

Article

Fault Diagnosis Based on an Approach Combining a Spectrogram and a Convolutional Neural Network with Application to a Wind Turbine System

Wenxin Yu ^{1,2,*}, Shoudao Huang ¹  and Weihong Xiao ³

¹ College of Electrical and Information Engineering, Hunan University, Changsha 410082, China; hsd1962@hnu.edu.cn

² School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, China

³ School of Information Engineering, Hunan Province Cooperative Innovation Center for Wind Power Equipment and Energy Conversion, Xiangtan University, Xiangtan 411105, China; 201610171777@smail.xtu.edu.cn

* Correspondence: slowbird@sohu.com; Tel.: +86-181-6363-7607

Received: 28 August 2018; Accepted: 24 September 2018; Published: 26 September 2018



Abstract: To investigate problems involving wind turbines that easily occur but are hard to diagnose, this paper presents a wind turbine (WT) fault diagnosis algorithm based on a spectrogram and a convolutional neural network. First, the original data are sampled into a phonetic form. Then, the data are transformed into a spectrogram in the time-frequency domain. Finally, the data are sent into a convolutional neural network (CNN) model with batch regularization for training and testing. Experimental results show that the method is suitable for training a large number of samples and has good scalability. Compared with Back Propagation Neural Network (BPNN), Support Vector Machine (SVM), Extreme Learning Machine (ELM), and other fault diagnosis methods, the average diagnostic correctness rate is higher; so, the method can provide more accurate reference information for wind turbine fault diagnosis.

Keywords: spectrogram; convolutional neural network; wind turbine; fault diagnosis

1. Introduction

With the growing shortage of energy, wind power generation has become the preferred alternative energy source. Whether wind turbine generators, which are the core devices of wind power generation technology, can be operated safely for a long period of time is directly related to benefits for power generation companies and local economic development. Fatigue life determines the length of the service life of a wind turbine under normal working conditions, and it is very important to predict the fatigue life accurately. References [1,2] present some new approaches based on stochastic processes, apply these methods to estimate the fatigue loads on wind turbines, and achieve significant results. In reference [3], a method for estimating the fatigue life of components by combining the non-homogeneous Poisson stochastic process function with the theory of concomitant damage is put forward. It is applied to the estimation of the fatigue life of wind turbine blades. The results show that this method is reliable and effective, which provides a new path for the fatigue reliability design of wind turbine blades. Reference [4] proposes a novel scheme for extending the lifetime of a wind energy conversion system (WECS) by integrating an online damage evaluation model into a control strategy for structural load reduction.

Although scholars have conducted extensive research on wind power life and proposed many ways to extend its operation, in the actual process, due to the complex structure of wind power

equipment and the extremely harsh working conditions, a fault diagnosis involves the comprehensive application of many professional disciplines, such as modern dynamics, methods for recognizing the failure mechanisms of wind power generation equipment, fault diagnosis methods, signal acquisition, signal processing, pattern recognition, and network communication. Thus, it is very difficult to conduct fault diagnosis. There are many kinds of fault diagnosis methods for wind power generation equipment. These include the three major types of diagnostic methods: mechanical signal fault diagnosis, electrical signal fault diagnosis, and artificial intelligence fault diagnosis. These methods use measurement equipment to obtain the corresponding electrical or non-electrical quantity signals and then use advanced high-efficiency signal processing technology to analyze and process the signals, after which feature information is accurately extracted to reflect the nature of the fault and the degree of failure.

Mechanical Signal Method

Faults, such as generator bearing failures, shaft misalignments, unbalanced rotor masses, loose seats, and rotor eccentricities, can be diagnosed by monitoring generator vibration, temperature, and speed signals. This is currently the most mature and most widely discussed technology. This fault diagnosis method has been successfully applied to the health monitoring of and fault diagnosis in key components, such as gearboxes, bearings, and blades, in wind turbines.

In 2014, the study in reference [5] proposed a noise reduction method for the characteristics of strong nonlinear noise in wind turbine vibration signals. They then constructed effective features based on the information after de-noising and modeled wind turbines using manifold learning algorithms allowing for early weak fault diagnosis. In 2016, the study in reference [6] proposed that the signals of vibrations of wind turbines in the fault signal extraction process are weak. It was proposed that the vibration signal-to-noise ratio of the motor can be increased and then de-noised using the empirical wavelet transform, after which the fault feature can be extracted. To monitor wind turbine vibrations, in 2017, reference [7] presented normal behaviour models to predict tower-top accelerations and drivetrain vibrations. In reference [8], the implementation and performance analysis of a vibration-response-based Gaussian Mixture Model Random Coefficient (GMM-RC) model-based structural health monitoring (SHM) framework for structures with time-dependent dynamics under significant uncertainty (operational and environmental) have been presented. The experimental results demonstrate that the GMM-RC model-based framework has consistently good performance.

Electrical Signal Method

An electrical fault in a generator is usually identified by monitoring the stator coil temperature, stator voltage, stator and rotor current, generator output rate, rotor speed, and other parametric signals. Relative to the vibration signal, the fault-related signals contained in electrical signals (such as current signals) are often relatively weak, and they are usually overwhelmed by electrical signals inherent in the motor and random noise. Thus, the signal-to-noise ratio is relatively low, and it is difficult to compare extracted fault features. At present, common methods of electrical fault diagnosis include stator current detection, partial discharge monitoring, flux detection, and current harmonics.

In 2014, reference [9] proposed a new online approach based on the instantaneous power spectrum of the rotor to detect doubly fed induction generator stator winding faults. A theoretical analysis showed that, compared with the common current and voltage spectrum analysis methods, the instantaneous power spectrum has an advantage insofar as external sensors and hardware devices are not needed. Moreover, the spectrum has immunity to interference. Once some slip tracking or calculation errors occur, the characteristic frequency will not change because the characteristic frequency is not affected by slip. In 2015, upon examining the problem that internal faults in doubly fed induction generators are not easily diagnosed and identified, reference [10] proposed a method for fault diagnosis in and positioning of doubly fed wind generators based on electromagnetic and wavelet transforms from the perspective of electromagnetic changes. In 2016, the study in reference [11] used

a synchronous sampling method to extract fault features from current signals of varying conditions and then used correlation dimension analysis to quantitatively analyze the different faults of wind turbines. In 2017, aimed at the nonstationary and nonlinear characteristics of wind turbine vibration signals, reference [12] proposed a novel fault diagnosis method based on integral extension load mean decomposition multiscale entropy and a least squares support vector machine. In reference [13], the detection of electrical asymmetry in rotors in wind turbine WT doubly fed induction generators (DFIGs) has been investigated using a test rig under three different driving conditions, and then an effective extended Kalman filter (EKF)-based method was proposed to iteratively estimate the fault signature components (FSCs) and track their magnitude.

Artificial Intelligence method

The research method of this type of fault diagnosis method is to analyze one or more types of signals of wind turbines, construct a set of high-dimensional statistical features in the time domain, frequency domain, and time-frequency domain, and fuse the features with machine learning methods. The method then implements dimensionality reduction, classification and visualization analysis, and equipment fault diagnosis.

Reference [14] proposed a fault detection method for the main bearing of wind turbines based on existing supervisory control and data acquisition data using an artificial neural network. In reference [15], an artificial neural network and an empirical mode decomposition (EMD)-based condition-monitoring approach using the Simulink fatigue, aerodynamics, structures, and turbulence (FAST) software and the TurbSim software were presented. Reference [16] proposed a wind turbine fault diagnosis method based on a diagonal spectrum and clustering binary tree support vector machines (SVMs). Reference [17] proposed a radially uniform (RU) design to sample representative datasets from a large volume of wind turbine data to build accurate data-driven models. Five algorithms, namely neural networks, multivariate adaptive regression splines, support vector machines, k-nearest neighbors, and linear regression, were applied to model the wind turbine power output, drivetrain vibratory acceleration, and tower vibratory acceleration based on training and sampled datasets. Reference [18] presented an approach for fault diagnosis in wind turbines based on a multi-class fuzzy support vector machine classifier. Reference [19] investigated a data-driven fault detection and isolation design based on the fusion of several classifiers for a wind turbine benchmark second challenge. The proposed method was robust against different operational conditions and measurement errors. A novel wind turbine weak feature extraction method based on Cross Genetic Algorithm (CGA) optimal Mexican-Hat Wavelet (MHW) is proposed in [20]. Experimental results showed that the proposed method can not only suppress interruptions due to strong background noise but also can extract weak wind turbine features effectively.

Although there are already some methods available for fault diagnosis in wind turbines, there are still several problems:

(1) Fault “big data” must be handled effectively. Due to the large scale of the wind turbines being monitored, each wind turbine needs more measuring points, and the sampling frequency of each measuring point is high. In addition, data collection from the start of service to the end of life takes a long time, and thus the monitoring system acquires a large amount of data; consequently, health monitoring has entered the era of “big data”. Finding information from these data and efficiently and accurately identifying the health status of wind turbines represent new problems we are currently facing.

(2) There are difficulties in the extraction and selection of features. Large-sized wind turbines are electromechanical devices that operate under variable working conditions. Their operating laws are strongly nonlinear and nonstationary. Because of their complex operating conditions, fault feature information can easily be overwhelmed by environmental and operational noise. Wind turbines are generally installed on the sea or grasslands, where there are high winds. The surrounding environment is complex, and the operation of a wind turbine is typically nonlinear and nonstationary. To solve

the problem of nonlinear fault diagnosis in equipment, there is an urgent need to further study the extraction of primitive features that can sufficiently predict the performance of equipment in a short period of time. The optimization of features is the focus of the current research, the goal of which is to reduce the feature dimension of the extraction, reduce the amount of necessary calculations, and improve the accuracy of fault diagnosis. However, the commonly used methods have their limitations. They still have to rely on experience to determine the parameters; therefore, the training speed is slow and the calculation cost is high.

(3) There is a lack of high-precision fault diagnosis methods. The existing fault diagnosis classification methods and models are highly complex in principle and have large numbers of parameters. Some of them require considerable experience to construct and train. Thus, it is difficult to achieve high diagnostic accuracy in practical applications.

Deep learning stems from the field of bionics on a simulated brain system. It is actually a multi-layer sensor neural network that constructs multiple hidden layers. It is also called a deep neural network. Deep learning simulates the learning process of the human brain through deep neural networks and can form knowledge and judgments directly from raw data. These processes are all based on the multi-level abstraction mechanism of the human brain, which eliminates the need for artificial feature extraction and selection. This avoids the complexity and uncertainty associated with traditional feature extraction processes and enhances the intelligence of the recognition process.

Convolutional neural networks (CNNs) play an important role in deep learning. A CNN is a machine learning model under deep-supervised learning. This paper presents a fault diagnosis method for wind turbine gearboxes based on a convolutional neural network. The introduction of batch normalization reduces the over-fitting phenomenon in the feature extraction process and shortens the training time. The structure of this paper is as follows. In Section 1, we briefly discuss the spectrum used in this paper. Sections 2–4 describe the convolution neural network and batch regularization and analyze the design of our fault diagnosis method. Section 5 discusses an example of fault diagnosis in wind turbines and gives the results for and analysis of this example. Finally, conclusions are drawn in Section 6.

1.1. Spectrogram

Speech signals are a typical example of a nonstationary signal. However, the nonstationary nature of speech is caused by the physical motion of the organ through which a person speaks. This process is slower than the velocity of the acoustic vibration. It can be assumed that speech is smooth on the timescale of 10~30 ms. Fourier analysis is a powerful means of analyzing the steady-state characteristics of linear systems and stationary signals. Short-time Fourier analysis, also known as a time-dependent Fourier transform, is a method that can address nonstationary signals using steady-state analysis under short-term stationarity assumptions.

Let the discrete time-domain sampling signal be $x(n)$, $n = 0, 1, 2, \dots, N - 1$, where n is the time-domain sampling point number, and N is the signal length. If the signal is framed, then $x(n)$ is denoted as $x_n(m)$, $n = 0, 1, 2, \dots, N - 1$, where n is the frame number, m is the time sequence number of the frame synchronization, and N is the frame length (the number of sampling points within one frame). $\{x(n)\}$ short-time Fourier transforms are given by:

$$X(n, k) = \sum_{m=0}^{N-1} w(m)x(m)e^{-j\frac{2\pi km}{N}} \quad (1)$$

where $w(m) = 0.54 - 0.46 \cos \frac{\pi m}{N}$ is the Hamming window.

Then, $|X(n, k)|$ is the discrete Fourier transform used to obtain the short-term amplitude spectrum estimation.

Therefore, the power spectral function is:

$$p(n, k) = |X(n, k)|^2 = (X(n, k)) \times (\text{conj}(X(n, k))) \quad (2)$$

where $p(n, k)$ is a two-dimensional non-negative, real-valued function, which is the Fourier transform of the short-term autocorrelation function of signal $x(n)$. With n as the abscissa and k as the ordinate, the value of $p(n, k)$ is expressed as the gray-level composition of the image, which is a phonogram.

$$l(n, k) = \frac{B(n, k) - Base}{Max(B(n, k) \forall (n, k)) - Base} \quad (3)$$

$$\text{and } B(n, k) = \begin{cases} P(n, k), & P(n, k) > Base \\ Base, & P(n, k) < Base \end{cases} .$$

1.2. Pseudo-Color Spectrogram

For pseudo-color mapping on $p(n, k)$, a pseudo-color chromatogram with a higher resolution and better visual effects can be obtained. The maximum value $p_{\max}(n, k)$ of $p(n, k)$ is mapped to the normalized level of 1, and the minimum value $p_{\min}(n, k)$ is mapped to the normalized level of 0. Then, $p(n, k)$ is linearly mapped to the level L_i between 0 and 1. The computer monitor can then display it in pseudo-color mode according to the value of L_i in the phonogram. To obtain a better display effect, one can select the appropriate reference base value. Then, one can limit the values that are less than the base value to the reference level and map the values greater than the base value linearly to between 0 and 1 normalized color values. The mathematical representation of the color value matrix $L = \{l(n, k)\}$ is as follows:

$$l(n, k) = \frac{B(n, k) - Base}{Max(B(n, k) \forall (n, k)) - Base} \quad (4)$$

$$\text{and } B(n, k) = \begin{cases} P(n, k), & P(n, k) > Base \\ Base, & P(n, k) < Base \end{cases} .$$

2. Convolutional Neural Network

A convolutional neural network includes input layers, convolutional layers, pooling layers, fully connected layers, and classifiers. The convolution layer contains several sets of parameters that can be learned by the convolution kernel, which is at the core of convolutional neural network training. The convolution of the current layer is performed by convoluting a set of pictures, and then a new feature map is obtained through the activation function. The new pixel output in the convolution layer can be calculated by formula (5):

$$\begin{aligned} x_j^{l-1} &= f(u_j^l) \\ u_j^l &= \sum_{i \in M_j} x_i^{l-1} * W_{ij}^l + b_j^l \end{aligned} \quad (5)$$

where $f(\cdot)$ represents an activation function, x_i^{l-1} represents a pixel value of the previous layer of the feature image, W_{ij}^l represents a convolution kernel, M_j represents a subset of the feature image of the upper layer involved in the operation, b_j^l represents the offset term, and l represents the layer number.

Convolutional neural network learning and the training of convolution kernel parameters are achieved via a gradient back-propagation algorithm, which is a supervised learning algorithm. Assuming that the training of the network is not perfect, the training output and the actual output of the input signal will have some errors. These errors will be propagated through the layers in the gradient descent algorithm, and the network parameters will be updated layer-by-layer. Let the cost function of sample (x, y) be $J(W, b; x, y)$:

$$J(W, b; x, y) = \frac{1}{2} \|h_{W, b}(x) - y\|^2. \quad (6)$$

The actual increase in the L_2 -norm control over the fitting phenomenon will then be:

$$J(W, b) = J(W, b : x, y) + \frac{\lambda}{2} W^T W \quad (7)$$

where $h_{W,b}(x)$ is the output of input sample x after passing through the network, y is the labeled true value of the sample, and λ is the control intensity. The goal of training, obviously, is to make the cost function as small as possible. To update the weights of each layer W and offset b , any weight (assuming layer l) is updated as follows:

$$\begin{cases} W_{ij}^l = W_{ij}^l - \alpha \frac{\partial}{\partial W_{ij}^l} J(W, b) \\ b_i^l = b_i^l - \alpha \frac{\partial}{\partial b_i^l} J(W, b). \end{cases} \quad (8)$$

3. Batch Regularization

For deep neural networks, as the number of network layers continues to deepen, small perturbations in the previous training parameters are amplified, resulting in a change in the probability distribution of the current layer's characteristics. This is inconsistent with the probability distributions of the initial layer's features and makes the previous learning rates and weights no longer applicable. Therefore, this paper uses batch regularization to transform the probability distribution of each layer into a standard normal distribution to avoid parameter perturbation. In so-called batch regularization, the batch is a batch of samples in a random gradient. Regularization means that the probability distribution of each dimension of each layer's activation value is transformed into a stable probability distribution with a mean of zero and a variance of one.

Assuming that the input of a layer is $x = (x^{(1)} x^{(2)} \dots x^{(k)})$, k is a dimension, and a batch of samples is $B = \{x_1 x_2 \dots x_m\}$. The formula for batch regularization is as follows:

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad (9)$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (10)$$

$$\widehat{x}^{(k)} = \frac{x^k - \mu_B}{\sigma_B} \quad (11)$$

$$y^{(k)} = \gamma^{(k)} \widehat{x}^{(k)} + \beta^{(k)}. \quad (12)$$

In the above equations, $x^{(k)}$ is the k -dimension of the input x , μ_B represents the expectation of the sample set B , σ_B^2 represents the variance of the sample set B , $\widehat{x}^{(k)}$ represents the regularization result of the input x , and $y^{(k)}$ represents the result after the $x^{(k)}$ -batch regularization. $\gamma^{(k)}$ and $\beta^{(k)}$ represent the parameters to be learned and are solved by training iterations to increase the stability of the output probability distribution.

4. Implementation of the Algorithm

In this paper, the spectrogram and convolutional neural network are applied to analog circuit fault diagnosis. The algorithm includes two main parts: data acquisition and processing and training and testing the classification model. The algorithm flow chart is shown in Figure 1 and is outlined as follows:

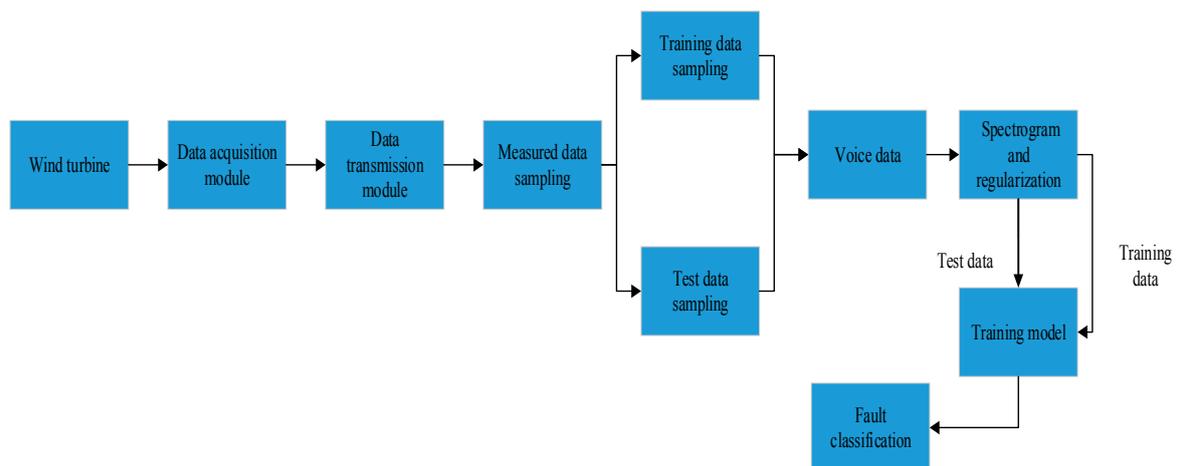


Figure 1. The algorithm flow chart.

(1) Select the fault through testing the wind turbine: set the fault and sample the fault data, then save the experimental raw data in text format.

(2) The original data for linguistic analysis are used to determine the low-level structure of the sentence and the composition of each word phoneme. Then, address the text data corresponding to the word or phrase extracted from the speech synthesis library. Transform the linguistic description into speech waveforms, and then save the data in WAV format.

(3) The speech data are analyzed to obtain the sampling signal. Then, use the discrete Fourier transform to obtain a short-term amplitude spectrum estimation.

(4) Calculate the power spectral function of the signal $p(n, k)$.

(5) Draw a two-dimensional image with the time n as the abscissa, k as the ordinate, and $p(n, k)$ as the value of the gray level. Then, the pseudo-color mapping on $p(n, k)$ can be obtained with a high resolution for a better visual effect of the pseudo-color chromatogram.

(6) The convolutional neural network with batch regularization trains and classifies faults.

5. Simulation Experiment and Discussion

To further verify the validity of the diagnostic model, this paper tests the 40 groups of wind turbine samples listed in [21]. Among the 40 groups of samples, each fault condition has 6 samples so that there are 240 sample data points and 8 failure categories. Table 1 lists some of the sample data.

Table 1. Data tables for some of the test samples.

Sample Fault Type	First Input	Second Input	Third Input	Fourth Input	Fifth Input	Sixth Input	Theoretical
	Parameter	Parameter	Parameter	Parameter	Parameter	Parameter	Output
Normal	-0.17570	3.203604	2.006353	-0.63307	-1.77743	-1.58773	0
	-0.15062	2.80914	1.8841	-0.46833	-1.61768	-1.53783	0
Gear wear	-0.13816	2.615372	1.843098	-0.37860	-1.45544	-1.40816	1
	-0.15903	3.074887	1.922826	-0.46939	-1.73008	-1.37910	1
Gear pitting	-0.23421	3.78483	2.143404	-1.37863	-1.80621	-1.52705	2
	-0.22194	3.616806	2.070489	-1.28463	-1.74630	-1.48818	2
Broken teeth	-0.11560	2.241891	1.640091	-0.28922	-1.19496	-1.25642	3
	-0.10466	2.018099	1.557112	-0.26057	-1.02524	-1.20379	3
Bearing inner ring fault	-0.14154	2.841607	1.980599	-0.19554	-1.72882	-1.51999	4
	-0.21623	4.14302	2.146404	-0.76796	-2.40306	-1.49559	4
Bearing outer ring fault	-0.13699	2.526633	1.834072	-0.37943	-1.30093	-1.45033	5
	-0.14466	2.569866	1.86299	-0.47906	-1.26774	-1.49878	5
Cage fault	-0.43317	9.501146	2.87289	-1.10344	-5.74552	-0.72027	6
	-0.36711	7.125565	2.78909	-1.41219	-4.12520	-1.51255	6
Rolling body fault	-0.13271	2.790753	1.829006	-0.16272	-1.72798	-1.36752	7
	-0.13261	2.649846	1.81267	-0.25177	-1.58189	-1.41919	7

In this algorithm, the parameter set in each part is as follows:

(1) Compared to other convolutional neural network models, the VGG16 mode [19] uses convolutional layers of multiple smaller convolution kernels instead of one convolutional core with a larger number of convolutional layers. On the one hand, the parameters can be reduced, and on the other hand, more nonlinear mappings can increase the network's ability to fit. Therefore, this article uses VGG16 as a training model.

(2) To effectively improve the performance of the convolutional neural network, a relatively high classification accuracy can be obtained with relatively few training samples. In this paper, we will generate a spectrogram for each fault condition. Thus, a total of 40 spectral maps are generated, of which 32 pictures are used for training and the remaining 8 pictures are used for network testing.

(3) The model consists of 13 convolution layers, 3 full connection layers, and 1 output layer. The 13 convolution layers are divided into 5 large convolution layers, and each pair has a pool layer between the elements in it.

(4) The specific settings of the parameters are as follows: the number of training steps is 15,000, the activation function is a ReLU (rectified linear) function, the gradient loss method is the application of the loss function, and the learning rate is 0.01.

Figures 2 and 3 show the short-term amplitude spectrum estimation and power spectral function graphs for the 240 samples, respectively.

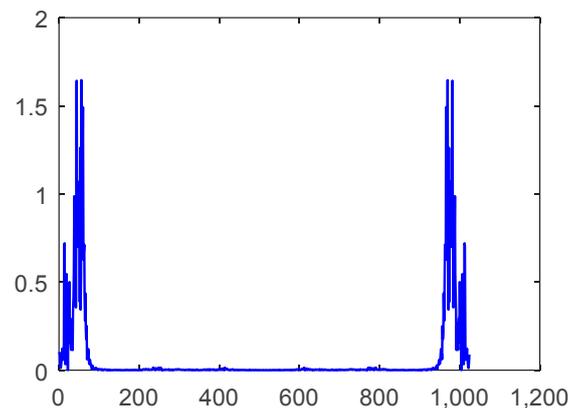


Figure 2. Short-term amplitude spectrum estimation graphs.

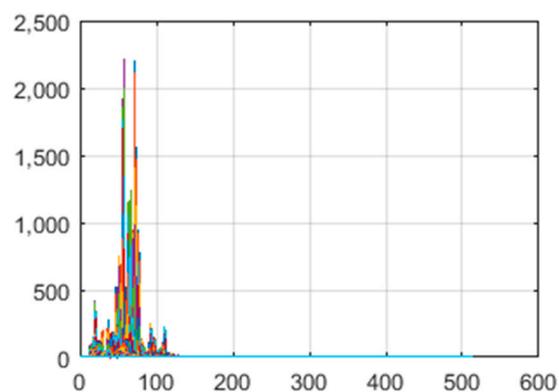


Figure 3. Power spectral function graphs.

After the above steps, in order to reduce the training time, 240 samples will generate 40 spectra and pseudo-color chromatograms. Figures 4 and 5 show spectrograms and pseudo-color spectra for the following sample data: $(-0.14154, 2.841607, 1.980599, -0.19554, -1.72882, -1.51999)$.

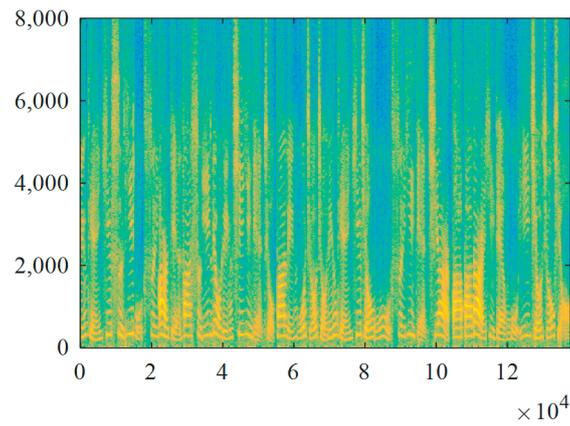


Figure 4. Spectrogram.

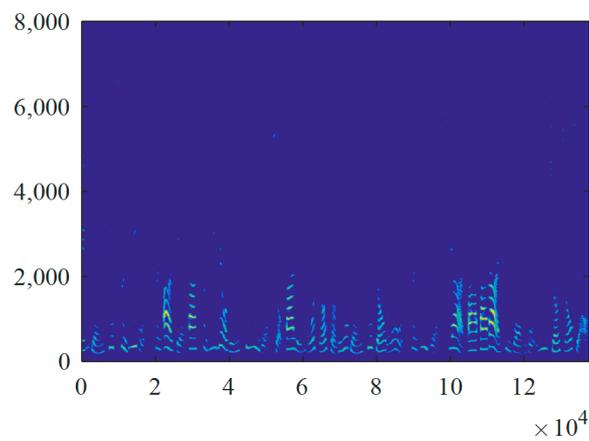


Figure 5. Pseudo-color spectrum.

Figures 6 and 7 show the accuracy and loss function curves of the training and the test data, respectively. From Figures 6 and 7, it can be seen that both the accuracy curve and the loss function curve fluctuate little during training and testing. In addition, the convergence speed is fast, and the convergence value is ideal.

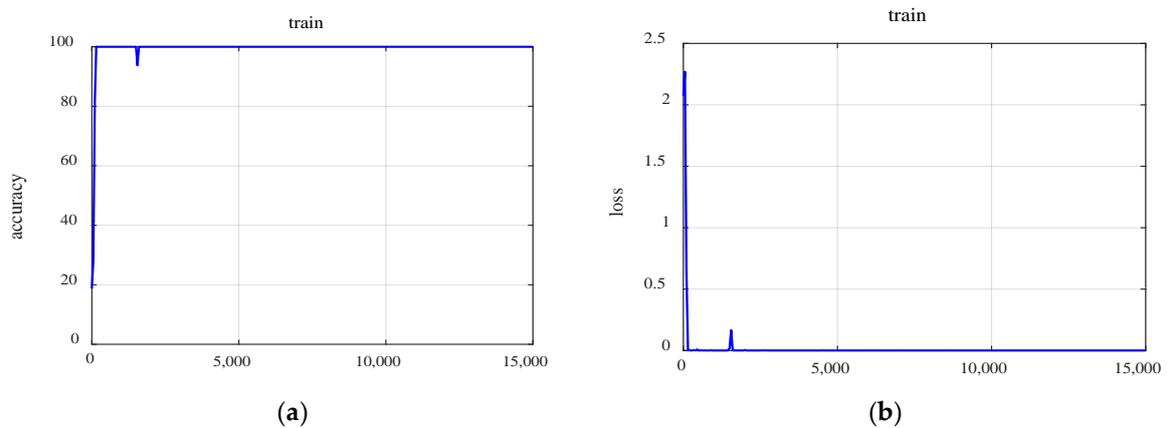


Figure 6. Accuracy and loss function curves for the training data. (a) Accuracy curve; (b) Loss function curve.

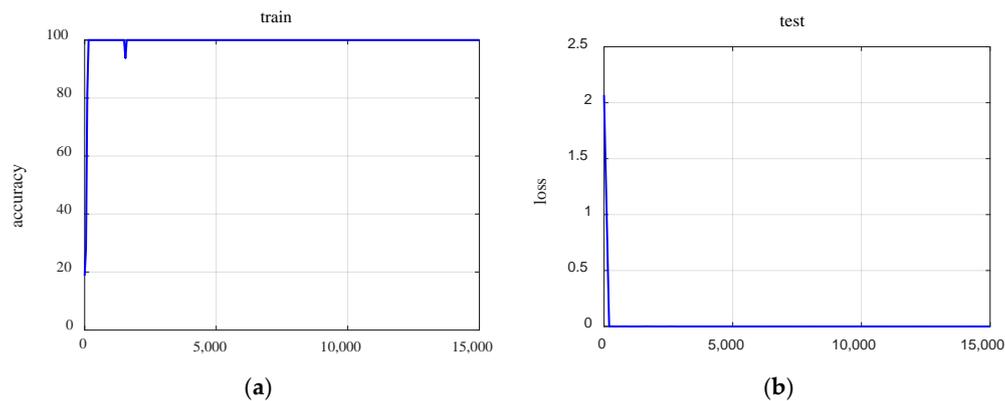


Figure 7. Accuracy and loss function curves for the test data. (a) Accuracy curve; (b) Loss function curves.

This shows that our proposed method, when applied to the 240 groups of wind turbine samples with 8 fault types, has 100% classification accuracy. From Table 2 [22], only the classification accuracy of the SVM is close to that of the proposed algorithm, and our algorithm is far more accurate than BPNN, ELM, Fixsize Sequence Extreme Learning Machine (FSSELM), and Incremental Extreme Learning Machine (IELM). This fact illustrates that, although the number of samples is small, the diagnostic accuracy of this method is relatively high. Therefore, the applicability of this algorithm to wind turbine fault diagnosis is very high.

Table 2. Accuracy comparison among six classifiers.

Algorithm	Training Accuracy	Testing Accuracy
BPNN	67.39%	61.2%
SVM	100%	98.75%
ELM	97.81%	97.5%
FSSELM	98.6%	98.48%
IELM	98.78%	98.54%
Proposed algorithm	100%	100%

In the above experiment with a small sample, the fault diagnosis correctness rate reached 100%. In order to verify the practicability of this method, we continued to increase the number of samples. As the number of samples in the training set and the test set continue to increase, the results show that the average correctness rate of this method remains high. Since batch regularization has been introduced and each fault condition is represented as a spectrogram, the method has the advantages of a short training time, a fast diagnosis speed, and a low error rate.

6. Conclusions

This paper emphasizes a way to achieve wind turbine fault diagnosis. First of all, the time-frequency spectrum is used for the extracted features, and then a convolutional neural network with batch regularization is applied as an optimized classifier, which is discussed in detail. Experimental results show that the method has the advantages of a short training time, a fast diagnosis speed, and a low error rate. So, the algorithm can be used for wind turbine fault diagnosis and will be highly valuable for relevant applications.

Author Contributions: W.Y. conceived of the project and proposed the methodological framework and implementation roadmap; S.H. reviewed and improved the methodological framework and implementation algorithm; W.X. revised the text and the simulations.

Acknowledgments: This work was supported by the National Key Research and Development Program of China (2016YFF0203400), by the China Postdoctoral Science Foundation-funded project No. 2017M622574, and by the Hunan Provincial Department of Education Science Research Project (Grant No. 16C0639).

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

References

1. Lind, P.G.; Wächter, M.; Peinke, J. Fatigue loads estimation through a simple stochastic model. *Energies* **2014**, *7*, 8279–8293. [[CrossRef](#)]
2. Rinn, P.; Heißelmann, H.; Wächter, M.; Peinke, J. Stochastic method for in-situ damage analysis. *Eur. Phys. J. B* **2013**, *86*, 1–5. [[CrossRef](#)]
3. Mi, L.; Cheng, H.; Quan, L. Estimation of Blade Fatigue Life of Wind Turbine Based on Poisson Stochastic Process. *J. Mech. Eng.* **2016**, *52*, 134–139. [[CrossRef](#)]
4. Njiri, J.G.; Beganovic, N.; Söffker, D. Consideration of Lifetime and Fatigue Load in Wind Turbine Control. *Renew. Energy* **2018**, *131*, 818–828. [[CrossRef](#)]
5. Tang, B.; Song, T.; Li, F.; Deng, L. Fault diagnosis for a wind turbine transmission system based on manifold learning and Shannon wavelet support vector machine. *Renew. Energy* **2014**, *62*, 1–9. [[CrossRef](#)]
6. Chen, J.; Pan, J.; Li, Z.; Zi, Y.; Chen, X. Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals. *Renew. Energy* **2016**, *89*, 80–92. [[CrossRef](#)]
7. Lind, P.G.; Vera-Tudela, L.; Wächter, M.; Kühn, M.; Peinke, J. Normal Behaviour Models for Wind Turbine Vibrations: Comparison of Neural Networks and a Stochastic Approach. *Energies* **2017**, *10*, 1944. [[CrossRef](#)]
8. Avendaño-Valencia, L.D.; Fassois, S.D. Damage/fault diagnosis in an operating wind turbine under uncertainty via a vibration response Gaussian mixture random coefficient model based framework. *Mech. Syst. Signal Process.* **2017**, *91*, 326–353. [[CrossRef](#)]
9. Ma, H.; Zhang, Z.; Shi, W.; Chen, J.; Chen, T. Doubly-fed Induction Generator Stator Fault Diagnosis Based on Rotor Instantaneous Power Spectrum. *Autom. Electr. Power Syst.* **2014**, *38*, 30–35.
10. Li, H. Blade Imbalance Fault Diagnosis of Doubly Fed Wind Turbines Based on Stator Current Feature Analysis. *Autom. Electr. Power Syst.* **2015**, *39*, 32–37.
11. Jin, X.; Qiao, W.; Peng, Y.; Cheng, F.; Qu, L. Quantitative Evaluation of Wind Turbine Faults under Variable Operational Conditions. *IEEE Trans. Ind. Appl.* **2016**, *52*, 2061–2069. [[CrossRef](#)]
12. Gao, Q.W.; Liu, W.Y.; Tang, B.P.; Li, G.J. A novel wind turbine fault diagnosis method based on intergral extension load mean decomposition multiscale entropy and least squares support vector machine. *Renew. Energy* **2018**, *116*, 169–175. [[CrossRef](#)]
13. Ibrahim, R.; Watson, S.; Djurovic, S.; Crabtree, C. An Effective Approach for Rotor Electrical Asymmetry Detection in Wind Turbine DFIGs. *IEEE Trans. Ind. Appl.* **2018**, *65*, 8872–8881. [[CrossRef](#)]
14. Zhang, Z.-Y.; Wang, K.-S. Wind turbine fault detection based on SCADA data analysis using ANN. *Adv. Manuf.* **2014**, *2*, 70–78. [[CrossRef](#)]
15. Malik, H.; Mishra, S. Artificial neural network and empirical mode decomposition based imbalance fault diagnosis of wind turbine using TurbSim, FAST and Simulink. *IET Renew. Power Gener.* **2017**, *11*, 889–902. [[CrossRef](#)]
16. Liu, W.; Wang, Z.; Han, J.; Wang, G. Wind turbine fault diagnosis method based on diagonal spectrum and clustering binary tree SVM. *Renew. Energy* **2013**, *50*, 1–6.
17. Tan, M.; Zhang, Z. Wind Turbine Modeling With Data-Driven Methods and Radially Uniform Designs. *IEEE Trans. Ind. Appl.* **2016**, *12*, 1261–1269. [[CrossRef](#)]
18. Hang, J.; Zhang, J.; Cheng, M. Application of multi-class fuzzy support vector machine classifier for fault diagnosis of wind turbine. *Fuzzy Sets Syst.* **2016**, *297*, 128–140. [[CrossRef](#)]
19. Pashazadeh, V.; Salmasi, F.R.; Araabi, B.N. Data Driven Sensor and Actuator Fault Detection and Isolation in Wind Turbine Using Classifier Fusion. *Renew. Energy* **2018**, *116*, 99–106. [[CrossRef](#)]
20. Ren, H.; Liu, W.; Jiang, Y.; Su, X. A novel wind turbine weak feature extraction method based on Cross Genetic Algorithm optimal MHW. *Measurement* **2017**, *109*, 242–246. [[CrossRef](#)]
21. Wu, B.; Xi, L.; Fan, S.; Zhan, J. Fault Diagnosis for Wind Turbine Based on Improved Extreme Learning Machine. *J. Shanghai Jiao Tong Univ. (Sci.)* **2017**, *22*, 466–473. [[CrossRef](#)]
22. Simonyan, K.; Zisserman, A. Very Deep Con-volutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.

