ISSN 2616-5767 Vol.6, Issue 12: 115-121, DOI: 10.25236/AJETS.2023.061217

Fault Diagnosis of Wind Turbine Based on CNN-LSTM Parallel Network Model

Peng Xin, Xun Zhang, Chenglei Yuan, Chaoran Li

Jilin Institute of Chemical Technology, Jilin, 132000, China

Abstract: *In response to the limited data feature extraction capability of a single neural network and data feature loss in the serial connection of multiple neural networks in fault diagnosis, a parallel network structure model comprising Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks is proposed, called CNN-LSTM parallel network model. The model's key is to simultaneously extract features from spatial and temporal dimensions of Supervisory Control and Data Acquisition (SCADA) data and make state judgments. Additionally, a partial ensemble learning meta-model was established to Identify fault types that are difficult to distinguish due to small differences in data. The CNN-LSTM parallel network model is employed for fault detection of the wind turbine using SCADA data. It's verified that the fault detection accuracy using the CNN-LSTM parallel network model is up to 99.60%, which is higher than the fault detection accuracy using single neural network models CNN and LSTM, as well as the CNN-LSTM serial connection model. Besides, the model outperforms the other models in terms of evaluation metrics, such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).*

Keywords: *Fault diagnosis; Convolutional Neural Networks; Long Short-Term Memory Network; SCADA data; Deep learning*

1. Introduction

With the continuous development of wind power technology, wind energy has been widely applied as a new type of clean energy [1]. A wind turbine is composed of diverse mechanical, electrical, and control components, and they operate continuously in challenging environments, which increases their susceptibility to various malfunctions. If left unaddressed, these faults can result in significant economic losses. In recent years, the theory of Deep Learning (DL) and deep network technologies have made significant breakthroughs in various fields[2]. Deep learning methods have been widely used in the fault diagnosis research of wind turbines [3,4].

Reference [5] proposed a new multivibrator fusion technology based on the Deep Belief Network (DBN), which uses different acceleration vibration signals of wind turbines as input vectors to diagnose bearing faults. Many researchers have studied wind turbine faults using SCADA data [6]. Reference [7] used CNN model to extract spatial features from SCADA data to identify the operating state of wind turbines. CNN is used for spatial feature extraction, but it has certain limitations in extracting features from time series data. Reference [8-9] proposed an improved CNN-LSTM serial connection model to classify and recognize 6 different operating states of wind turbine rolling bearings. The model utilizes CNN to extract spatial features from the data and then employs the LSTM network to extract temporal features from the extracted feature data. The drawback of this method is that CNN may discard some useful information between data, resulting in incomplete data obtained by the LSTM network, thereby affecting the recognition results.

A CNN-LSTM parallel network model is proposed. To solve the limitations of a single neural network and the issue of losing useful feature information when using serial combination methods. The model acquires temporal and spatial features of SCADA data in a supervised learning manner to determine the type of wind turbine fault. A partial ensemble learning meta-model is established, it can accurately recognize fault types that are difficult to distinguish due to small differences in data, thus improving the accuracy of fault diagnosis.

ISSN 2616-5767 Vol.6, Issue 12: 115-121, DOI: 10.25236/AJETS.2023.061217

2. CNN-LSTM parallel network model

2.1 CNN model

CNN consists of three components: convolutional layer, pooling layer, and fully connected layer. The function of the convolutional layer is to explore the intrinsic connections between data through local perception and weight sharing. The 1D-CNN model is used to process the one-dimensional SCADA data. The mapped features obtained from the convolution layer are then sent to the pooling layer for downsampling. Finally, the processed features are fed into the fully connected layer for recognition. The computational formula is as

$$X = [X_1, X_2, \ldots, X_k, \ldots, X_n], \tag{1}$$

$$X_k = \left[x_k^{(1)}, x_k^{(2)}, \ldots, x_k^{(j)}\right], \tag{2}$$

$$z_k = \varphi(W * X_k + b_k)\,, \tag{3}$$

where X is the dataset; n is the number of data in the dataset; j is the number of features in a single data; z_k is the convolutional output; $\varphi()$ is the activation function; W is the convolutional kernel; $*$ is the convolutional process; b_k is the bias term.

2.2 LSTM model

LSTM is considered a special type of RNN (Recurrent Neural Network) [10]. Its constituent is an Input Gate, Output Gate, and Forget Gate. These gates function are control and propagate information, addressing gradient vanishing and exploding during backpropagation in a time series, which are realized by (4), (5), (6), and (7). LSTM has advantages in handling time series prediction and classification problems. The basic unit of LSTM is shown in Fig. 1.

$$f_1 = \mathrm{sigmoid}\left(w_1 \begin{bmatrix} S_{t-1} \\ x_t \end{bmatrix} + b_1\right), \tag{4}$$

$$f_2 = \mathrm{sigmoid}\left(w_2 \begin{bmatrix} S_{t-1} \\ x_t \end{bmatrix} + b_2\right) * \tanh\left(\widehat{w}_2 \begin{bmatrix} S_{t-1} \\ x_t \end{bmatrix} + \hat{b}_2\right), \tag{5}$$

$$\mathrm{c_t} = f_1 * \mathrm{c_{t\text{-}1}} + f_2, \tag{6}$$

$$z_t{}^{(2)} = \mathrm{sigmoid}\left(w_3 \begin{bmatrix} S_{t-1} \\ x_t \end{bmatrix} + b_3\right) * \tanh(c_t), \tag{7}$$

where f_1 and f_2 are the calculation results of the Forget Gate and Output Gate at the time t , respectively; s_{t-1} is the "short-term memory" at the time $t-1$; x_t is the input content at the time t; c_{t-1} and c_t are the "long-term memory" at time $t-1$ and t, respectively; z_t is the output result at the time t; w_1,w_2,$w_3$and $\widehat{w}_2$ are the weight matrices of the Forget Gate, Input Gate, Output Gate, and cell state, respectively; b_1,b_2,$b_3, and\ \hat{b}_2$ are the bias terms of the Forget Gate, Input Gate, Output Gate, and cell state, respectively; sigmoid is a function that takes values between 0 and 1. It performs element-wise multiplication, masking out elements with a value of 0; tanh is the hyperbolic tangent function.

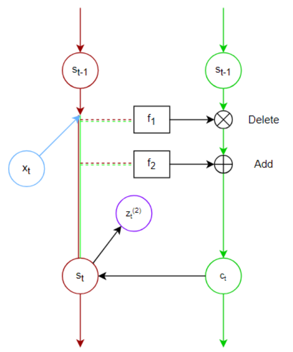


Figure 1: Structure diagram of the basic unit of LSTM

ISSN 2616-5767 Vol.6, Issue 12: 115-121, DOI: 10.25236/AJETS.2023.061217

2.3 Partial Ensemble Learning meta-model

A partial ensemble learning meta-model is used to reprocess small, difficult-to-distinguish fault types based on the predictions of multiple models [11]. In this paper, the prediction results of CNN and LSTM networks as input features to train the targeted meta-model. The entire process is shown in Fig. 2.

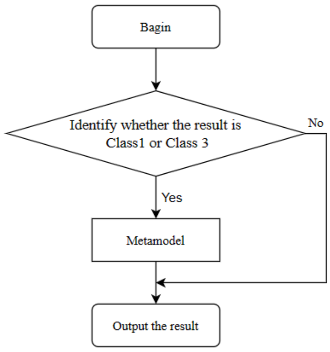


Figure 2: Flowchart of partial ensemble learning

2.4 CNN-LSTM parallel network model

The following research presents a CNN-LSTM parallel network model. The input layer, CNN module, LSTM module, partial ensemble learning meta-model, and output layer are the five distinct components of this new model. Fig. 3 depicts the overall structure.

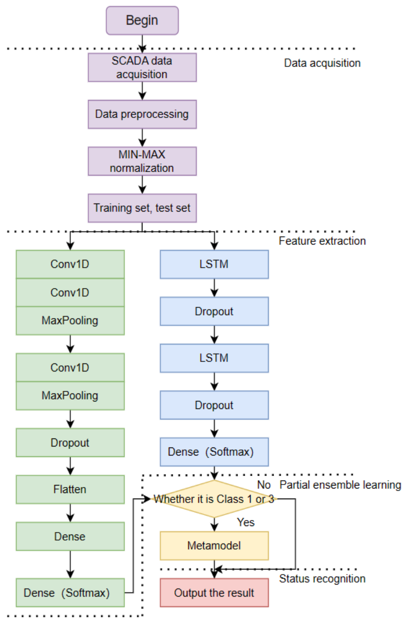


Figure 3: Block diagram of the overall structure of the CNN-LSTM parallel network model

3. Experimental Verification

3.1 Data Preprocessing

The research data in this thesis is derived from SCADA data from wind turbines in a wind farm in Jilin Province, China, in 2018. The data was collected every 10 seconds, yielding 10,800 data samples for 9 different fault categories, each comprising 1,200 data samples.

To divide the time series data for the LSTM model, we used a time window with a timing length of 5 and a step size of 1 [12]. The CNN model used the same number of samples as the LSTM model.

ISSN 2616-5767 Vol.6, Issue 12: 115-121, DOI: 10.25236/AJETS.2023.061217

Finally, the training and testing sets were roughly divided into an 8:2 ratio. The training set had 8,800 samples, whereas the testing set had 1,964 samples. The data was normalized using the Min-Max method. The method is given by

$$y' = \frac{y - x_{\min}}{x_{\max} - x_{\min}}, \tag{8}$$

where y' is the normalized data; y is the original data; $x_{\max}$ and $x_{\min}$ are the maximum value and minimum value in the original dataset, respectively.

3.2 Experimental setup

3.2.1 CNN structure parameter settings

This work investigates the effect of the number of layers in a convolutional layer on performance. It was found that the use of 3 convolutional layers is most appropriate. The size of the convolutional kernel is usually chosen between 3 and 5 based on previous experience [13]. Three identical 3x1 convolutional kernels were chosen based on the study's data properties. The number of filters in the first convolutional layer was set to 32 to better extract input information, and the following layers were doubled in number to increase network width.

3.2.2 Dropout parameter settings

The Dropout layer serves to prevent overfitting and reduce computational load. Setting different Dropout rates can lead to different performance effects of the model [8]. We conducted five sets of experiments using different dropout rates (0.1, 0.2, 0.3, 0.4, and 0.5), and the results are shown in Fig. 4.

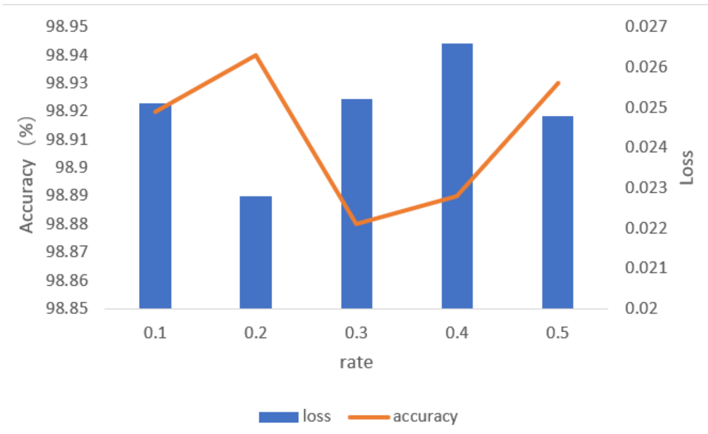


Figure 4: Comparison diagram of different dropout ratio results

The line graph indicates the accuracy, while the bar graph reflects the loss value, as illustrated in Fig 4. When the Dropout rate is 0.2, it obtains the maximum accuracy and has the smallest loss value. This means that a dropout rate of 0.2 produces the best model performance. As a result, the dropout rate in the study is set to 0.2.

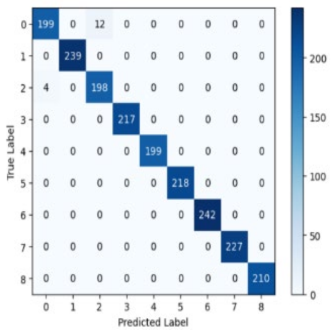
3.2.3 Comparative experiment settings

Four algorithmic models are used in comparison experiments with the same dataset. Experiment 1 used a single LSTM network model [10]. Experiment 2 used a 1D-CNN network model for temporal feature sampling [13]. Experiment 3 used a CNN network model for spatial feature sampling [14]. Experiment 4 used a CNN-LSTM serial network model [8,15]. Each model was run 5 times with 30 iterations per run. The results are shown in Table 3.

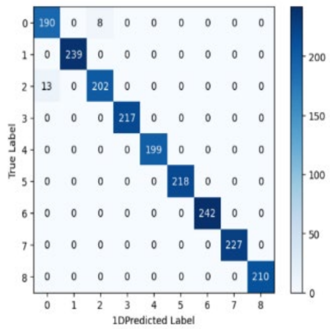
3.3 Experimental results and analysis

The recognition results of the two models in the fifth experiment are shown in Fig. 5.

ISSN 2616-5767 Vol.6, Issue 12: 115-121, DOI: 10.25236/AJETS.2023.061217



(a) CNN confusion matrix



(b) LSTM confusion matrix

Figure 5: Identify the resulting confusion matrix diagram

As illustrated in Fig 5, there were mistakes in the prediction of the first and third fault classes, whereas the accuracy for the other classes was 100%. As a result, the first and third fault classes were chosen as the targets for the partial ensemble learning meta-model. A binary meta-model was built specifically for these two fault classes. The binary meta-model, which is made up of three convolutional layers, achieved a recognition rate of 98.3% for the two fault classes. Fig. 6 shows the meta-model's recognition results.

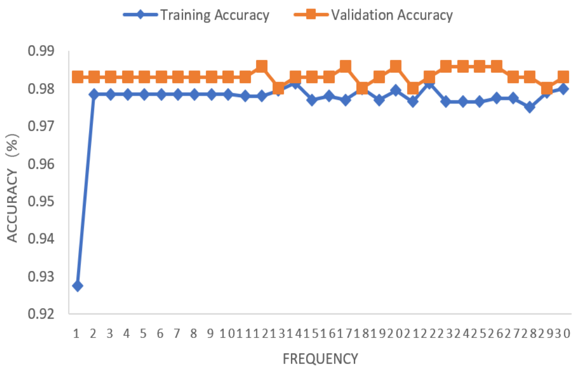


Figure 6: Metamodel recognition rate change chart

The formula for calculating the overall fault recognition rate is given as

$$A_1 = (1 - (N_2 * (1 - A_2))/N_1) * 100\%, \tag{9}$$

where A_1 and A_2 are the overall model recognition rate and meta-model recognition rate, respectively; N_1 and N_2 are the number of overall samples and meta-model input overall samples, respectively.

Table 1 shows the results of 10 experiments performed on the CNN-LSTM parallel network model.

Table 1: Experimental results of CNN-LSTM parallel network model

Numble	Test Loss	Test Accuracy(%)
1	0.0095	99.762
2	0.0083	99.864
3	0.0082	99.887
4	0.0096	99.864
5	0.0095	99.841
6	0.0095	99.841
7	0.0094	99.875
8	0.0083	99.864
9	0.0085	99.864
10	0.0082	99.909
average	0.0089	99.857

To showcase the superior diagnostic accuracy of the CNN-LSTM parallel network model, it was compared to four different algorithmic models, as detailed in Section 2.2.4 Experimental Setup. Each of these models was executed 5 times with 30 iterations and the results of these executions are shown in Table 2.

Table 2: Compare experimental results

Numble	Name	Training Accuracy(%)	Test Accuracy(%)
1	LSTM	99.12	98.93
2	1D-CNN	99.17	99.19
3	CNN	99.22	99.19
4	CNN-LSTM Serial network model	99.11	99.39
5	CNN-LSTM parallel network model	99.80	99.86

These model evaluations are shown in Table 3.

Table 3: Model evaluation

Name	MSE	RMSE	MAE	r2_score
LSTM	0.001554	0.039426	0.003415	0.983250
1D-CNN	0.001062	0.032582	0.002368	0.988595
CNN	0.001125	0.033539	0.002701	0.987837
CNN-LSTM Serial network model	0.000978	0.031267	0.002203	0.989017
CNN-LSTM Parallel network model	0.000654	0.025583	0.001658	0.992782

Experimental Results Analysis: From Table 2 and Table 3, it can be observed that the proposed CNN-LSTM parallel network model has a significant advantage in terms of fault type recognition accuracy. It also outperforms the other four models to varying degrees in all four evaluation criteria. Compared to the other four control experiments, the proposed method reduces MSE by 57.9%, 38.4%, 41.8%, and 33.1% respectively; RMSE by 35.1%, 21.5%, 23.7%, and 18.1% respectively; MAE by 51.4%, 30.0%, 38.6%, and 24.7% respectively; and increases r2_score by 0.97%, 0.42%, 0.50%, and 0.38% respectively. In conclusion, the CNN-LSTM parallel network model performs exceptionally well.

4. Summary

In-depth research was conducted to address the limitations of a single neural network in extracting multi-dimensional data features, as well as the drawbacks of the sequential combination approach that may overlook useful information and make it difficult to distinguish between similar fault types. The following conclusions were drawn:

(1) Constructing a CNN-LSTM parallel network model allows for feature extraction from both the spatial and temporal aspects of the data. This maximizes the utilization of data information, avoids wasting data information, and improves fault recognition accuracy.

(2) By using a partial ensemble learning meta-model, the prediction results of the two models are further processed to obtain new judgment results. The approach addresses the issue of low recognition accuracy caused by small differences between fault types in multi-type recognition. As a result, the recognition accuracy is further improved.

References

[1] LI Songfeng. Development Status and Development Trend of China's Wind Power Equipment Industry [J]. Value Engineering, 2019, 38(33): 37-38.
[2] Wang J, Ma Y, Zhang L, et al. Deep learning for smart manufacturing: Methods and applications[J]. Journal of Manufacturing Systems, 2018, 48.
[3] PENG Cheng ,LI Fengjuan,JIANG Jinyuan. Summary of fault diagnosis and prediction methods based on deep learning [J]. Modern Electronics Technique, 2022, 45(03): 111-120.
[4] WEN Chenglin, LÜ Feiya. Review on Deep Learning Based Fault Diagnosis [J]. Journal of Electronics & Information Technology, 2020, 42(01): 234-248.
[5] Jie T, Yilun L, Dalian Y. Bearing Fault Diagnosis Based on Deep Belief Network and Multisensor Information Fusion[J]. Shock and Vibration, 2016, 2016(Pt.6).
[6] LIANG Tao ,HUANG Yue. The Design of SCADA System for Wind Power Plants [J]. Hydropower and New Energy, 2023, 37(01): 1-4+10.

ISSN 2616-5767 Vol.6, Issue 12: 115-121, DOI: 10.25236/AJETS.2023.061217

[7] Ren Yan. Research on anomaly identification method of wind turbine generator based on convolution neural network(CNN) [D]. Northeastern University, 2018.
[8] CAO Zhengzhi, YE Chunming. Application of Improved CNN-LSTM Model in Fault Diagnosis of Rolling Bearings [J]. Computer Systems & Applications, 2021, 30(03): 126-133.
[9] Zhou B, Khosla A, Lapedriza À, et al. Learning Deep Features for Discriminative Localization[J]. CoRR, 2015, abs/1512.04150.
[10] GUO Xiaojing, YIN Yuxuan, YUN Yujing. Aeroengine Life Prediction Method Based on Improved LSTM [J]. Machine Tool & Hydraulics, 2022, 50(20): 185-193.
[11] Jinming Huang,Meng Du. Build a used car price prediction model based on Stacking ensemble learning [J]. CHINA SCIENCE AND TECHNOLOGY INFORMATION,2023(14):88-89.
[12] Wang Tingshao, Ji Tianyao, Jiang Yuzi, Wang Jin, et al. Fault diagnosis method for wind turbines based on de-noise auto-encoder and one-dimensional convolution network [J]. Electrical Measurement & Instrumentation, 2023, 60(01): 87-93+173.
[13] CHEN Haipeng, ZHOU Yuehao, WANG Chenlu, WANG Junqi, HAN Hao, et al. Economic Analysis of System Spinning Reserve Based on Improved CNN-LSTM Short Term Wind Power Prediction [J]. High Voltage Engineering, 2022, 48(02): 439-448.
[14] XIN Peng, YANG Kaixun, WEN Xiaoqiang. Research on Wind Turbine Fault Diagnosis Method based on Improved SE-CNN [J].Journal of Jilin Institute of Chemical Technology,2023,40(01):34-40.
[15] LU Tongchao, MA Lele, FEI Ziyan, et al. Fault Identification of Aviation Engine State Perception Network Based on GMM－CNN－LSTM [J]. Industrial Control Computer, 2021, 34(10): 15-17.