

Article

Pre-trained 1DCNN-BiLSTM Hybrid Network for Temperature Prediction of Wind Turbine Gearboxes

Kejia Zhuang¹, Cong Ma¹, Heung-Fai Lam², Li Zou¹ and Jun Hu^{3,4,*} ¹ School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430062, China; zhuangkj@whut.edu.cn (K.Z.); macong@whut.edu.cn (C.M.); zouli@whut.edu.cn (L.Z.)² Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong; paullam@cityu.edu.hk³ School of Civil Engineering and Architecture, Wuhan University of Technology, Wuhan 430062, China⁴ Sanya Science and Education Innovation Park, Wuhan University of Technology, Sanya 572025, China

* Correspondence: junhu22@whut.edu.cn

Abstract: The safety and stability of a wind turbine is determined by the health condition of its gearbox. The temperature variation, compared with other characteristics of the gearbox, can directly and sensitively reflect its health conditions. However, the existing deep learning models (including the single model and the hybrid model) have their limitations in dealing with nonlinear and complex temperature data, making it challenging to achieve high-precision prediction results. In order to tackle this issue, this paper introduces a novel two-phase deep learning network for predicting the temperature of wind turbine gearboxes. In the first phase, a one-dimensional convolutional neural network (1DCNN) and a bidirectional long short-term memory (BiLSTM) network are separately trained using the same dataset. The two pre-trained networks are combined and fine-tuned to form the 1DCNN-BiLSTM model for the accurate prediction of gearbox temperatures in the second phase. The proposed model was trained and validated by measured datasets from gearboxes from an existing wind farm. The effectiveness of the model presented was showcased through a comparative analysis with five traditional models, and the result has clearly shown that the proposed model has a great improvement in its prediction accuracy.



Citation: Zhuang, K.; Ma, C.; Lam, H.-F.; Zou, L.; Hu, J. Pre-trained 1DCNN-BiLSTM Hybrid Network for Temperature Prediction of Wind Turbine Gearboxes. *Processes* **2023**, *11*, 3324. <https://doi.org/10.3390/pr11123324>

Academic Editors: Liming Dai and Jialin Tian

Received: 23 October 2023

Revised: 25 November 2023

Accepted: 26 November 2023

Published: 29 November 2023



Copyright: © 2023 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/>).

Keywords: temperature prediction; wind turbine gearbox; one-dimensional convolutional neural network; bidirectional long short-term memory network

1. Introduction

With the further intensification of global warming, clean energy has experienced exponential growth for low-carbon development [1–3]. The rapid growth of the wind power sector poses maintenance challenges for critical components of wind turbines such as the gearbox, blade, and brake rotors [4,5]. Among them, the gearbox, as the fault-prone component, determines the stable operation of the wind turbine [6,7]. Once the wind turbine's gearbox breaks down, a huge economic loss will happen [5,8]. Thus, monitoring the operational condition of the gearbox is crucial for enhancing the wind turbine's reliability and safety. In the actual application, the temperature variation is a suitable index reflecting the health condition of the wind turbine gearbox [9]. For example, operating under an environment with a high temperature for a long time will directly lead to the degradation of the viscosity of lubricating oil and increase the friction of the gear mesh [10,11]. As a result, the temperature will continue to rise, which will result in a vicious cycle, eventually causing damage to the gearbox. Therefore, it is essential to monitor and predict the operation temperature of the wind turbine gearbox [12]. Although some temperature prediction methods have already been applied in actual applications, the prediction accuracy of the results is not satisfying and can hardly extract the characteristics from the complex nonlinear signals. Therefore, this paper introduces an innovative deep learning

network to address the aforementioned challenge. The proposed model outperforms the five classical comparison models in terms of prediction accuracy, meeting the demands for high prediction accuracy in actual engineering applications.

In recent decades, a variety of temperature prediction methods have been proposed and can be categorized into two groups, namely, the traditional mathematical model-based methods and deep learning-based methods [13,14]. Traditional approaches have significantly contributed to the advancement of temperature forecasting. For example, Li and coworkers [15] introduced a data processing technique that involves outlier identification, missing value imputation, and random error reduction. This method was employed to enhance the accuracy of temperature prediction in large-scale concrete applications. To handle locally stationary time series data, Das and coworkers [16] proposed a model-free temperature prediction method. In the literature, One-step-ahead point prediction models and prediction intervals were developed with the aim of improving prediction accuracy compared to the widely utilized RAMPFIT algorithm in the context of locally stationary long time series data. To deal with predictions of the thermal characteristics of CPU, Wang and coworkers have presented an enhanced linear regression method [17]. In the proposed method, the conventional linear regression model was enhanced by incorporating the correlation between time series data and the model's autocorrelation. Although these methods have achieved great success at temperature prediction, disadvantages still limit their further application. Basically, these methods were developed mainly based on prior knowledge and expert experience. Therefore, it is challenging to precisely characterize nonlinear data, such as temperature data, and extract deep abstract features [18].

On the contrary, machine learning-based methods can automatically mine the features of the data in a proper manner when dealing with nonlinear data [18,19]. Typical machine learning methods include convolutional neural networks (CNN), recurrent neural networks (RNN), long short-term memory (LSTM) networks, and so on [20]. Among these networks, LSTM is designed with a specific focus on addressing prediction challenges within time series data. The three-gate structure can handle the time series data well and efficiently avoid gradient disappearance and gradient explosion [21]. Considering temporal-spatial correlation in traffic systems, Zhao and coworkers [22] developed an LSTM-based traffic forecasting method for short-term traffic prediction. Vlachas and coworkers have introduced a data-driven forecasting method based on LSTM for high-dimensional chaotic systems [23]; the results indicated that the LSTM neural network is effective in nonlinear data processing. Jia and coworkers [24] developed an LSTM model for the long-term and seasonal temperature prediction of the sea, considering the impact of the input length. The experimental results indicate that selecting an appropriate input length can enhance the performance of the model.

Despite the advantages of the LSTM model over the classical CNN and RNN models in processing time series data, it falls short in its ability to extract periodic features, potentially impacting the prediction accuracy. To overcome this problem, the bidirectional long short-term memory (BiLSTM) network is proposed. The BiLSTM model has the advantage in nonlinear temporal signal processing with its bidirectional structure, combining the forward sequences and the reverse sequences of LSTM; thus, it can capture more valuable features from both the forward direction and the reverse direction. Therefore, the prediction accuracy can be significantly improved. Wu and coworkers [25] proposed a tool wear prediction model based on the BiLSTM model. Based on the dataset processed by singular value decomposition (SVD), the BiLSTM neural network can extract the periodic features of SVD features and achieve a higher prediction accuracy than the comparison models. In actual application, the performance of the BiLSTM model is closely tied to the expertise of professionals in the segmentation of time series data. Thus, Jiang and coworkers [26] proposed a method combining the elitist preservation genetic algorithm (EGA) with the BiLSTM model to predict battery temperature. The EGA method is employed to derive an optimized data segmentation strategy, thus enabling the BiLSTM neural network to achieve enhanced prediction accuracy. Farah and coworkers [27] utilized BiLSTM for time series

prediction, specifically focusing on confirmed cases, deaths, and recoveries in ten major countries affected by COVID-19. The result confirms that the BiLSTM neural network can achieve robustness and higher prediction accuracy.

Although the BiLSTM model has been successfully applied in many areas, the disadvantages of a single deep learning network are gradually becoming apparent. As the collected temperature data are nonlinear and complex with multiple features (e.g., temporal features, periodic features), it is challenging for a single model to fully capture all the features within the data. Thus, the hybrid model is often used in actual applications, since it successfully preserves the strengths of each model and attains improved prediction accuracy [28]. Xiao and coworkers [29] proposed a hybrid model that combines the LSTM neural network with an attention mechanism to handle the time series data to extract more important features from the historical data. The experimental results illustrate that the suggested hybrid model outperforms individual models in terms of robustness and prediction accuracy. Chen and his coworkers [30] developed a multi-scale CNN and LSTM model for fault diagnosis; this model has shown higher classification accuracy than some traditional intelligent algorithms, particularly in noisy environments. To process the nonlinear data of the concentration of PM2.5, a CNN-LSTM hybrid model is proposed by Li and coworkers [31]. The hybrid model retains the advantages of CNN and LSTM to achieve a higher prediction accuracy than the single models. Qiao and coworkers [32] proposed a hybrid model combining the wavelet transform (WT), LSTM, and a stacked autoencoder (SAE) to predict the electricity of America; the hybrid model can achieve higher robustness.

In this paper, a pre-trained 1DCNN-BiLSTM hybrid network is developed for the temperature prediction of the wind turbine gearboxes. In the proposed network, the hybrid model merges the 1DCNN's ability to process sequential signals and extract spatial features with the BiLSTM network's capability to extract periodic features. This allows the spatial and periodic features to be well extracted from the actual temperature datasets measured from the target wind turbine gearbox. Moreover, with the pre-training method, the parameters of the model are not randomly initialized, so the problem of local minima can be basically prevented, thus significantly improving the accuracy of the prediction result. In the proposed method, a 1DCNN and a BiLSTM network are pre-trained by the pre-training dataset. Then, the pre-training-based 1DCNN-BiLSTM model is constructed by the above two pre-trained networks. Finally, the wind turbine gearbox dataset is fed into the pre-trained hybrid model for model fine-tuning. To assess the effectiveness of the proposed model, a set of experiments is devised using three temperature datasets; the results indicate that the proposed model leads to improved prediction accuracy and enhanced robustness. The contributions of this paper are summarized as follows:

1. A novel hybrid 1DCNN-BiLSTM model is designed in this paper. Thus, the most important spatial and periodic features contained in the temperature data can be completely extracted by the hybrid model; thus, higher prediction accuracy can be achieved.
2. The pre-training method is creatively introduced into the model training. As a result, it is not necessary to randomly initialize the parameters of the hybrid model. This effectively prevents the training process from being trapped by local minima and significantly improves the prediction accuracy of the trained network.
3. Several experiments are conducted to evaluate the effectiveness of the proposed model by the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset and the real wind turbine temperature dataset. Evidently, the proposed model outperforms certain classical models for temperature prediction.

The remains of this article are organized as follows. In Section 2, the basic theory of the proposed model is presented. In Section 3, the implementation of the proposed method is reported in detail. In Section 4, several experiments are conducted to evaluate the performance of the proposed model. Finally, the conclusion is given in Section 5.

2. Theoretical Background

2.1. One-Dimensional Convolutional Neural Network

2.1.1. Traditional Convolutional Neural Network (CNN)

The structure of a traditional CNN is shown in Figure 1. A CNN is a feedforward neural network that incorporates convolutional layers, pooling layers, and fully connected layers. Through the convolutional layers and pooling layers, it captures the inherent features within the data and subsequently conducts feature classification by the fully connected layers [33]. The weight sharing and local connection characteristics of a CNN can reduce the connection between network layers, and there is also the possibility of overfitting [34].

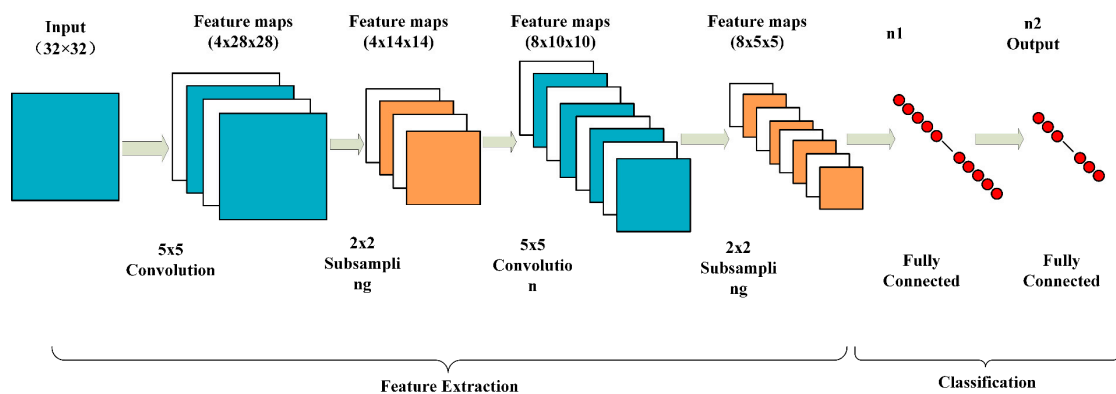


Figure 1. The structure of a traditional convolutional neural network.

As shown in Figure 1, a 32×32 matrix is the input to a CNN for feature classification. In the first convolutional layer, the size of the convolution kernel is 5×5 , the number of channels is 4, and the step size is 1. Being convolved through the sliding window method, the input matrix can be mapped into a $4 \times 28 \times 28$ feature map. In the subsequent pooling layer, the pooling kernel size is 2×2 , and the step size is 1. Only the maximum value in each set of pooling operations is preserved for sampling; thus, the feature map is converted into $4 \times 14 \times 14$. The following convolutional layer and pooling layer are the same as the aforementioned layers. Finally, the feature classification will be realized by fully connected layers.

2.1.2. The One-Dimensional Convolutional Neural Network

The primary application of a traditional CNN is in the recognition of two-dimensional (2D) images, so the input is a 2D matrix. When dealing with one-dimensional (1D) time-domain signals, various transformation methods are available for converting a 1D vector to a 2D matrix. No matter which transformation method is used, the correlations among the input data points will be altered. Thus, the accuracy of the prediction results may be seriously affected. Thus, the one-dimensional convolutional neural network (1DCNN) is developed to handle the 1D time-domain signals for retaining the correlation among data points and so as to ensure the features in the input data can be completely extracted [35].

Like the traditional CNN, the 1DCNN consists of an input layer, convolutional layers, pooling layers, fully connected layers, and an output layer. The input of the 1DCNN is 1D vectors, while the output is usually 1D feature vectors [36].

In the 1DCNN, the convolutional layer is the most important component among the five kinds of layers. The developed convolution kernel is designed to handle a 1D input, so the features can be completely extracted and preserved. The formula of the 1D convolutional layer is described as follows [36]:

$$z_i^l = \sigma \left(b_i^l + \sum_j z_j^{l-1} \times w_{ij}^l \right) \quad (1)$$

where z_i^l stands for the i th features in the l th layer; z_j^{l-1} is the j th feature in the $(l-1)$ th layer; w_{ij}^l represents the weights of the convolutional kernel connected to the j th feature in the l th layer; and b_i^l is the bias value of this feature. In this network, ReLU is employed as the activation function, and it can be expressed as follows:

$$\sigma(x) = \max(0, x) \quad (2)$$

Similar to other traditional CNNs, the inclusion of the pooling layer simplifies calculations by reducing the network's complexity. The commonly used pooling layer is max pooling or mean pooling. The proposed model employs max pooling, which can be described as follows:

$$z_i^l = \max_{p \in \Omega_i} z_p^{l-1} \quad (3)$$

where Ω_i stands for the pooling region with index i .

The extracted features are nonlinearly combined and fed into the fully connected layer to achieve feature classification and learning. In the CNN, the fully connected layer plays a crucial role in efficiently amalgamating the localized information extracted by the aforementioned two layers. The ReLU function is utilized for activating neurons in this layer to improve the performance of the fully connected layer [37].

The model structure of the adopted 1DCNN model, which is employed to extract the spatial features of temperature data, is shown in Figure 2. Features contained in the data will be extracted by the convolutional layer and pooling layer, alternatively. Subsequently, the feature vector, comprising the extracted features, is transmitted to the fully connected layer for regression analysis.

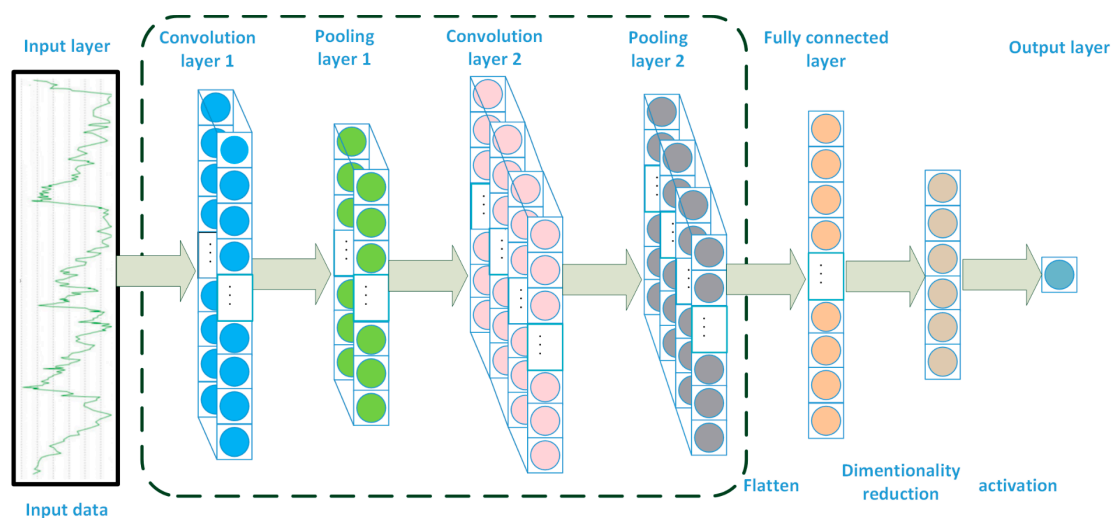


Figure 2. The structure of the 1DCNN model.

2.2. The Bi-Direction Long Short-Term Memory Network

To process time series data, the RNN model endows the network with memory ability. The output of the network can memorize information from the previous layer, so the outputs of the previous layer and the current layer are combined to form a feature vector to preserve more valuable features to improve prediction accuracy. However, the issues related to gradient disappearance and gradient explosion significantly affect the prediction accuracy and ultimately constrain the actual applicability.

LSTM stands out as a distinct category within the realm of RNNs; it is capable of effectively handling long-term dependencies in data while averting concerns related to gradient vanishing and explosion, which are common in standard RNN models. Three new gate structures (including the input gate, output gate, and forgetting gate) are designed in

the LSTM deep learning network to preserve and regulate the information flow throughout the time series data. The neural network can decide which information is useful in the long-term and short-term and, therefore, makes it suitable for processing long time series data [38]. The structure of the LSTM neural network is shown in Figure 3.

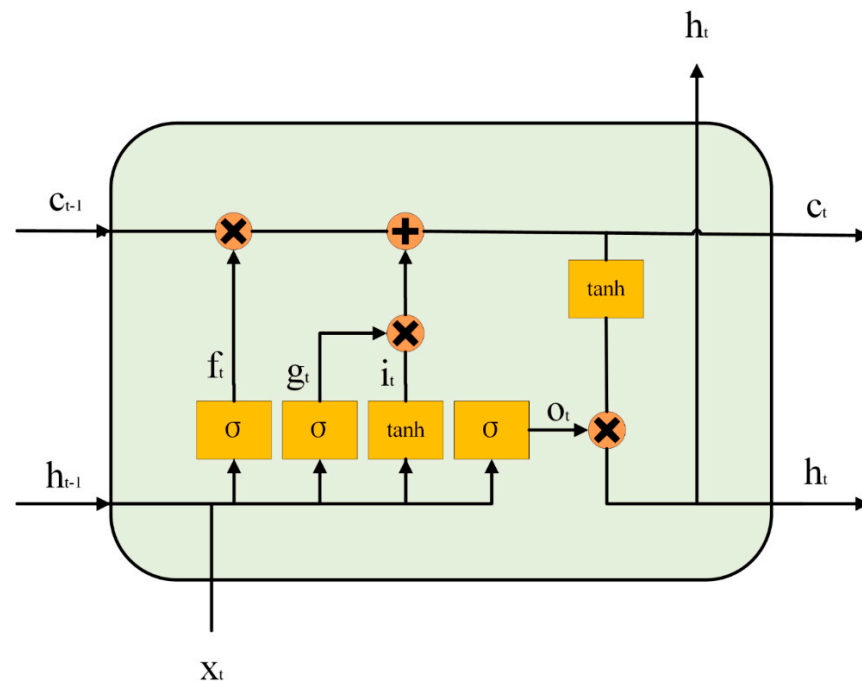


Figure 3. The structure of the LSTM neural network.

The input gate i_t , forgetting gate f_t , and output gate o_t can be described as follows [39]:

$$i_t = \text{sigmoid}(w_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

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

$$o_t = \text{sigmoid}(w_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

where h_{t-1} is the output information of the previous hidden layer, x_t is the input information of the current layer, h_t is the output information of the current layer, w is the weight of the reset gate, and b is the bias value.

The mathematical formulas of the activation functions are as follows:

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (7)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (8)$$

The output results of the long-term memory c_t and the short-term memory h_t are described as follows [39]:

$$c_t = f_t \otimes c_{t-1} \oplus i_t \otimes \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \quad (9)$$

$$h_t = o_t \otimes \tanh(c_t) \quad (10)$$

where \otimes is the multiplication of elements and \oplus is the addition of elements.

The inability to extract periodic features limits the capacity of the LSTM model to capture periodic patterns within time series data. To address the previously mentioned issue, a developed LSTM model is introduced [40]. The overall structure of the BiLSTM model is shown in Figure 4. Based on the LSTM model, the BiLSTM model merges a forward LSTM with a reverse LSTM, facilitating the extraction of features from both the forward and reverse directions; thus, the periodic features can be obtained from the training data. Like the LSTM model, the BiLSTM model also contains a three-gate structure [41].

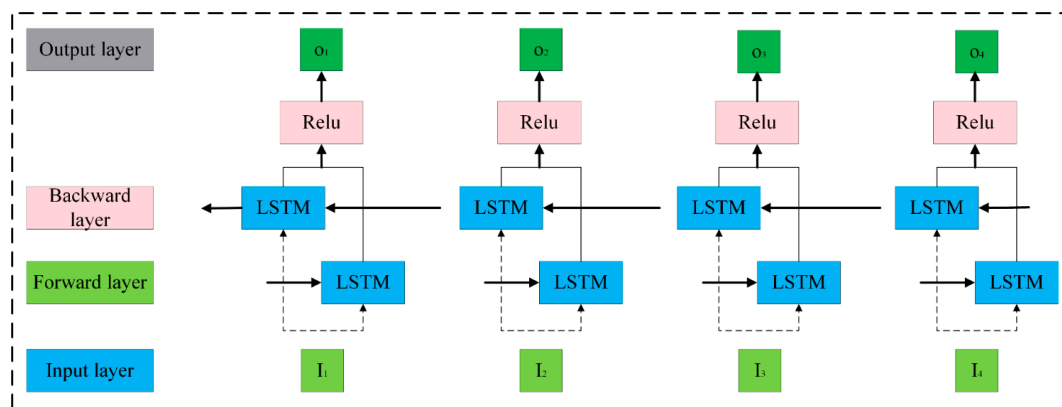


Figure 4. The structure of the BiLSTM neural network.

2.3. The Basic Theory of Pre-Training

The presence of local minima poses a significant challenge during model training because the model parameters are initially randomized in the conventional training process. This will substantially impair the prediction accuracy of the network. In addressing this issue, a pre-training method is introduced to fine-tune the model parameters, ultimately enhancing the precision of the prediction result.

The essence of the pre-training method is that the model parameters are not randomly initialized. The model training process is divided into two parts: commonality learning and feature learning [42,43]. The model is first pre-trained with pre-training data, enabling the model to obtain the initial parameters. Then, the parameters will be fine-tuned to enhance prediction accuracy in the subsequent temperature forecasting. The characteristics of commonality learning and feature learning significantly enhance the model's scalability. At present, model pre-training has been employed in a variety of machine learning projects, such as image classification and sentence relationship judgment.

The pre-training method is employed in the proposed model. That is, the 1DCNN model and the BiLSTM model are firstly pre-trained to obtain a set of initial model parameters. Then, the two pre-trained models are combined to form the hybrid model with the initial parameters being retained. Compared to the 1DCNN-BiLSTM model without pre-training, it is clear that the model pre-training is helpful to improve the prediction accuracy [42,44].

3. The Proposed Pre-trained 1DCNN-BiLSTM Model

3.1. Overview of the Proposed Hybrid Model

This paper presents an innovative pre-trained 1DCNN-BiLSTM network designed for temperature prediction. The hybrid model is created by integrating a pre-trained 1DCNN model with a pre-trained BiLSTM neural network. The pre-training phase yields a set of initial model parameters, mitigating the risk of encountering local minima. Subsequently, temperature data are input into the pre-training-based hybrid model for both model training and regression analysis. The overall architecture of the proposed model is depicted in Figure 5. The 1DCNN and BiLSTM models undergo initial pre-training using a temperature dataset to acquire the initial parameters. Following this, the relevant convolutional layers,

pooling layers, and bidirectional long short-term layers are extracted to formulate the pre-trained models. Finally, the pre-trained model undergoes fine-tuning using an experimental dataset. The process of temperature prediction and the detailed procedures are further elaborated in the subsequent subsections.

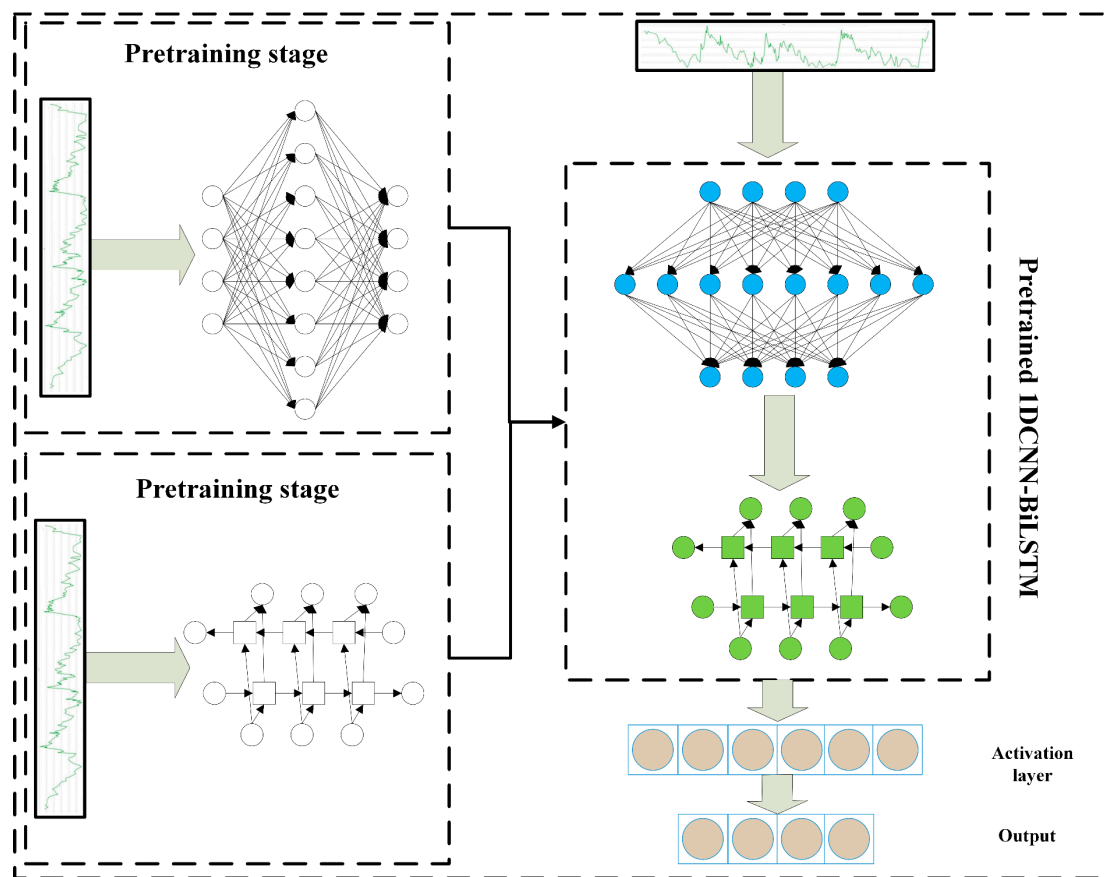


Figure 5. The process of temperature prediction. The line represents the temperature curve, the arrow represent the model training flow. The blue cycle represents the pretrained cells of 1DCNN, and the green represents the pretrained cells of BiLSTM. The last cycles represent the pro-cessed data and then will output as predicted temperature.

3.2. Data Collection and Processing

In this section, three experiments are performed based on three datasets. The first one is the turbine engine dataset published by the National Aeronautics and Space Administration (NASA) in 2020 [45]. The data are the simulated data of the turbofan aircraft engine obtained by the C-MPASS platform. It is one of the most widely used datasets for RUL predictive studies. The dataset consists of four sub-datasets and each of them contains several kinds of data collected from 100 engines; the temperature dataset was used to assess the performance of the proposed model.

Another two datasets contain temperature data collected from an actual wind turbine gearbox in an existing wind farm, and the measurement duration is two minutes. Two comparative experiments have been structured to assess the effectiveness of the proposed model using the provided datasets. The location and actual view of the wind farm are shown in Figure 6.



Figure 6. Location and actual view of an existing wind farm.

The preprocessing methods employed for the data include data normalization and data slicing. Initial temperature data are normalized firstly to avoid the occurrence of the problem that large numbers cover up small numbers when the difference in values is too large. Meanwhile, since the gradient descent will be used for the model optimization, data normalization can avoid the appearance of a poor solution path. The min-max normalization was employed for data processing. Taking the wind turbine gearbox temperature data normalization as an example, the normalized data are shown in Figure 7. The temperature data were normalized to a range of [0, 1].

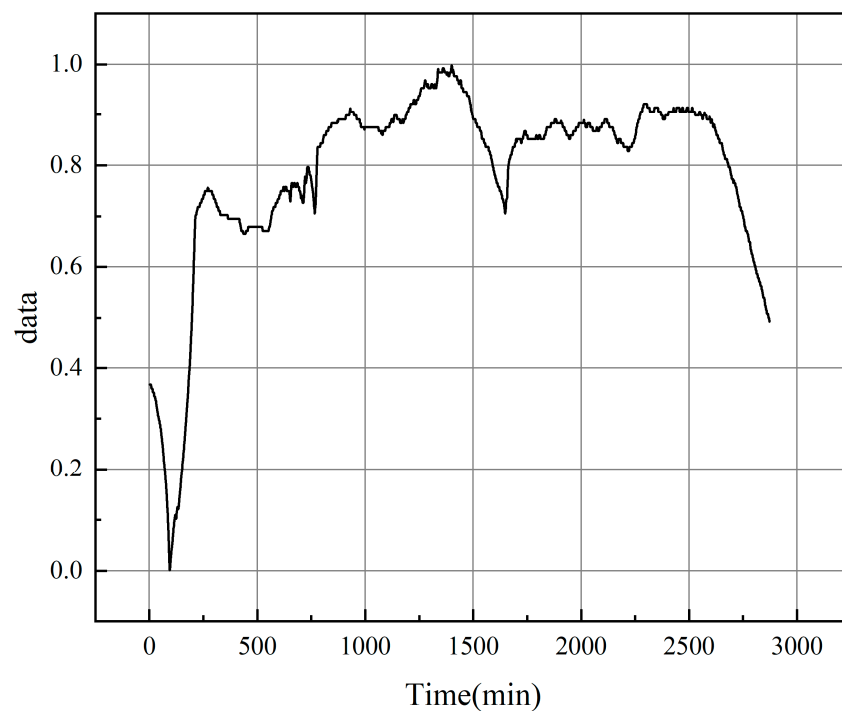


Figure 7. Data normalization.

After the data normalization, the dataset should be partitioned into training and test sets with an 8:2 split ratio. For a dataset with N data points, the data slicing process will generate m training-label data pairs. The process starts as follows:

1. The first n data points are selected as a training sample and the next data point as the label for this training sample. The value of n can be selected according to the complexity of the input data.
2. Then, the first data point in the dataset is removed.
3. The next data pair is generated by repeating steps one and two until the number of data points is less than or equal to n .

Eventually, m training-label data pairs can be formed for model training. An example of data slicing with $n = 5$ is shown in Figure 8.



Figure 8. An example of data slicing with $n = 5$.

3.3. The Components of the Hybrid Model

The proposed model utilizes the 1DCNN model as the initial step to process the data and extract spatial features. In the 1DCNN model, two convolutional layers with a kernel number of 32 and two pooling layers with a pooling size of two are stacked, alternatively. Both the convolutional layers and max pooling layers utilized a stride size of one. The detailed parameters of the 1DCNN model are shown in Table 1. The BiLSTM model is mainly employed to extract the periodic features that are contained in the temperature data. The detailed parameters of the BiLSTM model are shown in Table 2.

The BiLSTM model is mainly employed to extract the periodic features that are contained in the temperature data. The detailed parameters of the BiLSTM model are shown in Table 2. The two models are the components of the proposed hybrid model, and both of them are pre-trained before being combined into a hybrid model.

Table 1. The parameters of the 1DCNN model.

The Definition of Layers	The Parameters of Layers
Input layer	Input size: (1, 5)
Convolutional layer 1	Channel number: 32, nuclear size: 1
Pooling layer 1	Nuclear size: 2, step size: 2
Convolutional layer 2	Channel number: 32, nuclear size: 1
Pooling layer 2	Nuclear size: 2, step size: 2
Activation layer 1	Activation function: ReLU
Flatten layer	One-dimension feature vector
Output layer	Output size: 1

Table 2. The parameters of BiLSTM.

The Definition of Layers	The Parameters of Each Layer
Input layer	Input size: (1, 16)
BiLSTM layer	Channel number: 64, activation function: ReLU
Output layer	Output size: 1

3.4. Building the Hybrid Model

To leverage the advantages of the 1DCNN and BiLSTM models, the hybrid model is constructed by combining the two pre-trained networks. The parameters of each layer have already been adjusted after pre-training and the initial parameters will be retained in the hybrid model. In the target problem, the parameters of the pre-training-based hybrid model will be fine-tuned to achieve the optimal configuration. The detailed parameters of the 1DCNN-BiLSTM model are shown in Table 3.

Table 3. The parameters of the pre-trained 1DCNN-BiLSTM model.

The Definition of Layers	The Parameters of Each Layer
Input layer	Input size: (1, 5)
Convolutional layer 1	Channel number: 32, nuclear size: 1
Pooling layer 1	Nuclear size: 2, step size: 2
Convolutional layer 2	Channel number: 32, nuclear size: 1
Pooling layer 2	Nuclear size: 2, step size: 2
Activation layer 1	Activation function: ReLU
BiLSTM layer	Channel number: 64, activation function: ReLU
Output layer	Output size: 1

4. Case studies of Temperature Prediction

The proposed model is evaluated using the C-MAPSS public dataset and two measured temperature datasets.

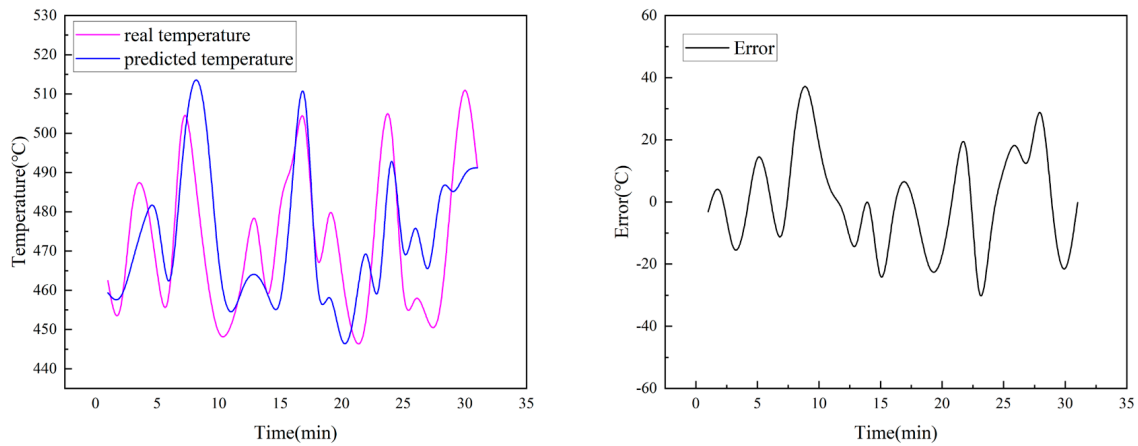
4.1. Case 1: Experiments Using the C-MAPSS Dataset

The C-MAPSS dataset contains simulated data generated from 100 turbine engines. The blade temperature data of turbine engine 81 are employed for model validation. For comparison purposes, five different models are considered in this study:

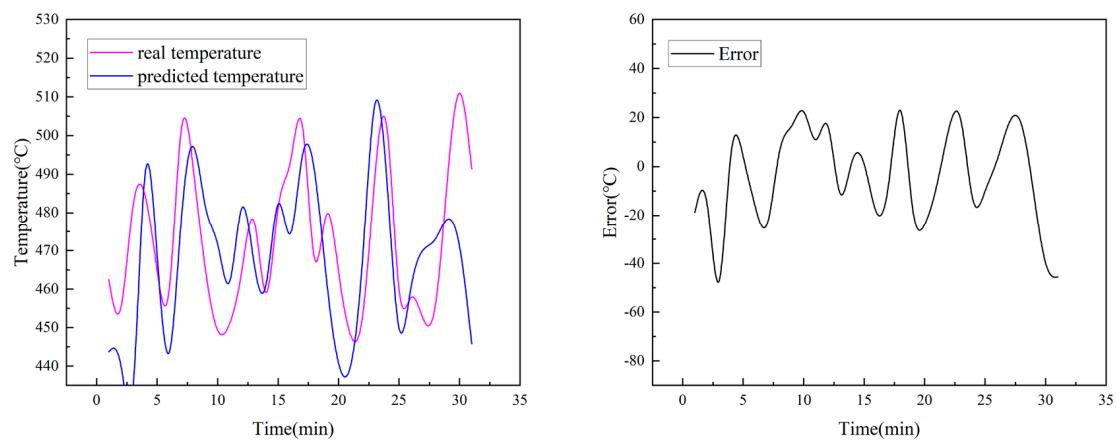
- M0: proposed pre-trained 1DCNN-BiLSTM model,
- M1: CNN model,
- M2: RNN model,
- M3: CNN-LSTM model,
- M4: 1DCNN-BiLSTM model without pre-training, and
- M5: residual shrink neural network.

These six models are used to predict the temperature of a turbine engine. M1 is a 1DCNN model with the same parameters as the 1DCNN model in M0. M4 is an ordinary

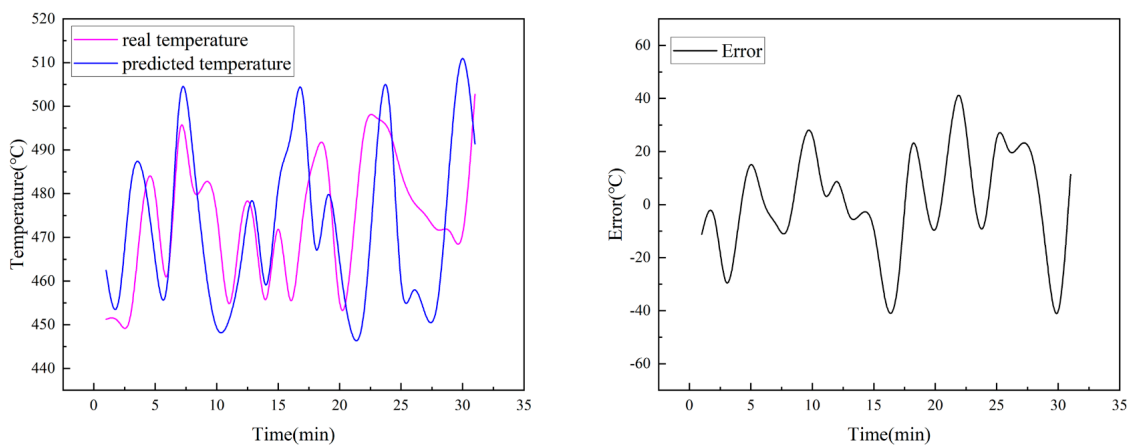
1DCNN-BiLSTM hybrid neural network without pre-training. M5 is a residual shrink neural network proposed by Zhao and coworkers [46] in 2020. This model employs a soft-threshold denoising technique, endowing it with robust noise-resistance capabilities. Figure 9 shows the comparison of the predicted and actual values of the engine blade temperature and its error for M0 to M5.



(a) M0



(b) M1



(c) M2

Figure 9. Cont.

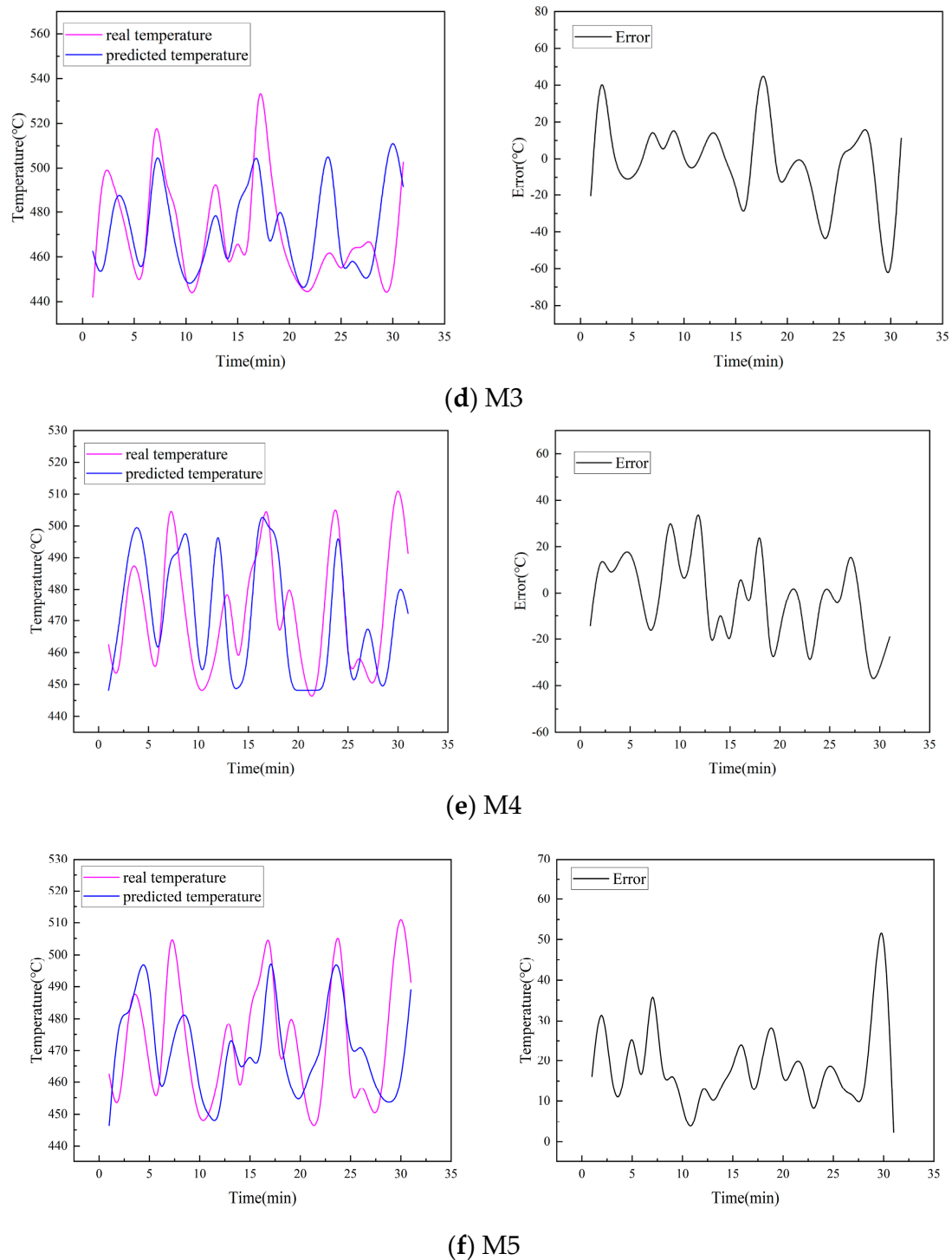


Figure 9. The comparison of the six models in Case 1.

As shown in Figure 9, it is clear that the proposed model (M0) can predict the trend of temperature change well. The prediction accuracies of all six models are similar, while the maximum prediction error of M0 is significantly smaller than that of the other five methods. To further evaluate the performance of different models, the root mean square

error (RMSE) and the mean absolute error (MAE) of the prediction results are employed. Their mathematical formulations are as follows [25]:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{X}_i - X_i)^2} \quad (11)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{X}_i - X_i| \quad (12)$$

where \hat{X}_i and X_i are the i th predicted value and the actual value, and N is the number of predicted values. The indicators corresponding to the five models are shown in Table 4.

Table 4. The comparison of performance among the five models.

Model	RMSE	MAE
M0	22.61107	17.46969
M1	27.59789	22.37025
M2	26.48077	21.16881
M3	30.95766	22.54774
M4	26.64763	20.69926
M5	23.58911	18.61125

Table 4 presents the prediction error values obtained by the six models. The overall prediction performance of the proposed model surpasses that of the CNN, indicating that the hybrid model retains the advantages of the CNN and BiLSTM and achieves better performance than a single model. When comparing M0 to M4, the prediction error is significantly reduced by the pre-training process, and it can be concluded that the pre-training method can help neural networks select a better optimization path to obtain the optimal parameters, therefore significantly improving the performance of temperature prediction. When comparing M0 to M5, it can be concluded that the proposed model can achieve a more stable prediction result. The maximum error is clearly lower than M5, showing the strong feature extraction and noise-resistance capabilities of the proposed model.

4.2. Experiments Using Measured Dataset from Hejiashan Wind Farm

4.2.1. Case 2a: Wind Turbine Gearbox Temperature

In this part, measured gearbox temperature data collected from an existing wind turbine are utilized to validate the practicality and reliability of the proposed model. The same validation approach is used to verify the validity of the model based on a real temperature dataset.

Figure 10 shows the comparison of the predicted temperature obtained by the proposed model (M0), CNN model (M1), RNN model (M2), CNN-LSTM model (M3), 1DCNN-BiLSTM model without pre-training (M4), and residual shrink neural network (M5). Figure 11 shows the training data, predicted data, and prediction error, representing the variance between the predicted and actual temperatures, across the five models.

It is clear from Figure 10 that all six models can predict the trend of temperature changes. However, the values of the predicted temperature are different. It can be seen from Figure 11 that temperatures predicted by the proposed model have the lowest prediction error (i.e., the highest accuracy) among all models considered in this study. The maximum error appears at the time when the temperature changes sharply. It is the most important time point to which more attention needs to be paid in actual temperature monitoring. Detailed information has been acquired from the prediction error curves, and the maximum error value of the proposed model exhibits a smaller maximum error value compared to other models. The error values are less than 2 °C.

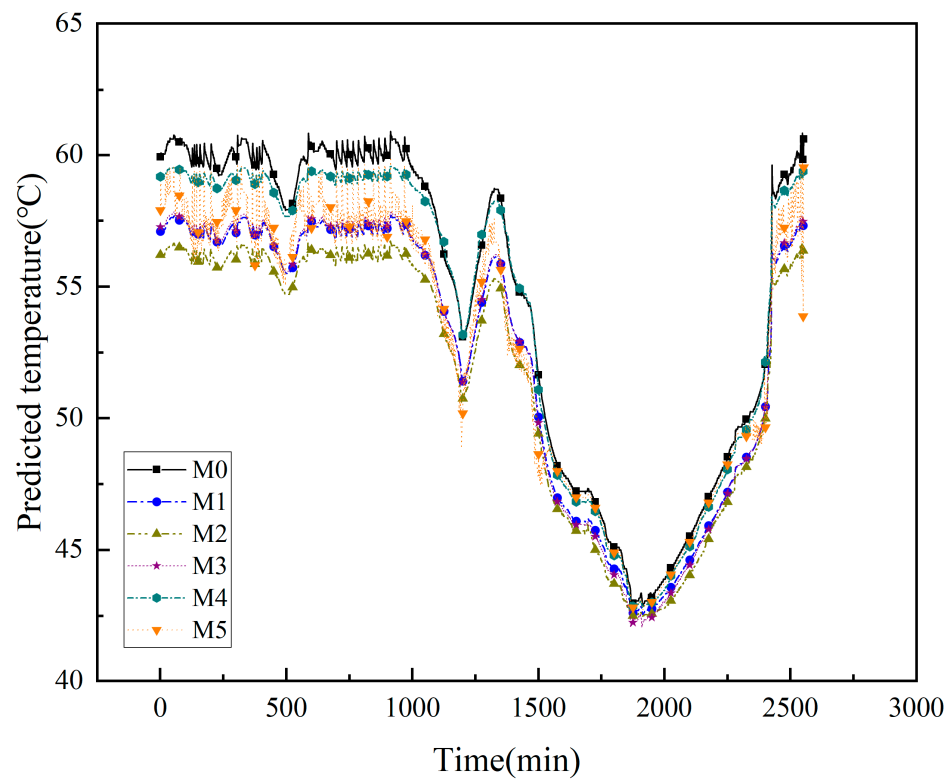


Figure 10. The predicted temperature of the six models in Case 2a.

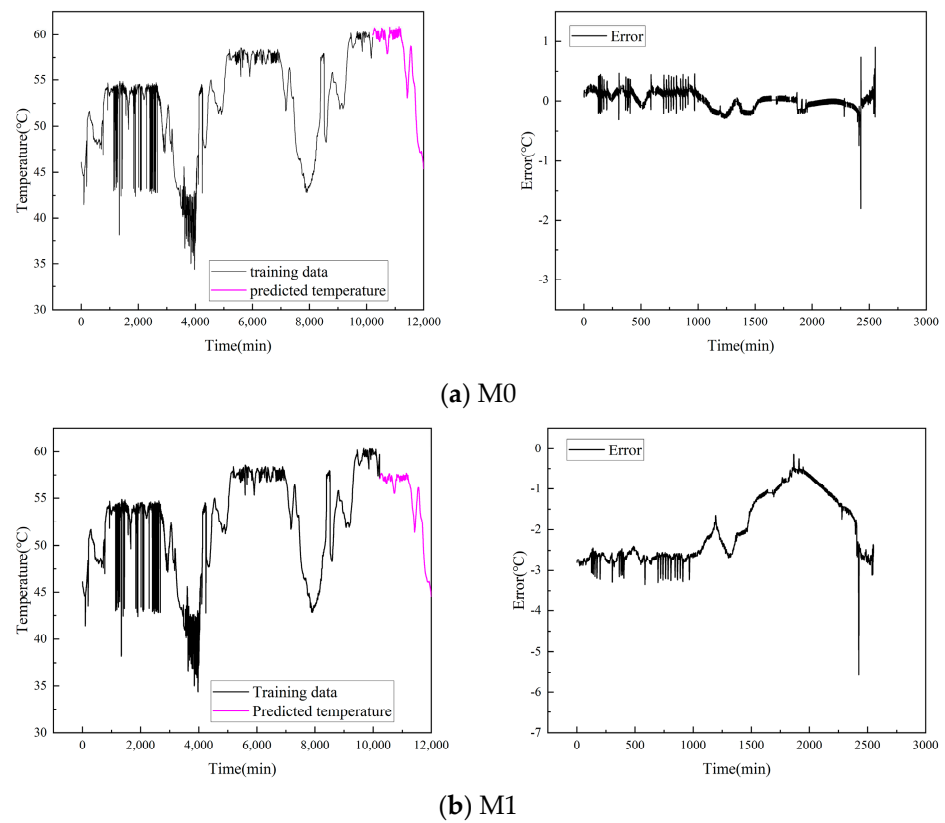
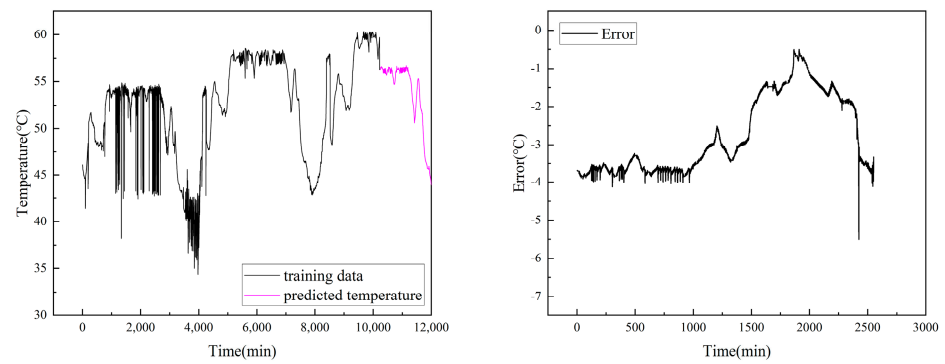
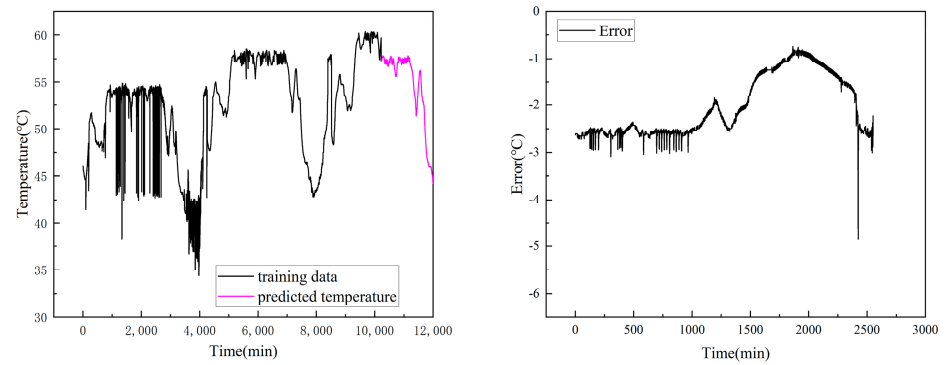


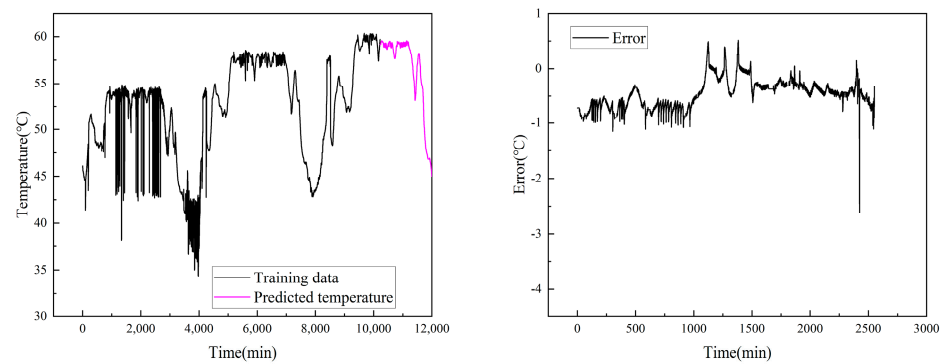
Figure 11. Cont.



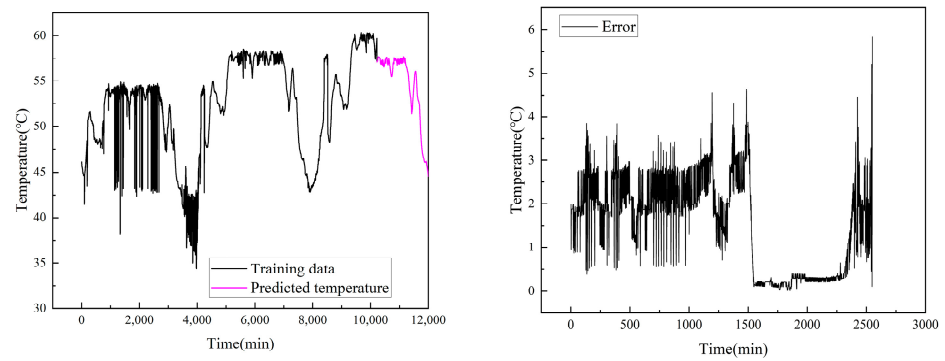
(c) M2



(d) M3



(e) M4



(f) M5

Figure 11. The comparison of the six models in Case 2a.

In the experiment, although all six models can predict the temperature of the gearbox, the prediction errors of M1 (CNN model) and M2 (RNN model) are higher than 3 °C on

average, and the maximum error is about 6 °C. It can be concluded that the ability of a single model to handle complex signals is weaker than hybrid models. M1 can hardly extract the temporal feature and the periodic feature contained in the temperature data, and the RNN model (M2) can hardly process long time series signals. Therefore, the prediction errors of these two single models are higher than the other hybrid models.

Even though the accuracies of M3 and M4 are higher than those of M1 and M2, the prediction accuracy is terrible at the point where the temperature sharply changes. The prediction errors are about 3 °C and 5 °C, but notably, M0 (the proposed model) achieves a prediction error of less than 2 °C. The result in Case 2a underscores the effectiveness of the pre-training method in enhancing the performance of the model.

When comparing M0 to M5, from Figure 11, it is evident that the predictive results of M5 exhibit significant volatility, indicating that the predictive performance of this model is inferior to the approach proposed in this paper.

In this experiment, RMSE, MAE, and the coefficient of determination R2 are employed to evaluate the performance of the models. The mathematical formulation of the coefficient of determination R2 is as follows [35]:

$$R2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i \left(y_i - \frac{1}{n} \sum_{i=1}^n y_i \right)^2} \quad (13)$$

where f_i is the predicted value of the model, and y_i is the true value. The value range of the R2 coefficient is [0, 1]. A higher value of R2 represents a better effect of the regression model. The results of the five models are summarized in Table 5.

Table 5. The comparison of results among the five models in Case 2a.

Model	RMSE	MAE	R2
M0	0.12252	0.07602	0.99959
M1	1.73922	1.59984	0.89635
M2	2.04201	1.85558	0.84766
M3	1.71908	1.45505	0.89737
M4	0.57687	0.52392	0.99061
M5	1.91532	1.55412	0.87133

Table 5 makes it evident that M0 (the proposed model) outperforms all other models with significantly lower RMSE and MAE values, along with the highest R2 coefficient. The prediction errors of M1 and M2 are higher than M3 and M4. The experimental results indicate that the hybrid models enhance the feature extraction capability as well as the prediction accuracy. Meanwhile, the MAE of M0 has an average of 589% reduction when compared to M4 (the 1DCNN-BiLSTM model without pre-training). The three metrics for M0 are superior to those of M5, indicating that the model's adaptability is the highest among the six models. The pre-training-based 1DCNN-BiLSTM model is more suitable for actual wind turbine temperature prediction.

4.2.2. Case 2b: Wind Turbine Bearing Temperature

The bearing is a key component of a wind turbine gearbox, but it can be easily affected by environmental factors. Abnormal bearing temperature will lead to the deterioration of its function, resulting in a vicious cycle during operation. Furthermore, the deterioration of bearing function may lead to gearbox damage.

Therefore, in this experiment (Case 2b), a measured bearing temperature dataset of a wind turbine in an existing wind farm is utilized to assess the performance of the proposed model. Similar to the previous experiment (Case 2a), M1 to M5 are utilized to demonstrate the performance of the proposed model (M0) in temperature prediction. Figure 12 shows the comparison of the temperatures predicted by the five models (i.e., M0 to M5). Figure 13 shows the training data, the predicted data, and the error of the predicted temperature.

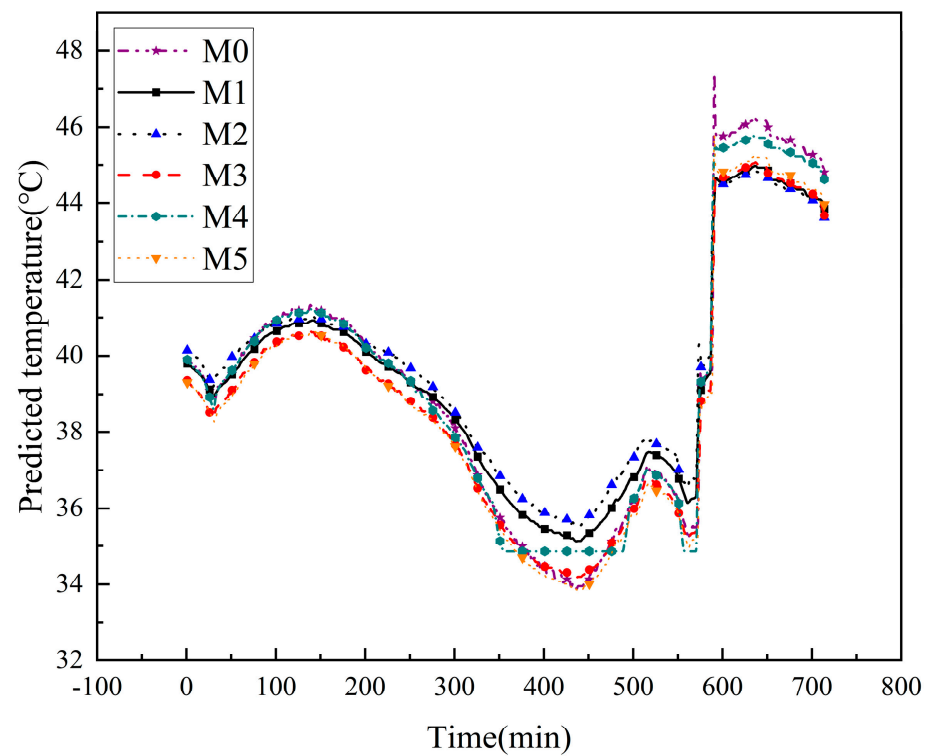


Figure 12. The predicted temperature of the five models in Case 2b.

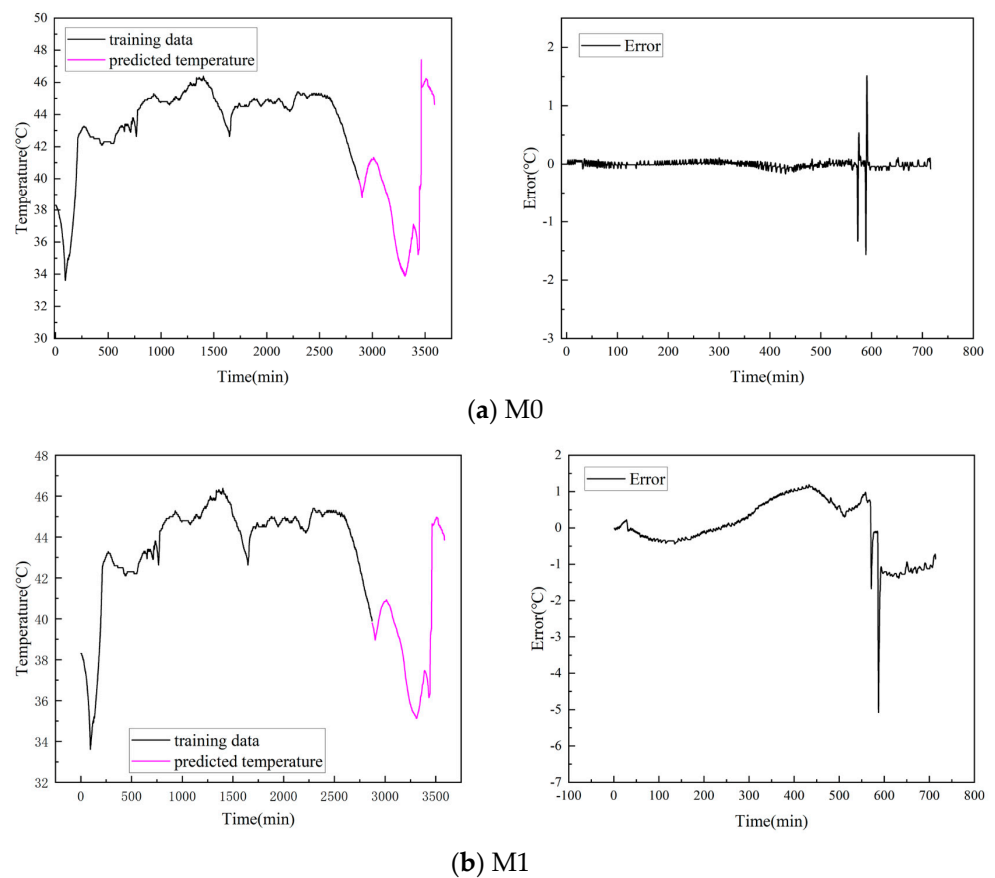
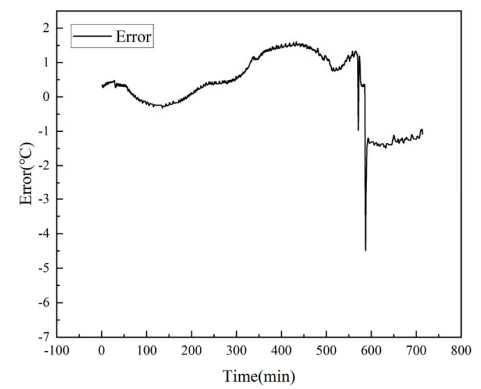
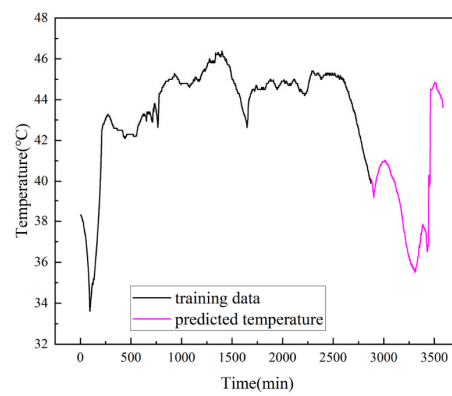
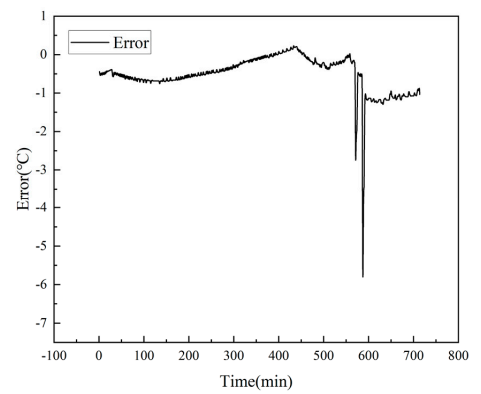
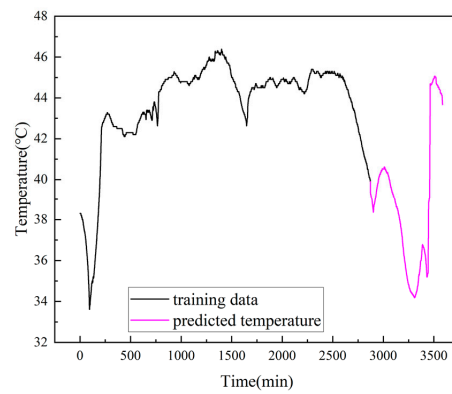


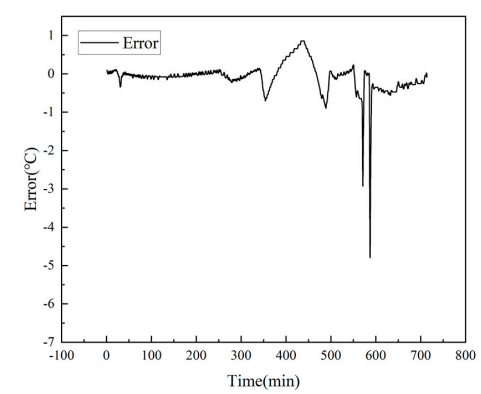
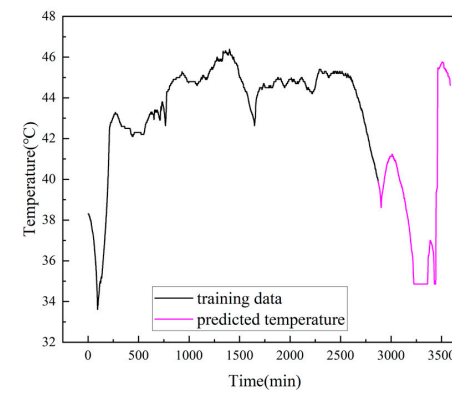
Figure 13. Cont.



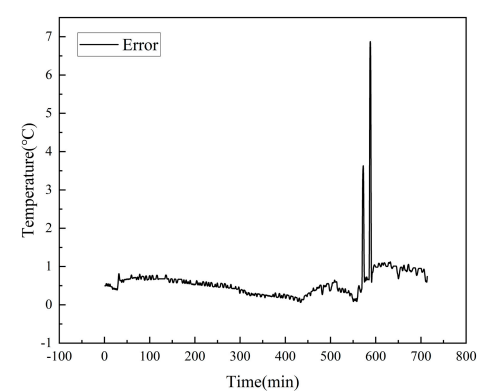
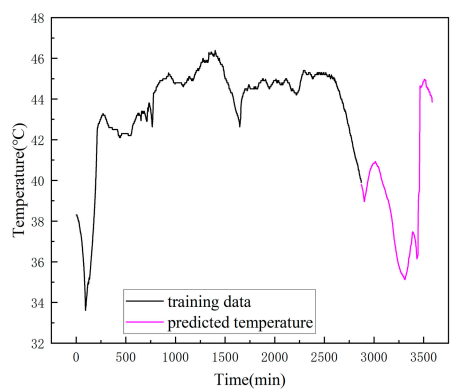
(c) M2



(d) M3



(e) M4



(f) M5

Figure 13. The comparison of the six models in Case 2b.

The trend of the temperature changes predicted by the five models can be observed in Figure 12. The predicted trends of the temperature changes by all six models are similar while the values are different. It can be concluded that the performances of different models are different.

From Figure 13, it is evident that M0 (the proposed model) exhibits a relatively low prediction error. Even though the highest prediction error is about 3 °C, the prediction error of M0 is lower than 1 °C in most of the considered time duration. The mean error and the highest error of M0 are obviously lower than those of the other five models. When the results from M0 are compared to those of M4, it can be concluded that the performance of the hybrid model is enhanced by the pre-training method. The analysis results in this case are consistent with Case 2a in that the proposed model can achieve a higher prediction accuracy. Next, the results from M1 and M2 are discussed. The prediction errors of M1 and M2 are more than 3 °C, and the maximum error is over 6 °C. These results are consistent with those in Case 2a. When comparing M0 to M5, the maximum error of M5 is obviously larger than that of M0, indicating that the stability of the proposed method is higher.

Similar to Case 2a, the prediction accuracy is very low when the temperature sharply changes. Even though the prediction error of M0 is less than 2 °C, it can be concluded that the pre-training method can improve the performance of the model.

Similarly, MAE, RMSE, and R2 are used to evaluate the performance of the five models. The analysis results are summarized in Table 6.

Table 6. The comparison of results among the five models in Case 2b.

Model	RMSE	MAE	R2
M0	0.29843	0.06297	0.99349
M1	0.80648	0.62013	0.92599
M2	0.91263	0.72302	0.89618
M3	0.72388	0.51454	0.95233
M4	0.43028	0.22917	0.98541
M5	0.67138	0.57061	0.96228

It can be concluded from Table 6 that the values of MAE and RMSE of M0 are on average 263~1048% and 44.2~205%, respectively, lower than those of other models. Meanwhile, the R2 coefficient also improved by using the proposed model in M0.

5. Conclusions

In regression analysis, the problems based on deep learning, exemplified by the temperature prediction, nonlinearity, and complexity inherent in the data, pose demanding challenges to the feature extraction performance of the model. In order to enhance the feature extraction capability of models, scholars often resort to constructing residual networks or hybrid networks. However, these models may encounter the issue of being trapped in local minima during the training process. Simultaneously, as the depth of the model increases, the associated problems of gradient vanishing or exploding also significantly diminish the effectiveness of the model. Therefore, the objective of this study is to innovatively introduce a pre-training approach while enhancing the feature extraction capability, aiming to circumvent the aforementioned issues.

In this paper, a pre-training-based 1DCNN-BiLSTM model is proposed for temperature prediction of the wind turbine gearbox. The proposed model retains the advantages of the 1DCNN model and the BiLSTM model. Furthermore, the pre-training method, which can pre-adjust the parameters of the model so that it is not necessary to randomly generate initial model parameters, is innovatively employed in the hybrid model. After combining the pre-trained 1DCNN model and the pre-trained BiLSTM model, the model parameters are fine-tuned in the subsequent training. The incorporation of the hybrid model along with the application of the pre-training method results in a remarkable enhancement in the performance of temperature prediction.

Then, three experiments were devised to assess the effectiveness of the proposed model; by comparing the error curves and three key indicators, it can be observed that the method proposed in this paper demonstrates superior performance in terms of both the error upper limit and error fluctuation. Meanwhile, by comparing the proposed method with the 1DCNN-BiLSTM model without pre-training, it is evident that the pre-training method can significantly enhance the model's performance. It demonstrated superior accuracy in prediction compared to five other existing models, achieving the highest level of prediction accuracy. Based on the experimental case studies, the following can be concluded:

- (a) The hybrid model retains the advantages of the two single models, and therefore, its more useful features can be extracted by the hybrid model, improving its performance at temperature prediction.
- (b) The pre-training method can help the model to obtain a better optimization path to obtain optimal parameters. Thus, the ability of the anti-interference of the model can be improved.
- (c) From the results of the experiments based on measured temperature data, the appropriateness of the proposed model in real applications is demonstrated.

Meanwhile, the research exhibits certain limitations that necessitate additional investigation. The following aspects, in particular, warrant further research:

- (a) The deep learning-based approach demonstrates strong generalization capabilities; it is reasonable to do further research on its universality. Employing the proposed method to predict the operational temperature of some other mechanical equipment, such as cement production machinery, aerospace engines, and so on, is worthy of further research.
- (b) The predictive accuracy of the proposed model in Case 1 is observed to be lower than in the other two cases. This is probably attributed to the insufficient data volume and the high complexity of the dataset. Therefore, in subsequent research, further refinement of the model's structure and parameters can be performed. Additionally, designing a dynamic loss function to capture dynamic biases could be considered, eventually achieving higher prediction accuracy based on a small sample size dataset.

Author Contributions: K.Z.: writing—review and editing, supervision; C.M.: writing—original draft, methodology, experiment, investigation, visualization, writing-review and editing; L.Z.: writing—review and editing, methodology; J.H.: writing—review and editing, conceptualization, supervision; H.-F.L.: project administration, funding acquisition, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: The work described in this article was partially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China [Project No. R5020-18 (RIF 8799008)], the Hainan Provincial Natural Science Foundation of China [project NO. 522CXTD517], and the Hainan Provincial Natural Science Foundation of China [project NO. 522RC879].

Data Availability Statement: Data are contained within the article.

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

References

- Han, J.; Chang, H. Development and opportunities of clean energy in China. *Appl. Sci.* **2022**, *12*, 4783. [\[CrossRef\]](#)
- Zhang, W.; Li, B.; Xue, R.; Wang, C.; Cao, W. A systematic bibliometric review of clean energy transition: Implications for low-carbon development. *PLoS ONE* **2021**, *16*, e0261091. [\[CrossRef\]](#) [\[PubMed\]](#)
- Liu, H.; Liang, D. A review of clean energy innovation and technology transfer in China. *Renew. Sustain. Energy Rev.* **2013**, *18*, 486–498. [\[CrossRef\]](#)
- Rinaldi, G.; Thies, P.R.; Johanning, L. Current status and future trends in the operation and maintenance of offshore wind turbines: A review. *Energies* **2021**, *14*, 2484. [\[CrossRef\]](#)
- Gao, Z.; Liu, X. An overview on fault diagnosis, prognosis and resilient control for wind turbine systems. *Processes* **2021**, *9*, 300. [\[CrossRef\]](#)

6. Guo, Z.; Pu, Z.; Du, W.; Wang, H.; Li, C. Improved adversarial learning for fault feature generation of wind turbine gearbox. *Renew. Energy* **2022**, *185*, 255–266. [\[CrossRef\]](#)
7. Heydari, A.; Garcia, D.A.; Fekih, A.; Keynia, F.; Tjernberg, L.B.; De Santoli, L. A hybrid intelligent model for the condition monitoring and diagnostics of wind turbines gearbox. *IEEE Access* **2021**, *9*, 89878–89890. [\[CrossRef\]](#)
8. Yu, X.; Tang, B.; Zhang, K. Fault diagnosis of wind turbine gearbox using a novel method of fast deep graph convolutional networks. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 6502714. [\[CrossRef\]](#)
9. Zhang, X.; Zhong, J.; Li, W.; Bocian, M. Nonlinear dynamic analysis of high-speed gear pair with wear fault and tooth contact temperature for a wind turbine gearbox. *Mech. Mach. Theory* **2022**, *173*, 104840. [\[CrossRef\]](#)
10. Liu, H.; Yu, C.; Yu, C. A new hybrid model based on secondary decomposition, reinforcement learning and SRU network for wind turbine gearbox oil temperature forecasting. *Measurement* **2021**, *178*, 109347. [\[CrossRef\]](#)
11. Yang, Y.; Liu, A.; Xin, H.; Wang, J. Fault early warning of wind turbine gearbox based on multi-input support vector regression and improved ant lion optimization. *Wind Energy* **2021**, *24*, 812–832. [\[CrossRef\]](#)
12. Wang, A.; Qian, Z.; Pei, Y.; Jing, B. A de-ambiguous condition monitoring scheme for wind turbines using least squares generative adversarial networks. *Renew. Energy* **2022**, *185*, 267–279. [\[CrossRef\]](#)
13. Foley, A.M.; Leahy, P.G.; Marvuglia, A.; McKeogh, E.J. Current methods and advances in forecasting of wind power generation. *Renew. Energy* **2012**, *37*, 1–8. [\[CrossRef\]](#)
14. Tran TT, K.; Bateni, S.M.; Ki, S.J.; Vosoughifar, H. A review of neural networks for air temperature forecasting. *Water* **2021**, *13*, 1294. [\[CrossRef\]](#)
15. Baoyu, L.; Guoxing, L.; Guiyu, W.; Guofeng, Z.; Man, Y. Research on CART Model of Mass Concrete Temperature Prediction Based on Big Data Processing Technology. *IEEE Access* **2022**, *10*, 32845–32854. [\[CrossRef\]](#)
16. Das, S.; Politis, D.N. Predictive inference for locally stationary time series with an application to climate data. *J. Am. Stat. Assoc.* **2021**, *116*, 919–934. [\[CrossRef\]](#)
17. Wang, N.; Li, J.Y. Efficient multi-channel thermal monitoring and temperature prediction based on improved linear regression. *IEEE Trans. Instrum. Meas.* **2021**, *71*, 9500809. [\[CrossRef\]](#)
18. Zou, L.; Lam, H.F.; Hu, J. Adaptive resize-residual deep neural network for fault diagnosis of rotating machinery. *Struct. Health Monit.* **2022**, *22*, 2193–2213. [\[CrossRef\]](#)
19. Mu, R.; Zeng, X. A review of deep learning research. *KSII Trans. Internet Inf. Syst.* **2019**, *13*, 1738–1764. [\[CrossRef\]](#)
20. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaria, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 1–74. [\[CrossRef\]](#)
21. Sezer, O.B.; Gudelek, M.U.; Ozbayoglu, A.M. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Appl. Soft Comput.* **2020**, *90*, 106181. [\[CrossRef\]](#)
22. Zhao, Z.; Chen, W.; Wu, X.; Chen, P.C.; Liu, J. LSTM network: A deep learning approach for short-term traffic forecast. *IET Intell. Transp. Syst.* **2017**, *11*, 68–75. [\[CrossRef\]](#)
23. Vlachas, P.R.; Byeon, W.; Wan, Z.Y.; Sapsis, T.P.; Koumoutsakos, P. Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2018**, *474*, 20170844. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Jia, X.; Ji, Q.; Han, L.; Liu, Y.; Han, G.; Lin, X. Prediction of Sea Surface Temperature in the East China Sea Based on LSTM Neural Network. *Remote Sens.* **2022**, *14*, 3300. [\[CrossRef\]](#)
25. Wu, X.; Li, J.; Jin, Y.; Zheng, S. Modeling and analysis of tool wear prediction based on SVD and BiLSTM. *Int. J. Adv. Manuf. Technol.* **2020**, *106*, 4391–4399. [\[CrossRef\]](#)
26. Jiang, L.; Yan, C.; Zhang, X.; Zhou, B.; Cheng, T.; Zhao, J.; Gu, J. Temperature prediction of battery energy storage plant based on EGA-BiLSTM. *Energy Rep.* **2022**, *8*, 1009–1018. [\[CrossRef\]](#)
27. Shahid, F.; Zameer, A.; Muneeb, M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos Solitons Fractals* **2020**, *140*, 110212. [\[CrossRef\]](#)
28. Zheng, H.; Lin, F.; Feng, X.; Chen, Y. A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 6910–6920. [\[CrossRef\]](#)
29. Kun, X.; Shan, T.; Yi, T.; Chao, C. Attention-based long short-term memory network temperature prediction model. In Proceedings of the 2021 7th International Conference on Condition Monitoring of Machinery in Non-Stationary Operations (CMMNO), Guangzhou, China, 11–13 June 2021; pp. 278–281. [\[CrossRef\]](#)
30. Chen, X.; Zhang, B.; Gao, D. Bearing fault diagnosis base on multi-scale CNN and LSTM model. *J. Intell. Manuf.* **2021**, *32*, 971–987. [\[CrossRef\]](#)
31. Li, T.; Hua, M.; Wu, X.U. A hybrid CNN-LSTM model for forecasting particulate matter (PM_{2.5}). *IEEE Access* **2020**, *8*, 26933–26940. [\[CrossRef\]](#)
32. Qiao, W.; Li, Z.; Liu, W.; Liu, E. Fastest-growing source prediction of US electricity production based on a novel hybrid model using wavelet transform. *Int. J. Energy Res.* **2022**, *46*, 1766–1788. [\[CrossRef\]](#)
33. Jun, T.J.; Eom, Y.; Kim, D.; Kim, C.; Park, J.H.; Nguyen, H.M.; Kim, Y.H.; Kim, D. TRk-CNN: Transferable ranking-CNN for image classification of glaucoma, glaucoma suspect, and normal eyes. *Expert Syst. Appl.* **2021**, *182*, 115211. [\[CrossRef\]](#)
34. Ilesanmi, A.E.; Ilesanmi, T.O. Methods for image denoising using convolutional neural network: A review. *Complex Intell. Syst.* **2021**, *7*, 2179–2198. [\[CrossRef\]](#)

35. Bai, Y.; Liu, S.; He, Y.; Cheng, L.; Liu, F.; Geng, X. Identification of MOSFET Working State Based on the Stress Wave and Deep Learning. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 9003209. [[CrossRef](#)]
36. Ye, W.; Jiang, Z.; Li, Q.; Liu, Y.; Mou, Z. A hybrid model for pathological voice recognition of post-stroke dysarthria by using 1DCNN and double-LSTM networks. *Appl. Acoust.* **2022**, *197*, 108934. [[CrossRef](#)]
37. Liu, Z.; Wang, H.; Liu, J.; Qin, Y.; Peng, D. Multitask learning based on lightweight 1DCNN for fault diagnosis of wheelset bearings. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 3501711. [[CrossRef](#)]
38. Lara-Benítez, P.; Carranza-García, M.; Riquelme, J.C. An experimental review on deep learning architectures for time series forecasting. *Int. J. Neural Syst.* **2021**, *31*, 2130001. [[CrossRef](#)] [[PubMed](#)]
39. Latifoğlu, L. A novel combined model for prediction of daily precipitation data using instantaneous frequency feature and bidirectional long short time memory networks. *Environ. Sci. Pollut. Res.* **2022**, *29*, 42899–42912. [[CrossRef](#)]
40. Chen, B.; Zheng, H.; Wang, L.; Hellwich, O.; Chen, C.; Yang, L.; Liu, T.; Luo, G.; Bao, A.; Chen, X. A joint learning Im-BiLSTM model for incomplete time-series Sentinel-2A data imputation and crop classification. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *108*, 102762. [[CrossRef](#)]
41. Zhou, D.; Zhuang, X.; Zuo, H.; Cai, J.; Zhao, X.; Xiang, J. A model fusion strategy for identifying aircraft risk using CNN and Att-BiLSTM. *Reliab. Eng. Syst. Saf.* **2022**, *228*, 108750. [[CrossRef](#)]
42. Rangasamy, K.; As'ari, M.A.; Rahmad, N.A.; Ghazali, N.F. Hockey activity recognition using pre-trained deep learning model. *ICT Express* **2020**, *6*, 170–174. [[CrossRef](#)]
43. Zhou, T.; Lu, H.; Yang, Z.; Qiu, S.; Huo, B.; Dong, Y. The ensemble deep learning model for novel COVID-19 on CT images. *Appl. Soft Comput.* **2021**, *98*, 106885. [[CrossRef](#)] [[PubMed](#)]
44. Gao, J.; Song, Z.; Gui, J.; Yuan, S. Gas-bearing prediction using transfer learning and CNNs: An application to a deep tight dolomite reservoir. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 3001005. [[CrossRef](#)]
45. Saxena, A.; Goebel, K.; Simon, D.; Eklund, N. Damage propagation modeling for aircraft engine run-to-failure simulation. In Proceedings of the 2008 International Conference on Prognostics and Health Management, Denver, CO, USA, 6–9 October 2008; pp. 1–9.
46. Zhao, M.; Zhong, S.; Fu, X.; Tang, B.; Pecht, M. Deep residual shrinkage networks for fault diagnosis. *IEEE Trans. Ind. Inform.* **2019**, *16*, 4681–4690. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.