossRef\]](#)
23. Yucesan, Y.A.; Viana, F.A. A hybrid physics-informed neural network for main bearing fatigue prognosis under grease quality variation. *Mech. Syst. Signal Process.* **2022**, *171*, 108875. [\[CrossRef\]](#)
24. Rezamand, M.; Kordestani, M.; Orchard, M.E.; Cariveau, R.; Ting, D.S.-K.; Saif, M. Improved remaining useful life estimation of wind turbine drivetrain bearings under varying operating conditions. *IEEE Trans. Ind. Inform.* **2020**, *17*, 1742–1752. [\[CrossRef\]](#)
25. Zheng, Y.-F.; Gao, Z.-H.; Shen, J.; Zhai, X.-S. Optimising Automatic Text Classification Approach in Adaptive Online Collaborative Discussion-A perspective of Attention Mechanism-Based Bi-LSTM. *IEEE Trans. Learn. Technol.* **2022**, 1–14. [\[CrossRef\]](#)
26. Yin, X.; Wu, G.; Wei, J.; Shen, Y.; Qi, H.; Yin, B. Multi-stage attention spatial-temporal graph networks for traffic prediction. *Neurocomputing* **2021**, *428*, 42–53. [\[CrossRef\]](#)
27. Hu, J.; Zheng, W. Multistage attention network for multivariate time series prediction. *Neurocomputing* **2020**, *383*, 122–137. [\[CrossRef\]](#)
28. Su, X.; Shan, Y.; Li, C.; Mi, Y.; Fu, Y.; Dong, Z. Spatial-temporal attention and GRU based interpretable condition monitoring of offshore wind turbine gearboxes. *IET Renew. Power Gener.* **2022**, *16*, 402–415. [\[CrossRef\]](#)
29. Xiang, L.; Wang, P.; Yang, X.; Hu, A.; Su, H. Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism. *Measurement* **2021**, *175*, 109094. [\[CrossRef\]](#)
30. Xiao, X.; Liu, J.; Liu, D.; Tang, Y.; Dai, J.; Zhang, F. SSAE-MLP: Stacked sparse autoencoders-based multi-layer perceptron for main bearing temperature prediction of large-scale wind turbines. *Concurr. Comput. Pract. Exp.* **2021**, *33*, e6315. [\[CrossRef\]](#)
31. Zhang, H.; Huang, H.; Han, H. A Novel Heterogeneous Parallel Convolution Bi-LSTM for Speech Emotion Recognition. *Appl. Sci.* **2021**, *11*, 9897. [\[CrossRef\]](#)
32. Cambron, P.; Masson, C.; Tahan, A.; Pelletier, F. Control chart monitoring of wind turbine generators using the statistical inertia of a wind farm average. *Renew. Energy* **2018**, *116*, 88–98. [\[CrossRef\]](#)
33. Xiao, X.; Liu, J.; Liu, D.; Tang, Y.; Zhang, F. Condition Monitoring of Wind Turbine Main Bearing Based on Multivariate Time Series Forecasting. *Energies* **2022**, *15*, 1951. [\[CrossRef\]](#)
34. Pei, S.; Qin, H.; Yao, L.; Liu, Y.; Wang, C.; Zhou, J. Multi-step ahead short-term load forecasting using hybrid feature selection and improved long short-term memory network. *Energies* **2020**, *13*, 4121. [\[CrossRef\]](#)
35. Hossain, M.A.; Chakraborty, R.K.; Elsayah, S.; Ryan, M.J. Very short-term forecasting of wind power generation using hybrid deep learning model. *J. Clean. Prod.* **2021**, *296*, 126564. [\[CrossRef\]](#)