applied sciences

MDPI

Article

Data-Driven Model Space Method for Fault Diagnosis of High-Speed Train Air Brake Pipes

Weigang Ma *, Jing Wang *, Xin Song, Jiaqi Qi, Yaping Yu and Dengfang Hu

Faculty of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China; songxin@xaut.edu.cn (X.S.); 2211221140@stu.xaut.edu.com (J.Q.); 2211221076@stu.xaut.edu.com (Y.Y.); 2211220048@stu.xaut.edu.cn (D.H.)

* Correspondence: mwg_markey@xaut.edu.cn (W.M.); wangjingsgz@163.com (J.W.); Tel.: +86-15991728902 (W.M.); +86-19834405728 (J.M.)

Abstract: A data-driven fault diagnosis method is proposed in this study to address the challenge of handling a large volume of pressure data in the air brake pipe of high-speed trains. The suggested method utilizes a BP (back propagation) neural network to transform the time series pressure data into model elements in the model space, ensuring simplicity and stability. Various fitting functions, including Fourier basis, Gaussian basis, polynomial basis, sine basis, and others, are employed to accurately fit the pressure curve of the air brake pipe. The fault diagnosis process involves two steps: classifying the fault based on an optimal approximation equation and diagnosing it by analyzing the topological relationship of the model elements in the model space. The proposed method achieves an average fault diagnosis accuracy of 89.8%, with high accuracy rates for different fault states: 98% for normal state, 88% for blockage state, 84% for leakage state, and 96% for compressor fault state. Compared to the hidden Markov model method, the proposed method improves the average diagnostic accuracy by 2% for known working conditions and 4.87% for all working conditions, demonstrating its effectiveness and reliability. The fault diagnosis of the air brake tube in high-speed trains is of great significance, which aims to realize accurate fault diagnosis and prediction through sensor data monitoring and signal processing technology, so as to ensure the safe operation of high-speed trains. These studies provide an important theoretical and practical basis for the improvement and application of fault diagnosis methods.

Keywords: data driven; model space; BP neural network; fault diagnosis; air brake pipe

check for updates

Citation: Ma, W.; Wang, J.; Song, X.; Qi, J.; Yu, Y.; Hu, D. Data-Driven Model Space Method for Fault Diagnosis of High-Speed Train Air Brake Pipes. *Appl. Sci.* **2023**, *13*, 8335. https://doi.org/10.3390/app13148335

Academic Editor: Suchao Xie

Received: 22 May 2023
Revised: 3 July 2023
Accepted: 17 July 2023
Published: 19 July 2023

1. Introduction

High-speed trains have been rapidly developed and are widely used worldwide due to their fast, safe, and comfortable features [1]. The air brake system plays an important role in high-speed trains because of its high safety and reliability. As the operating mileage increases, components such as air ducts, air cylinders, and compressors in the air brake system inevitably age and become damaged. If not repaired in time, it can easily cause failure, which is a huge challenge for the safe operation of high-speed railways. Therefore, research on the air brake system in high-speed trains has important practical significance.

The self-learning ability of neural networks can summarize the rules from a large amount of data. Even if the operating mechanism of the neural network is not clear, as long as the neural network is trained, a better input and output model can be established. Through self-memory and analysis in the neural network, the data types are classified and processed accordingly. The unique structure and learning ability of a neural network make it possible to combine the pressure data of high-speed train air brake pipes and improve the network according to the measured air pipe pressure data, so as to ensure the accuracy of the predicted data. At present, there are many prediction-based neural network models, including the single model and hybrid optimization model.

The single model includes a radial basis neural network, limited learning neural network, and a BP neural network. Hebei Railway University used the self-developed settlement deformation test instrument based on hydraulic differential measurements to create a real-time acquisition and automatic monitoring system using a wireless sensor network on the Harbin–Dalian passenger dedicated line and after the completion of the Beijing–Shanghai high-speed rail track laying.

By eliminating interference noise, Kalman filter extracts realistic and effective information, which greatly improves the efficiency in data processing. Kalman filtering is used to estimate and predict the deformation, which fully reflects the motion state of the deformed body, provides reliable data for the project, and ensures the safe construction of the project.

The hybrid optimization network overcomes the shortcomings of a single model and networks learn from each other. Research on various combination models is widely carried out, such as the grey system and neural network combination model, grey system and ARMA combination model, wavelet analysis and neural network combination model, and so on. In recent years, crash prediction models (CPMS) have become a useful tool for railway planning, mainly for accident prediction; Zhai et al. [2] used the optimized BP neural network to predict the future track of the ship. The results show that the prediction accuracy is higher than the traditional prediction model, and the practicability is strong. In addition, there are high-order neural network (HON-PILE model) and SMT-GP combination model, BPARX combination model, etc. These provide new impetus and new reference for the development of the combined model.

At present, scholars at home and abroad have performed a lot of research on the fault diagnosis method of the EMU braking system and have achieved fruitful results. Tai et al. [3] introduced the structure and composition of the EMU brake control system and the network topology of the fault detection and diagnosis module, proposed a fault detection algorithm, and conducted experiments on a type of EMU brake system in China to verify the effectiveness of the algorithm. Zhou et al. [4] proposed a fault detection index with mutual variable variance, completed the feasibility analysis of fault detection, and proposed a fault isolation method. Finally, experiments were carried out on the EMU brake test bench at Qingdao Sifang Rolling Stock Research Institute Co., LTD, to verify the effectiveness of the fault diagnosis method. Sang et al. [5] proposed an EMU brake cylinder early leakage fault diagnosis method for the brake cylinder in the EMU brake system and conducted experiments on the brake test bench to verify the effectiveness of the fault diagnosis method. Huang et al. [6] used a neural network to diagnose sensor faults in the EMU braking control system. Two different learning methods were used in the neural network algorithm used in the fault diagnosis process, and the diagnosis results obtained by the two different learning methods were compared based on the electric locomotive air brake system, the important modules were modelled, such as air cylinders and air ducts, and their characteristics were analyzed. In addition to using the simulation model and intelligent algorithm to explore the train air brake system, Sun et al., based on the analysis of the characteristic parameters of the control valve and brake cylinder, proposed the air brake based on the train. The multi-parameter mathematical simplification method of the feature calculates the train braking characteristics [7]. Wei et al. [8]. developed the train air brake system combined with longitudinal dynamics to simulate the important parameters of various trains in the running process. Du et al. proposed a vibration-based fault detection and isolation algorithm based on the three fault characteristics of the master cylinder pressure, vehicle longitudinal acceleration, and wheel speed. The frequency response difference between the braking cycle and the normal driving cycle (non-braking) is used to improve the robustness of the algorithm. The experimental results show that this method can better diagnose the thickness change fault in the car body and isolate the fault to each car body rotation angle. Ji et al. [9] proposed a fault detection method combining the four-stage division idea with the new combined statistics for fault detection in the high-speed train air brake system and compared it with the traditional IVV method to prove the superiority of the proposed method. Sang et al. conducted a linear transformation of the data from the actual characteristics for three early faults in the air brake system: sensor deviation fault, leakage fault, and brake cylinder assembly. After obtaining the approximate stationary property, the multimodal data is mapped to a tight domain, and the fault signal ratio (FSR) is used to reflect the sensitivity of the proposed detection statistic to the fault. Finally, experiments were carried out on the braking test platform to verify the effectiveness of the proposed strategy [10]. Huang et al. [11] proposed an improved fully integrated empirical mode decomposition based on adaptive noise (ICEEMDAN) and one-dimensional convolutional neural network (1-D CNN) fault diagnosis method, which can simultaneously identify the fault state in high-speed train bogies and the location of fault components. Wang et al. [12] proposed a mixed dual discriminator GAN (M-D2GAN), which uses different methods to generate different types of variables. The GAN model was appropriately modified to add a normal class discriminator. Liu et al. [13] effectively integrated the weighted feature strategy and cost-sensitive learning into the multi-kernel support vector machine model, which solved the problems caused by class imbalance and signal heterogeneity.

The above research has achieved great results in the modeling and simulation of high-speed train air brake systems, air cylinders, ducts, and other important air brake components, but the results from research on the fault diagnosis are limited. Since the high-speed train air brake system is a complex engineering system, this will have a negative impact on the method of relying on traditional mathematical models for fault diagnosis, which will affect the diagnostic effect. In many traditional fault diagnosis methods, only a small number of known fault types are assumed, and no new fault types are mentioned. This argument is flawed [14]. Due to factors such as technological progress, new working environment, and different usage methods, there may be various unexpected failures in the air brake pipeline in high-speed trains in reality. If it only depends on the known fault type, the traditional fault diagnosis method cannot accurately determine the cause of the fault in the face of new fault types, resulting in inaccurate diagnosis results or even an ineffective diagnosis. Therefore, in order to improve the accuracy and reliability of fault diagnosis, a more flexible and adaptive method is needed to identify the known fault types, and new fault types can be found and diagnosed to improve the comprehensiveness and accuracy in fault diagnosis.

Aiming at the above problems, this paper proposes a model space based high-speed train air duct fault diagnosis method based on the existing research. Through data learning, function approximation, and other methods, the measured wind pipe pressure data in the time series space is transformed into the model elements in the continuous functional space, the element neighborhood is calculated, and then the topological relationship between the elements is used to diagnose the fault. At the same time, this paper defines three known fault types based on the existing experimental results and builds fault libraries based on these [15]. When an unrecognizable fault occurs, it is added to the fault pool according to the cause of the fault.

Fault diagnosis has a wide range of different research applications, such as the following scenarios. Automotive industry: Fault diagnosis is widely used in the automotive industry, including engine, transmission system, brake system, suspension system, and other fault diagnosis. Possible faults are diagnosed and fixes are recommended by the monitoring vehicle sensor data and system status. Energy field: In the process of energy production and supply, fault diagnosis can be used for fault monitoring and prediction of energy systems, such as power plants, wind farms, and solar farms. It can help detect faults in advance and take corresponding maintenance measures to ensure the reliable operation of the energy system. Manufacturing: In the manufacturing industry, fault diagnosis can be used for the fault detection and diagnosis of equipment and machines. Through the analysis of sensor data and signals during the production process, equipment status can be monitored in real time and potential failures can be detected in time, thereby reducing downtime and maintenance costs. Aerospace industry: In the aerospace sector, fault diagnosis is critical to ensuring flight safety. It can be used for fault detection and prediction in aircraft systems, including engines, hydraulic systems, electrical systems, etc. By monitoring aircraft sensor data and system status in real time, faults can be accurately diagnosed, and necessary measures can be taken to ensure the safe flight of the aircraft. Medical field: In medical devices and medical systems, fault diagnosis can be used to detect and diagnose equipment failures and ensure the safety and reliability of medical processes. For example, in medical imaging equipment, fault diagnosis can be used to detect image quality problems or equipment failures and provide recommendations for repair. Internet of Things (IoT): With the development of the Internet of Things technology, a large number of sensors and devices are connected to the network; fault diagnosis plays an important role in the Internet of Things system. It can be used to monitor and diagnose failures in various IoT devices, including smart home devices, smart city infrastructure, and more.

2. Model Space

Traditional fault diagnosis methods are mainly based on mathematical models. Specifically, by comparing the model with the actual calculation results [11], the time series data can better explain the fault diagnosis based on real-world scenarios. Therefore, in most traditional fault diagnosis methods, the establishment of mathematical models mainly depends on the processed time series data. However, due to the traditional fault diagnosis method using time series mathematical model to characterize the fault characteristics, if the time series data samples are not rich and the characteristics are not obvious, the model fitting effect is not ideal and the generalization performance is poor [12,13].

The model space is to transform a model layer above the time series data, transform the time series data into model elements in the model space, and replace the time series data with the model elements. The subsequent learning strategies are also directly applied to the model space instead of the time series space [12,13]. Replacing time-series data with model elements can make the data description more streamlined and stable, and the model space is a function space composed of all time series data involved in transforming model elements, so the model elements become point sets in the model space [12,13]. This idea can ensure that the learning strategy has a higher approximation accuracy in dynamic and uncertain environments [12,13]. The current thinking of the model space has also received extensive attention. Chen et al. used this method to solve various fault diagnosis problems in different fields, such as the Van der Boer oscillator and Barcelona water supply network system. Chen et al. uses the sliding window to convert all time series data into multiple model elements in the model space and uses the classification method to diagnose faults. In contrast, this paper converts the duct pressure data of the same working condition into data through data learning and function approximation. A model element in the model space calculates the neighborhood of the element and solves the problem in fault diagnosis through the neighborhood relationship. The model space frame is shown in Figure 1.

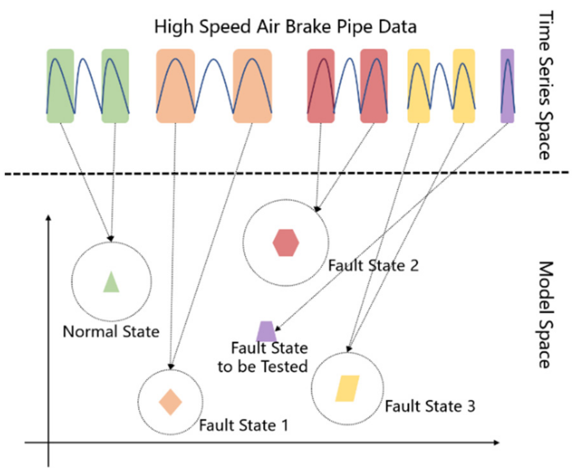

Figure 1. Framework of the model space.

In this paper, the duct pressure data used for conversion to model elements is obtained through BP neural network learning. The BP neural network is a "black box" for the user, and the learning results cannot be directly used for fault diagnosis. Therefore, you need to choose the appropriate method to describe the learning results.

First, in the continuous functional space, find a suitable set of basis and coefficients to express BP neural network learning results as function expressions. In this paper, four base (Fourier basis, Gaussian, polynomial basis, sinusoidal) design experiments are selected for the characteristics of duct pressure data, and the multi-order function approximation equation is calculated based on each basis. Finally, the optimality is selected by the fit K with the base and equation coefficients. The formula for calculating the fit K is as shown in Formula (1):

$$\mathrm{K} = \mathrm{SSE} + \mathrm{RMSE} + (1 - \mathrm{R\_SQUARE}) + (1 - \mathrm{ADJUSTED\_R\_SQUARE}) \quad (1)$$

where SSE represents the variance and RMSE represents the root mean square error, R-square represents the determined coefficient, and the adjusted R-square represents the corrected determined coefficient. The condition for selecting the best approximation equation using the fit K is that the order of the approximation equation is as small as possible, and K is suitable for selecting the basis and coefficient of the equation when it is suitable for the order of diagnosis.

After determining the optimal basis and coefficient, the function's best approximation equation is obtained. For the two model elements F_1 and F_2 in the model space, the norm between them is defined as shown in Formula (2):

$$L(F_1, F_2) = \|F_1 \cdot F_2\| \quad (2)$$

When the fault data to be tested is brought in, first calculate the norm between the fault data to be tested and the known fault model elements (air duct normal, air leakage fault, blocking fault, compressor fault) according to Formula (2), and then pass the known fault. The topological relationship of the model elements determines the type of fault. If the norm falls within the neighborhood of a known fault model element, it is considered that the fault type represented by the fault data to be tested belongs to the known fault type; if the norm does not fall within the neighborhood of all known fault model elements, it may be determined. The fault type represented by the data to be tested is an unknown fault, that is, the fault type indicated by the fault data to be tested belongs to a new fault.

The neighborhood definition of the model elements in the model space is shown in Equation (3), where R is the neighborhood of a known fault model element, and n is the number of measured duct pressure cycle data participating in the BP neural network learning of the model element. The number of measured duct pressure cycle data sets participating in the BP neural network learning of the model element, $G(x)$, is the best approximation equation for the model element, and $f(x)$ is the best approximation of the number of different measured duct pressures participating in the BP neural network learning of the model element. In the formula, $f_1(x)$ represents the best approximation equation for the first set of measured duct pressure cycle data, and $f_2(x)$ represents the best approximation equation for the second set of measured duct pressure data, and so on.

$$R = \frac{1}{n}\sum_{i=1}^{n} L(G(x), f_i(x)) \quad (3)$$

The fault diagnosis in the model space is shown in Formula (4), where $Class(j)$ represents the fault diagnosis result and j represents the number of known fault model elements. $G_j(x)$ represents the best approximation equation for the jth known fault model element, and $U(x)$ is the best approximation equation for the fault model element to be tested.

$$Class(j) = \begin{cases} 1, & L\big(G_j(x), U(x)\big) \le R_j \\ 0, & L\big(G_j(x), U(x)\big) > R_j \end{cases} \quad (4)$$

When *Class*(j) is 1, it indicates that the fault type represented by the data to be tested belongs to the known fault j. When *Class*(j) is 0, it indicates that the fault type represented by the data to be tested does not belong to the known fault j, only when there is $L >$ for all j's. R indicates that the type of fault indicated by the data to be tested belongs to an unknown fault.

3. Steps in Fault Diagnosis for Air Brake Pipes

In the fault diagnosis for the duct, since the entire diagnostic process requires two spaces (sequence space, model space), the diagnosis process is divided into two steps: data conversion in the time series space and fault diagnosis in the model space. The steps in high-speed train air brake pipe fault diagnosis is shown in Figure 2.

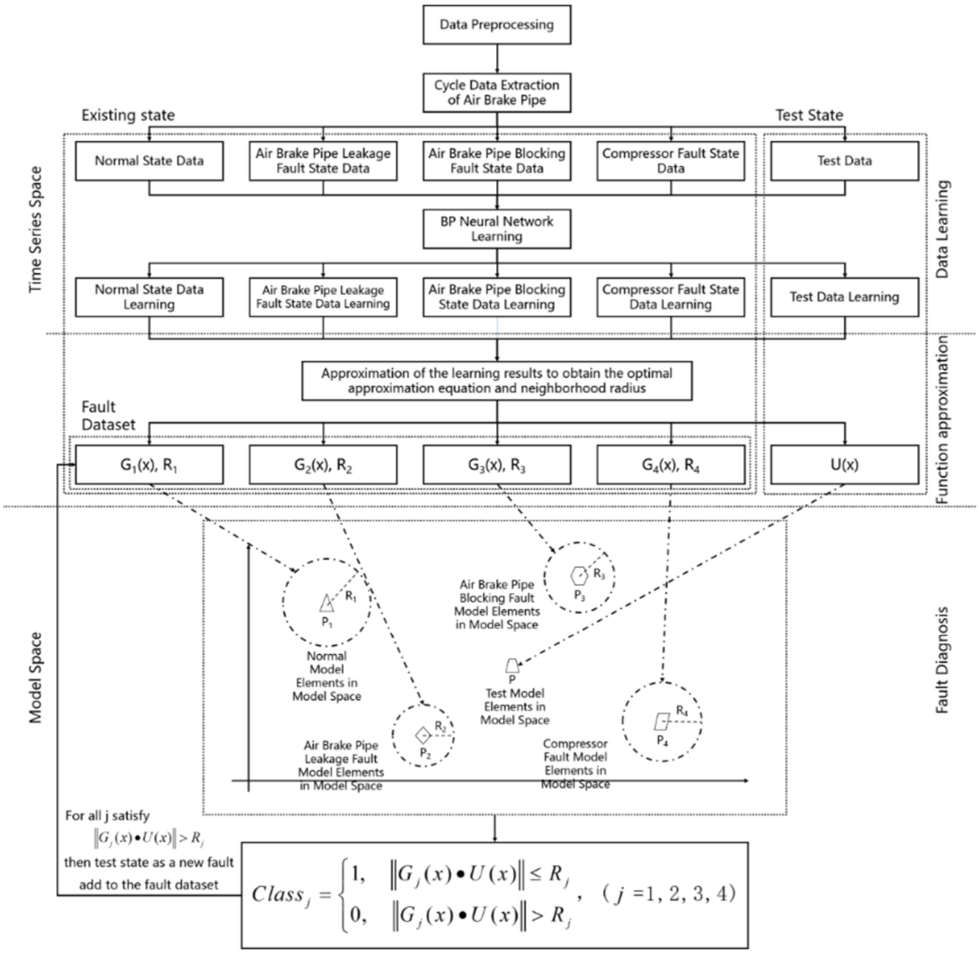

Figure 2. Steps in high-speed train air brake pipe fault diagnosis.

Data conversion in time series space includes data learning and function approximation. Data learning is divided into data preprocessing, air duct work cycle data extraction, and BP neural network learning. In actual operation, because the air duct pressure data has the characteristics of high redundancy and vacancy data, data preprocessing is required to obtain better diagnosis results. The type of duct failure is hidden in practice. No data can directly reflect the fault in the duct, but the duct pressure data shows periodic changes with the air charge and exhaust in the duct. The change corresponds to the type of duct failure. Unfortunately, the duct pressure data for the same fault type is different in terms of cycle and pressure values. This makes it impossible to mathematically establish separate models for different faults. Because the BP neural network has a strong nonlinear mapping ability and proves its effectiveness in fault diagnosis problems in various fields, BP neural network learning can learn different duct pressure data from the same fault as a duct pressure curve through which this type of fault is represented [13–17]. This is also the necessity of data extraction in the duct work cycle and BP neural network simulation. There are also other deep learning methods that play a guiding role in the fault diagnosis in high-speed trains [17–28].

Function approximation is to transform the learning results in the BP neural network into model elements in model space through a set of optimal bases and coefficients, and calculate the neighborhood of the model elements. This is because the learning result in the BP neural network is still a scatter value, and the inter-model norm cannot be calculated. Only by finding an appropriate set of bases and coefficients to represent the scatter value as an approximation equation can the fault diagnosis be performed by the norm. Model elements obtained by computing known fault function approximations can be saved as a fault library.

Fault diagnosis in the model space is achieved by the topological relationship between the model elements. Through data learning and function approximation in the known fault model, the normal model element P1, P2, P3, and P4 in the duct, and the neighborhood R1, R2, R3, and R4 in the model element are obtained.

When the fault data to be tested is input, data learning of the fault data to be tested is first performed, the data to be measured is preprocessed, the vacancy data is filled by the interpolation method, the data redundancy is reduced by using the smoothing function, and the air duct with the completion period is extracted. Regarding pressure data, the BP neural network is used to learn the duct pressure data.

Then, the function of the fault data to be tested is approximated by the function approximation to approximate the optimal equation U(x).

Finally, data transformation and fault diagnosis in the model elements are carried out, and the model element P to be tested in the model space is calculated. The norm L1, L2, L3, and L4 between the model element P to be tested and the known fault model element P1, P2, P3, and P4 is calculated by Formula (2). If the norm falls within the neighborhood of a known fault model element, the fault type represented by the data to be tested belongs to the known fault, and if the norm does not fall in all known within the neighborhood of the fault model element, the fault type represented by the data to be tested is added to the fault pool as a new fault.

4. Implementing Fault Diagnosis for Air Brake Pipes

In this chapter, the model space method proposed in this paper is used to simulate the fault diagnosis data and function approximation of the high-speed train air brake pipe, and the experimental results are visually displayed to verify the effectiveness and efficiency of the method.

4.1. Data Simulation

Considering the complexity of the high-speed train air brake system, according to the characteristics of the duct pressure data, a four-layer BP neural network is designed, which includes an input layer, two hidden layers, and one output layer. The network topology is shown in Figure 3. Among them, the input layer is the first layer of the neural network, which receives raw data or preprocessed feature vectors as inputs. The parameters of the input layer mainly include the input node and input feature vector, and each input feature corresponds to an input node, which is used to receive input data. The original data or features are converted into a numerical vector representation as the value of the input node, that is, the input feature vector. The hidden layer is a collection of one or more layers of neurons located between the input layer and the output layer. It is used to process input data and extract key feature representations. The parameters of the hidden layer mainly include a hidden node, weight, bias, and activation function, wherein each hidden node receives the output of the previous layer, calculates the weighted sum, and then obtains the output value through the activation function. The weight represents the strength of the connection between the input node and the hidden node and is used to calculate the weighted sum. Bias is a constant term for each hidden node and is used to adjust the activation threshold of the hidden node. The hidden layer usually uses nonlinear activation functions (such as Sigmoid, ReLU, etc.) for weighting and nonlinear mapping to increase the expressability of the network. The output layer is the last layer of the neural network, which produces the output of the network. The parameters of the output layer mainly include an output node, weight, bias, and activation function. Each output node corresponds to an output result, which is used to represent the prediction or classification result of the network. The weight represents the strength of the connection between the hidden node and the output node and is used to calculate the weighted sum. Bias is a constant term for each output node and is used to adjust the activation threshold of the output node. The activation function of the output layer is usually selected based on the specific problem, such as the Sigmoid function for binary classification problems and the Softmax function for multi-classification problems. In the training process of the BP neural network, the data set is usually divided into three parts: training set, verification set, and test set. This study uses cross-validation techniques to more accurately evaluate the performance of different network structures and parameter settings. By dividing the data set into multiple training/validation subsets and performing training and validation on different subsets, more reliable performance evaluation results are obtained. In this study, the training set accounted for 60%, the verification set accounted for 10%, and the test set accounted for 30%. Since the number of neurons directly affects the network training time, it is verified that the number of neurons in the hidden layer 1 is 20, and the number of neurons in the hidden layer 2 is 40.

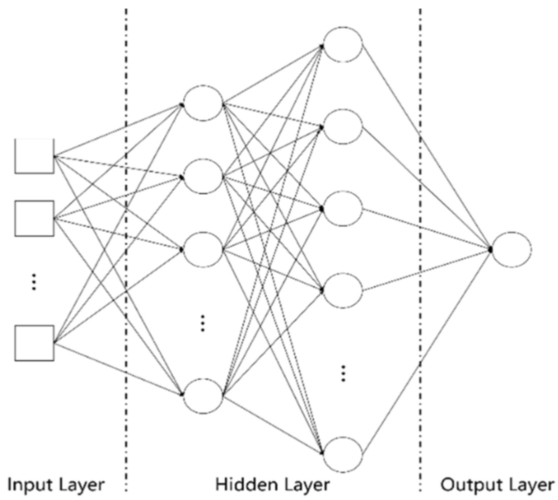

Figure 3. Structure of the BP neural network.

The optimal parameters for the number of layers and the number of neurons in a BP neural network is an iterative process that requires constant trial and evaluation. This study uses a larger data set, which is more suitable for deep networks because deep networks have greater expressiveness and learning ability and are able to learn more complex patterns and features. The number of neurons in the input layer should match the dimension of the input feature, that is, one input neuron for each input feature. The number of neurons in the hidden layer is usually determined by experiment and validation. Generally speaking, increasing the number of hidden layer neurons can improve the expression ability of the network, but it also easily leads to overfitting. This study selected the optimal number by gradually increasing the number of neurons and observing performance changes on the validation set. The number of neurons in the output layer depends on the type of problem. For the multi-classification problems involved in this study, the number of neurons in the output layer should be equal to the number of classes.

The same type is selected (normal duct, air duct leakage, duct obstruction, compressor failure), and the data on the duct pressure with the complete working period under different working conditions in different working conditions is learned. In order to speed up the convergence in the training network, the original data is normalized using a normalization function. At the same time, in order to obtain better learning results, the learning accuracy is limited, and the best learning results are found in 1000 data learning. The BP neural network learning effect is shown in Figure 4.

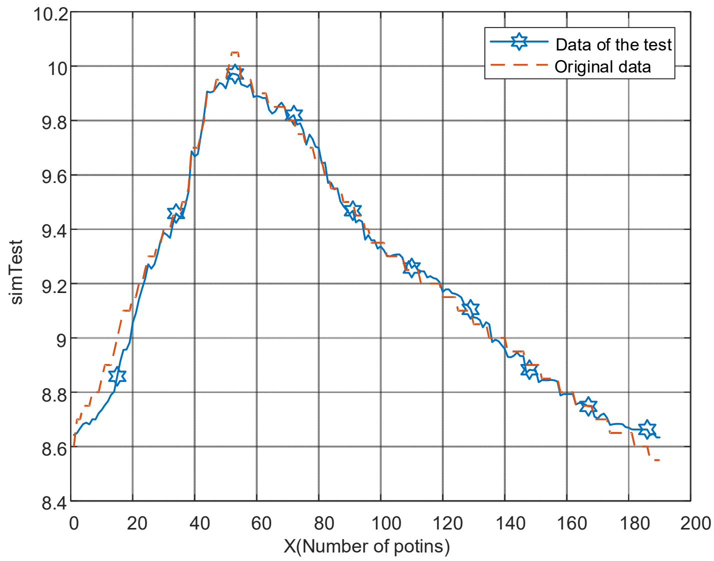

Figure 4. Learning curve of the BP neural network.

Figure 4 shows the effect of the BP neural network on the normal data learning of the duct. The horizontal axis represents the timing-based data collection point, and the vertical axis represents the duct pressure. The green curve in the figure represents the raw data for the duct pressure, and the blue curve represents the neural network learning data. It can be seen from the figure that the BP neural network is more effective for air duct learning, and the learning curve is more suitable for the original data curve. The BP neural network learning error check table for known faults is shown in Table 1.

Table 1. Learning error checks in the BP neural network.

Type	Quantity	Average Relative Error	Maximum Absolute Error
Normal State	0.7820	0.0021	0.0583
Air Brake Pipe Leakage Fault State	0.9605	0.0028	0.0028
Air Brake Pipe Blocking Fault State	0.7868	0.0027	0.0920
Compressor Fault State	1.2889	0.0037	0.0996
Average Error	0.9546	0.0028	0.0919

It can be seen from the error checklist that the BP neural network has a better learning effect on the duct pressure data and a higher degree of fitting. The average root mean square error is 0.9546, the average relative error is 0.28%, and the average maximum absolute error is 9.19%. However, the air duct obstruction fault and the compressor fault are larger than other state errors in the maximum absolute error and the root mean square error test. This is because there is a certain time in the several sets of training data that the duct pressure data points are higher than in the other groups of duct pressure data at that moment.

4.2. Function Approximation

Discrete learning data can be obtained by the BP neural network for duct pressure learning, but such data cannot directly calculate the norm, so it is necessary to select appropriate bases and coefficients to express such data.

According to the characteristics of the pipeline pressure data, Fourier basis, Gaussian basis, polynomial basis, and sine basis are used for the experimental design in the continuous function space. On the basis of each basis, the multi-order function approximation equation is calculated, and the optimality is selected by fitting the K basis coefficient and the equation coefficient, so that the learning result in the BP neural network is expressed as a function expression. The BP neural network is used to learn the pipeline pressure data. Through a set of optimal bases and coefficients, the learning results are transformed into model elements in the model space by function approximation, and the optimal equation is obtained, so as to prepare for the next fault diagnosis.

Taking the normal learning data of the duct as an example, the Fourier base, Gaussian, polynomial basis, and sine base are selected to perform the first to eighth order approximation in the BP neural network learning data, and the fit K is calculated.

The calculation results of the four-based function approximation error test table and the fit degree K are shown in Tables 2–5. The experimental results of the four-base 1 to 8 order function approximation are shown in Figures 5–8.

Table 2. Error checks and fit degree of the Fourier basis function.

Fourier Basis	SSE	R-Square	Adjusted R_Square	RMSE	K
1-order	4.197	0.8765	0.8745	0.1502	4.5962
2-order	0.428	0.9874	0.9871	0.04823	0.50173
3-order	0.198	0.9942	0.9939	0.03299	0.24289
4-order	0.1968	0.9942	0.9939	0.03307	0.24177
5-order	0.1896	0.9944	0.9941	0.03264	0.23374
6-order	0.1479	0.9956	0.9953	0.02898	0.18598
7-order	0.09264	0.9973	0.997	0.02307	0.12141
8-order	0.07201	0.9979	0.9977	0.02046	0.09687

Table 3. Error checks and fit degree of the Gaussian basis function.

Gaussian Basis	SSE	R-Square	Adjusted R_Square	RMSE	K
1-order	9.86	0.7098	0.7067	0.2296	10.6731
2-order	1.304	0.9616	0.9606	0.0842	1.466
3-order	0.8978	0.9736	0.9724	0.07043	1.02223
4-order	0.1958	0.9942	0.9939	0.03316	0.24086
5-order	0.1667	0.9951	0.9947	0.03087	0.20777
6-order	0.1623	0.9952	0.9948	0.03072	0.20302
7-order	0.1612	0.9953	0.9947	0.03088	0.20208
8-order	0.1622	0.9952	0.9946	0.03126	0.20366

Table 4. Error checks and fit degree of the polynomial basis function.

Polynomial Basis	SSE	R-Square	Adjusted R_Square	RMSE	K
1-order	26.13	0.2312	0.2271	0.3728	28.0445
2-order	10.26	0.698	0.6948	0.2343	11.1015
3-order	1.995	0.9413	0.9403	0.1036	2.217
4-order	1.781	0.9476	0.9464	0.09813	1.98513
5-order	1.244	0.9634	0.9624	0.08224	1.40044
6-order	0.4044	0.9881	0.9877	0.04701	0.47561
7-order	0.3481	0.9898	0.9894	0.04373	0.41263
8-order	0.2498	0.9926	0.9923	0.03715	0.30205

Table 5. Error checks and fit degree of the sinusoidal basis function.

Sinusoidal Basis	SSE	R-Square	Adjusted R_Square	RMSE	K
1-order	10.13	0.7018	0.6986	0.2328	10.9624
2-order	1.132	0.9667	0.9658	0.07845	1.27795
3-order	1.298	0.9618	0.9601	0.08467	1.46077
4-order	0.2164	0.9936	0.9932	0.03487	0.26447
5-order	0.3102	0.9909	0.9901	0.0421	0.3713
6-order	0.2541	0.9925	0.9918	0.03844	0.30824
7-order	0.2134	0.9937	0.9930	0.03553	0.26223
8-order	0.2721	0.9920	0.9909	0.04048	0.32968

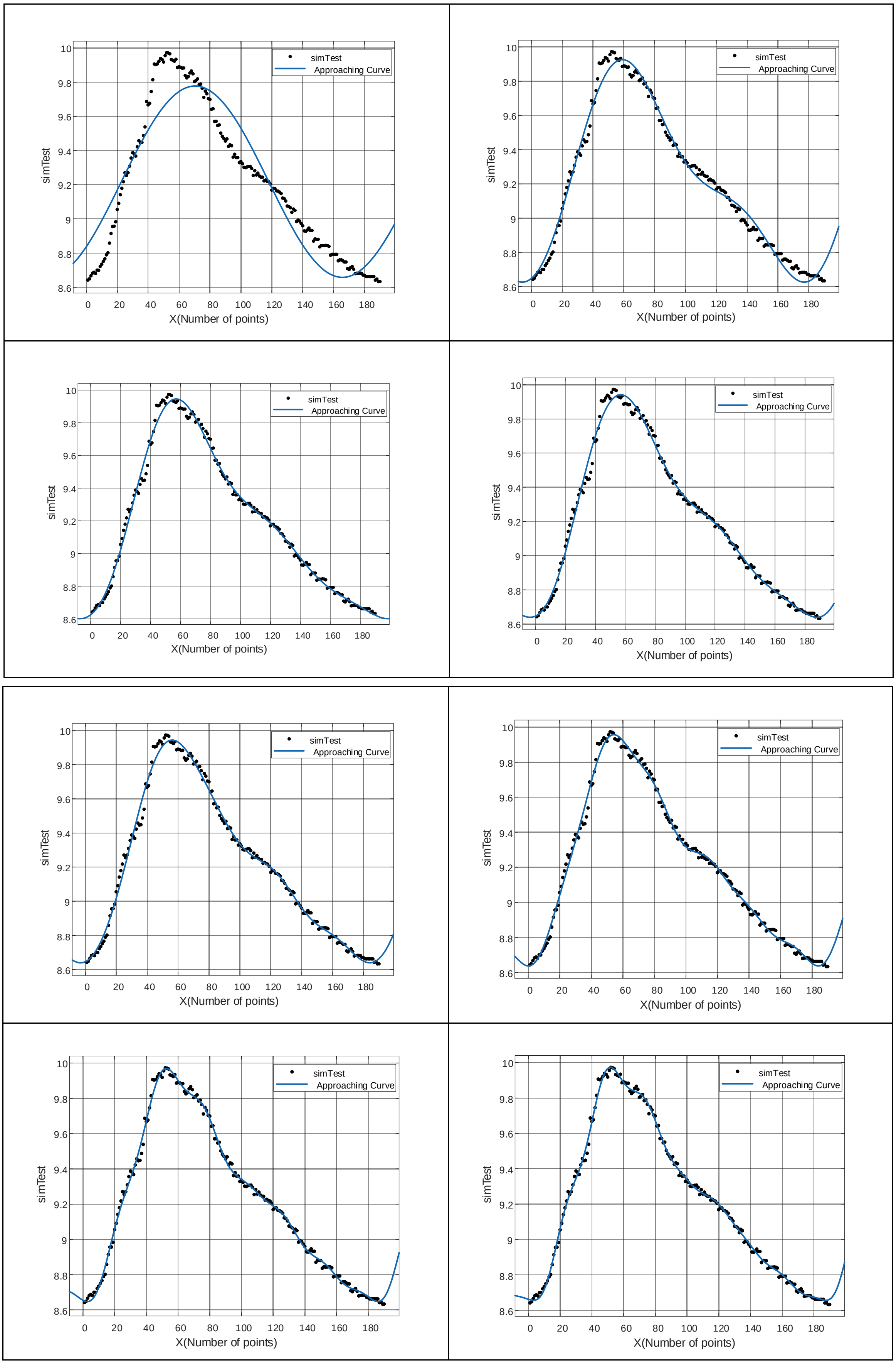

Figure 5. Function approximation curves of the first to eighth order of the Fourier basis function.

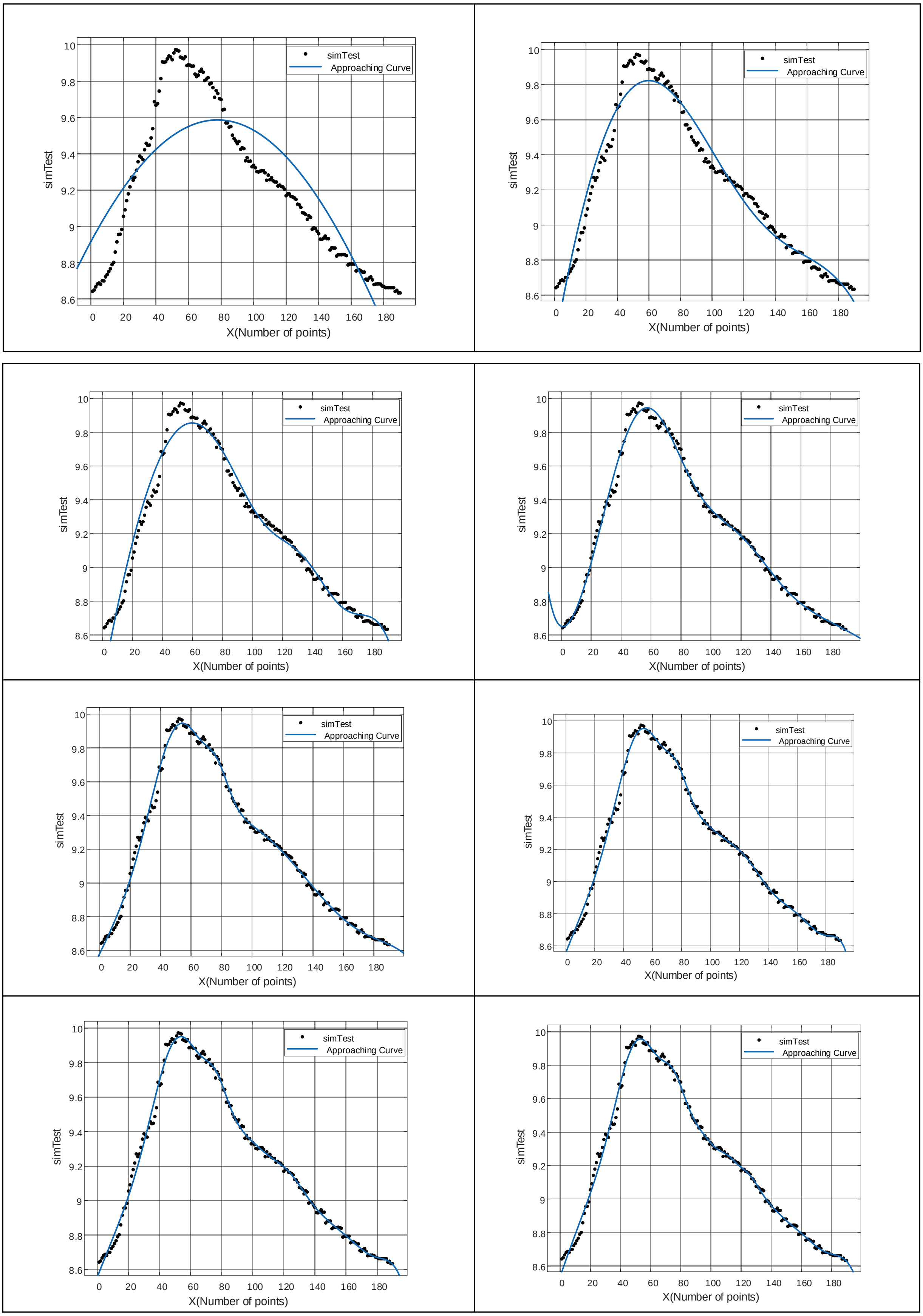

Figure 6. Function approximation curves of the first to eighth order of Gaussian basis function.

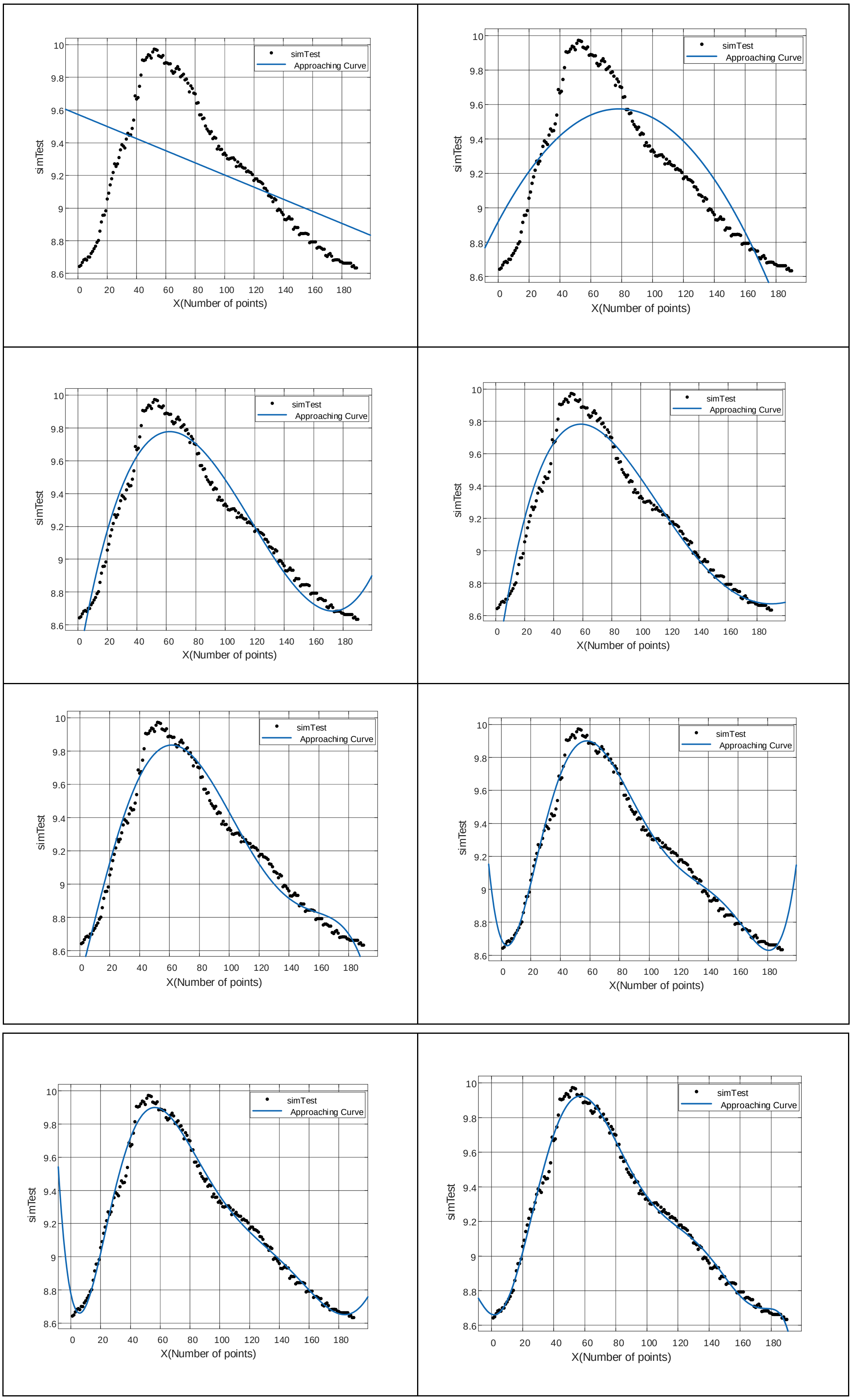

Figure 7. Function approximation curves of the first to eighth order of the polynomial basis function.

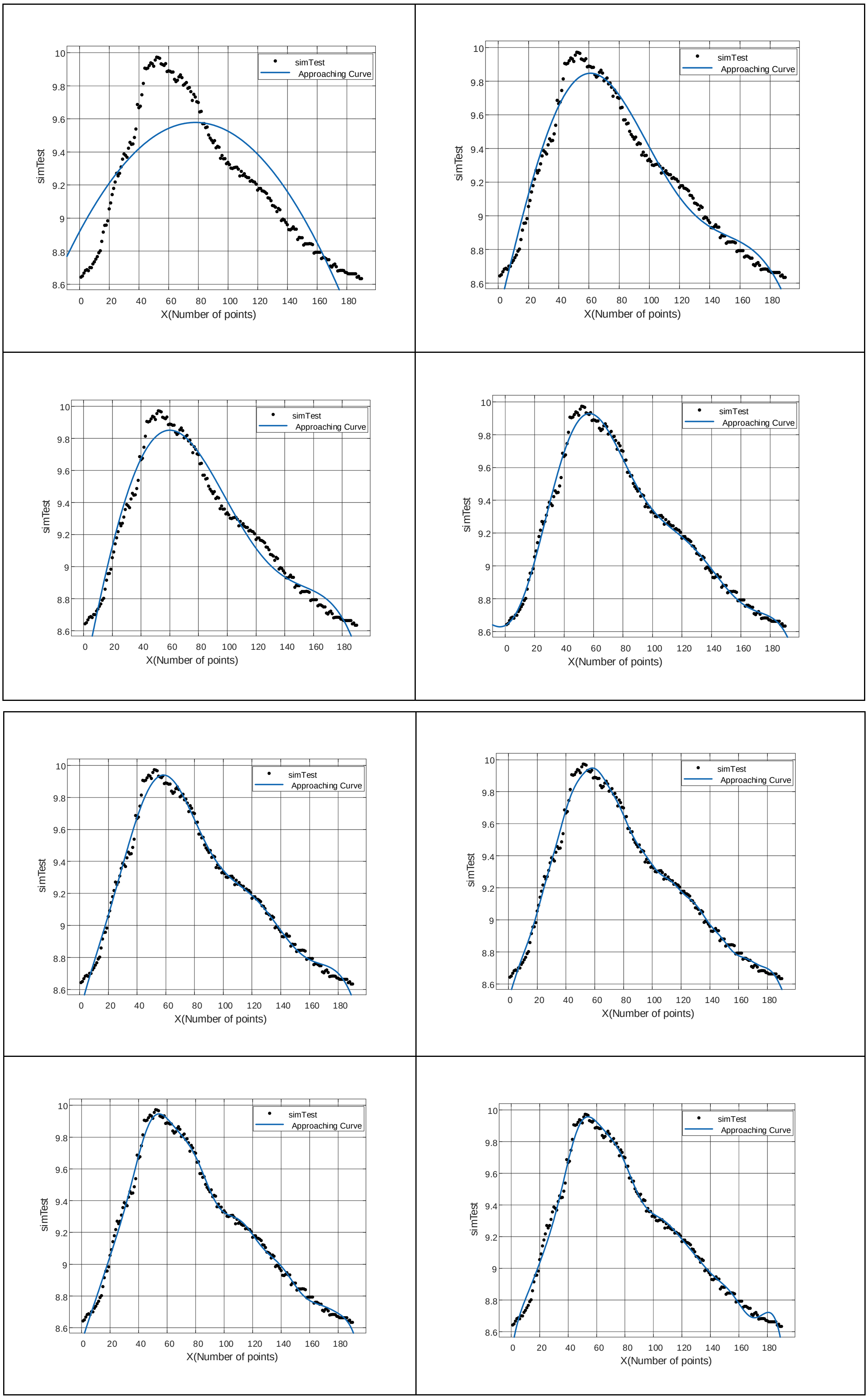

Figure 8. Function approximation curves of the first to eighth order of the sinusoidal basis function.

The solution is the optimal solution only if the fit K satisfies the diagnostic requirements and the approximation equation order is as small as possible. It is verified by experiments where the function approximation effect can meet the fault diagnosis requirements when K is approximately 0.2. Therefore, the optimal function approximation equations for the four duct states are shown in Table 6.

Table 6. Optimal equations in function approximation.

Type	Function Approximation	
	Approximation Method	Approximation Method
Normal State	3rd Order Fourier	$U(x) = 9.172 - 0.3338 \times \cos(x \times 0.0305) + 0.4737 \times \sin(x \times 0.0305) - 0.2006 \times \cos(2 \times x \times 0.0305) - 0.013 \times \sin(2 \times x \times 0.0305) - 0.0122 \times \cos(3 \times x \times 0.0305) - 0.0718 \times \sin(3 \times x \times 0.0305)$
Air Brake Pipe Leakage Fault State	1st Order Fourier	$U(x) = 9.149 - 0.5561 \times \cos(x \times 0.0359) + 0.3462 \times \sin(x \times 0.0359)$
Air Brake Blocking Fault State	3rd Order Fourier	$U(x) = 9.2 - 0.6427 \times \cos(x \times 0.0439) + 0.0169 \times \sin(x \times 0.0439) + 0.0368 \times \cos(2 \times x \times 0.0439) + 0.0275 \times \sin(2 \times x \times 0.0439) - 0.0204 \times \cos(3 \times x \times 0.0439) - 0.0701 \times \sin(3 \times x \times 0.0439)$
Compressor Fault State	2ndOrder Fourier	$U(x) = 9.17 - 0.5635 \times \cos(x \times 0.0318) + 0.18 \times \sin(x \times 0.0318) - 0.0791 \times \cos(2 \times x \times 0.0318) - 0.1252 \times \sin(2 \times x \times 0.0318)$

4.3. Fault Diagnosis

For the high-speed train duct failure, 25 sets of measured duct pressure data are taken as an example for diagnosis. Test1 to Test5 are the normal test data of the air duct, Test6 to Test10 are the air duct blocking fault test data, Test11 to Test15 are the compressor fault test data, Test16 to Test20 are the air duct leakage fault test data, and Test21 to Test25 are the unknown fault test. The data and the diagnosis results are shown in Table 7.

Table 7. Results of the air brake pipe fault diagnosis.

Test Data	Distance between Test Data and Different Fault State				Diagnostic Type	Diagnostic Result
	Normal State	Leakage Fault State	Blocking Fault State	Compressor Fault State		
Test1	37.2358	1919.9133	2177.4830	721.9825	Normal State	True
Test2	42.9296	1947.0545	2207.1937	685.5478	Normal State	True
Test3	56.7879	1891.1919	2152.4350	709.6894	Normal State	True
Test4	32.5843	1919.9875	2181.7044	700.9639	Normal State	True
Test5	46.3860	1906.5605	2171.6021	706.2858	Normal State	True
Test6	1957.1840	42.2385	421.4379	2554.5663	Leakage Fault State	True
Test7	1907.3857	67.8710	401.3378	2505.9051	Leakage Fault State	True
Test8	1940.9620	42.7162	421.0550	2533.5172	Leakage Fault State	True
Test9	1929.0792	66.0089	466.9328	2523.4337	Leakage Fault State	True
Test10	1956.9832	53.3107	381.9810	2558.9373	Leakage Fault State	True
Test11	684.1946	2521.5848	2794.3020	34.7988	Compressor Fault State	True
Test12	711.0597	2491.3055	2770.4532	57.9325	Compressor Fault State	True
Test13	701.0328	2541.9890	2815.0319	16.6492	Compressor Fault State	True
Test14	750.1892	2476.9892	2781.6865	82.7843	Compressor Fault State	True
Test15	743.5008	2515.6166	2798.7663	53.8872	Compressor Fault State	True
Test16	2174.7052	372.6418	44.4142	2785.1763	Blocking Fault State	True
Test17	2187.7097	366.8166	50.4448	2793.8970	Blocking Fault State	True
Test18	2204.5154	412.9392	25.7749	2813.7515	Blocking Fault State	True
Test19	2222.2124	392.3571	39.9392	2837.8202	Blocking Fault State	True
Test20	2213.5172	364.3250	49.3329	2827.1702	Blocking Fault State	True
Test21	2300.7249	640.2812	304.0638	2982.2402	Unknown Fault State	True
Test22	2891.1955	1142.1557	735.9846	3507.3801	Unknown Fault State	True
Test23	2895.6145	1134.3446	728.9325	3511.5424	Unknown Fault State	True
Test24	2913.3218	1110.7721	728.9557	3532.4020	Unknown Fault State	True
Test25	2906.7712	1105.9850	717.0546	3517.6925	Unknown Fault State	True

It can be seen from Table 7 that all test data fault diagnosis results are correct, and the norm of the same fault type is stable and maintained within the same order of magnitude. The difference between different fault type norms is large, and the data difference is as high as one to two orders of magnitude compared with the same fault type norm.

The distance between the unknown fault model element and the known fault model element is far from the same, and the empirical distance difference is also maintained at one to two orders of magnitude. In addition, the norm between the normal model element of the duct and the compressor failure model element is smaller than the norm between the normal model element of the duct and the air duct blocking fault model element and the air duct leakage fault model element. The compressor fault model element also has this feature.

According to the above known fault data (normal air duct, air duct obstruction, air duct leakage, compressor fault) and unknown fault data, the fault diagnosis experiment statistics are performed, and each of the four known faults completes 50 experiments, and the unknown fault is completed in six experiments. The diagnostic results are shown in Table 8.

Table 8. Statistical results in air brake pipe fault diagnosis.

Type	Normal State	Leakage Fault State	Blocking Fault State	Compressor Fault State	Unknown Fault State	Diagnostic Rate
Normal State	49	0	0	1	0	98%
Leakage Fault State	0	44	6	0	0	88%
Compressor Fault State	2	0	0	48	0	96%
Blocking Fault State	0	8	42	0	0	84%
Unknown Fault State	0	0	0	1	5	83%
Average Diagnostic Rate	89.8%					

It can be seen from Table 8 that the diagnosis method based on the model space proposed in this paper has a considerable diagnostic effect. The average diagnostic hit rate is 89.8%. In addition, the statistical results again prove the experimental results in Table 7. The diagnosis of the normal model type of the air duct and the fault type of the compressor are mutually causal, and the diagnosis of the fault type of the air duct blockage and the fault type of the air duct leakage are mutually causal.

The hidden Markov model is a probabilistic model of time series, which is used to describe the process of randomly generating an unobservable state random sequence from a hidden Markov chain, and then generating an observation from each state to generate an observation random sequence. HMM is determined by the initial probability distribution, state transition probability distribution, and observation probability distribution. Since the hidden Markov model is a parametric model, when HMM is applied to the field of fault diagnosis, if the basic parameters of the model are determined, the observed data can be used to describe the transfer process of the internal operating state of the air brake tube or to identify the operating state. Compared with the literature [21], the condition of using the hidden Markov model to diagnose the fault in the air duct is that the model is reliable, but the engineering system is often more complex in the production environment, resulting in the model may be inaccurate, which affects the diagnosis effect.

Therefore, a hit rate comparison experiment of two fault diagnosis methods is designed for the same set of data. The traditional hidden Markov method is compared with the fault diagnosis method of a high-speed train air brake tube based on the model space proposed in this paper. The results are shown in Table 9.

Table 9. Comparative results in fault diagnosis based on the model space and the HMM.

Type		Model Space Method	HMM Method
Existing Fault Type	Normal	98%	98%
	Blocking Fault	88%	84%
	Compressor Fault	96%	96%
	Leakage Fault	84%	80%
	Average Diagnostic Rate	91.5%	89.5%
Unknown Fault		83%	66.67%
Average Diagnostic Rate		89.8%	84.93%

Through the comparison table of diagnostic results, it can be seen that the diagnostic hit rate in the known methods is above 89%, but the hit rate in the new fault in the hidden

Markov model is lower, and the overall fault diagnosis in the model space diagnosis method has a slightly higher hit rate than the hidden Markov model. Compared with the fault diagnosis method based on the model, the model space method can improve the fault diagnosis rate to a certain extent and has the characteristics of identifying new faults.

5. Conclusions

This paper proposes a fault diagnosis method for high-speed train ducts based on the model space, establishes a BP neural network learning model for train ducts, and proposes a method for transforming measured air pressure data in the model space and model elements in the model space. High-speed train air duct fault diagnosis and new fault capture is based on the model element topological relationship. Through the proposed data-driven model space method, accurate fault diagnosis and prediction are further realized, so as to ensure the safe operation of high-speed trains. These studies provide an important theoretical and practical basis for the improvement and application of fault diagnosis methods.

In this study, according to the characteristics of the duct pressure data, the Fourier basis, Gaussian basis, polynomial basis, and sine basis are used to design the experiment in the continuous function space. Based on each basis, the multi-order function approximation equation is calculated, and the optimality is selected by fitting the K basis co-efficient and the equation coefficient, so that the learning result in the BP neural network is expressed as a function expression. The diagnosis process is divided into data learning and function approximation in the time series space. The BP neural network is used to learn the duct pressure data, and function approximation is used to transform the learning results into model elements in the model space through a set of optimal bases and coefficients to obtain the optimal equation. Finally, the topological relationship in the model space is used for fault diagnosis. The fault diagnosis results of the measured duct pressure data in high-speed trains show that the diagnostic accuracy in the model space method is higher than 84% in the diagnosis of normal duct, duct blockage fault, duct leakage fault, and compressor fault. The hit rate in fault diagnosis is 83%, and the average hit rate in overall fault diagnosis is 89.8%. Compared with the hidden fault diagnosis method based on the hidden Markov model, the average diagnostic hit rate in four known faults is increased by 2%, and the overall fault diagnosis rate is increased by 4.87%, which shows the effectiveness and reliability of the diagnosis method.

Aiming at the problem of diagnostic accuracy, generalization, and robustness of high-speed train duct fault diagnosis methods, this paper uses the model space idea based on the model element field to study the fault diagnosis method. Although more effective results have been achieved, there are still some shortcomings to be further improved. However, only the problem of disjoint neighborhoods of the model elements is considered in the fault diagnosis, and the intersection problem is not considered. Therefore, the next step will be to study this aspect. In addition, when establishing the fault model, the method does not take into account the mode conversion between fault types and needs to be further improved.

Author Contributions: Conceptualization, W.M.; Methodology, W.M.; Validation, D.H.; Data curation, X.S.; Writing—original draft, J.W.; Writing—review & editing, Y.Y.; Visualization, J.Q. All authors have read and agreed to the published version of the manuscript.

Funding: National Natural Science Foundation of China: U1734210.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request due to privacy.

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

References

1. Li, R.C. Development and progress of world's railway high speed trains in the past 50 years. *Foreign Roll. Stock.* **2014**, *51*, 1–10.
2. Zhen, R.; Jin, Y.X.; Hu, Q.Y.; Shi, C.J.; Wang, S.Z. Prediction of Ship Navigation Behavior Based on AIS Information and BP Neural Network. *China Navig.* **2017**, *40*, 6–10.

3. Tai, X.; Guo, T.; Chen, M.; Zhang, J.; Zhou, D. Understanding the Fault in EMU Braking System. In Proceedings of the 2019 CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS), Xiamen, China, 5–7 July 2019; pp. 176–179.
4. Zhou, D.; Ji, H.; He, X.; Shang, J. Fault Detection and Isolation of the Brake Cylinder System for Electric Multiple Units. *IEEE Trans. Control. Syst. Technol.* **2018**, *26*, 1744–1757. [CrossRef]
5. Sang, J.; Zhang, J.; Guo, T.; Tai, X.; Chen, M.; Zhou, D. Detection of Incipient Leakage Fault in EMU Braking System. In Proceedings of the 2019 CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS), Xiamen, China, 5–7 July 2019; pp. 1–5.
6. Huang, M.J.; Wang, J.Y.; Yang, R. BP Neural Network Based Fault Diagnosis in Vehicle Braking Control System. In Proceedings of the 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, China, 25–27 May 2018; pp. 794–798.
7. Sun, S.L.; Li, F.; Ding, J.J. The multi-parameter mathematic simplification method for air braking features of heavy haul trains. *Roll. Stock.* **2017**, *55*, 6–9.
8. Du, X.; Mai, L.; Kazemi, H.; Sadjadi, H. Fault Detection and Isolation for Brake Rotor Thickness Variation. In Proceedings of the Annual Conference of the PHM Society, PHM Society, Virtual, 9–13 November 2020.
9. Ji, H.; Zhou, D. Incipient fault detection of the high-speed train air brake system with a combined index. *Control. Eng. Pract.* **2020**, *100*, 104425. [CrossRef]
10. Sang, J.; Guo, T.; Zhang, J.; Zhou, D.; Chen, M.; Tai, X. Incipient Fault Detection for Air Brake System of High-Speed Trains. *IEEE Trans. Control. Syst. Technol.* **2020**, *29*, 2026–2037. [CrossRef]
11. Huang, D.; Li, S.; Qin, N.; Zhang, Y. Fault diagnosis of high-speed train bogie based on the improved-CEEMDAN and 1-D CNN algorithms. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3508811. [CrossRef]
12. Wang, C.; Liu, J.; Zio, E. A modified generative adversarial network for fault diagnosis in high-speed train components with imbalanced and heterogeneous monitoring data. *J. Dyn. Monit. Diagn.* **2022**, *1*, 84–92. [CrossRef]
13. Liu, J.; Hu, Y.; Yang, S. A SVM-based framework for fault detection in high-speed trains. *Measurement* **2021**, *172*, 108779. [CrossRef]
14. Cantone, L.; Ottati, A. A simplified pneumatic model for air brake of passenger trains. *Railw. Eng. Sci.* **2023**, *31*, 145–152. [CrossRef]
15. Chen, Z.; Peng, L.; Fan, J.; Chen, Z.; Peng, T.; Yang, C. Fault Injection Strategies for Air Brake System of High-speed Train with AMESim/Simulink Co-simulation. *IFAC-PapersOnLine* **2022**, *55*, 803–808. [CrossRef]
16. Günay, M.; Korkmaz, M.E.; Özmen, R. An investigation on braking systems used in railway vehicles. *Eng. Sci. Technol. Int. J.* **2020**, *23*, 421–431. [CrossRef]
17. Assadi, K.; Slimane, J.B.; Chalandi, H.; Salhi, S. Shunt faults detection and classification in electrical power transmission line systems based on artificial neural networks. *COMPEL Int. J. Comput. Math. Electr. Electron. Eng.* 2023, *ahead of print*. [CrossRef]
18. Cheng, C.; Chen, H.; Shao, J.; Wang, W.; Zhang, B.; Teng, W. Enhanced Fault Diagnosis Using Broad Learning for Traction Systems in High-Speed Trains. *IEEE Trans. Power Electron.* **2021**, *36*, 7461–7469. [CrossRef]
19. Cheng, C.; Liu, M.; Chen, H.; Xie, P.; Zhou, Y. Slow feature analysis-aided detection and diagnosis of incipient faults for running gear systems of high-speed trains. *ISA Trans.* **2021**, *125*, 415–425. [CrossRef]
20. Liu, J.; Tan, H.; Shi, Y.; Ai, Y.; Chen, S.; Zhang, C. Research on Diagnosis and Prediction Method of Stator Interturn Short-Circuit Fault of Traction Motor. *Energies* **2022**, *15*, 3759. [CrossRef]
21. Li, C.; Wei, F.J.; Wang, C.; Zhou, S.H. Fault diagnosis and prediction of complex system based on Hidden Markov model. *J. Intell. Fuzzy Syst.* **2017**, *33*, 2937–2944. [CrossRef]
22. Tian, R.; Chen, F.; Dong, S. Compound Fault Diagnosis of Stator Interturn Short Circuit and Air Gap Eccentricity Based on Random Forest and XGBoost. *Math. Probl. Eng.* **2021**, *2021*, 2149048. [CrossRef]
23. Zhu, J.; Li, S.; Dong, H. Running Status Diagnosis of Onboard Traction Transformers Based on Kernel Principal Component Analysis and Fuzzy Clustering. *IEEE Access* **2021**, *1*, 121835–121844. [CrossRef]
24. Liu, Z.; Yang, S.; Liu, Y.; Lin, J.; Gu, X. Adaptive correlated Kurtogram and its applications in wheelset-bearing system fault diagnosis. *Mech. Syst. Signal Process.* **2021**, *154*, 107511. [CrossRef]
25. Shi, L.; Zhu, Y.; Zhang, Y.; Su, Z. Fault Diagnosis of Signal Equipment on the Lanzhou-Xinjiang High-Speed Railway Using Machine Learning for Natural Language Processing. *Complexity* **2021**, *2021*, 9126745. [CrossRef]
26. Chen, H.; Jiang, B.; Ding, S.X. A Broad Learning Aided Data-Driven Framework of Fast Fault Diagnosis for High-Speed Trains. *IEEE Intell. Transp. Syst. Mag.* **2021**, *13*, 83–88. [CrossRef]
27. Xu, Y.; Liu, J. High-speed train fault detection with unsupervised causality-based feature extraction methods. *Adv. Eng. Inform.* **2021**, *49*, 101312. [CrossRef]
28. Liu, C.; Zhuo, F.; Wang, F. Fault Diagnosis of Commutation Failure Using Wavelet Transform and Wavelet Neural Network in HVDC Transmission System. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3525408. [CrossRef]