

Research Article

Intelligent Fault Diagnosis of Aeroengine Sensors Using Improved Pattern Gradient Spectrum Entropy

Huihui Li , Linfeng Gou , Hua Zheng, and Huacong Li

School of Power and Energy, Northwestern Polytechnical University, Xi'an 710072, China

Correspondence should be addressed to Linfeng Gou; goulinfengnwpu07@163.com

Received 2 September 2020; Revised 22 October 2020; Accepted 24 November 2020; Published 20 January 2021

Academic Editor: Paolo Castaldi

Copyright © 2021 Huihui Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Timely and effective fault diagnosis of sensors is crucial to enhance the working efficiency and reliability of the aeroengine. A new intelligent fault diagnosis scheme combining improved pattern gradient spectrum entropy (IPGSE) and convolutional neural network (CNN) is proposed in this paper, aiming at the problem of poor fault diagnosis effect and real-time performance when CNN directly processes one-dimensional time series signals of aeroengine. Firstly, raw fault signals are converted into spectral entropy images by introducing pattern gradient spectral entropy (PGSE), which is used as the input of CNN, because of the great advantage of CNN in processing images and the simple and rapid calculation of the modal gradient spectral entropy. The simulation results prove that IPGSE has more stable distinguishing characteristics. Then, we improved PGSE to use particle swarm optimization algorithm to adaptively optimize the influencing parameters (scale factor λ), so that the obtained spectral entropy graph can better match the CNN. Finally, CNN mode is proposed to classify the spectral entropy diagram. The method is validated with datasets containing different fault types. The experimental results show that this method can be easily applied to the online automatic fault diagnosis of aeroengine control system sensors.

1. Introduction

As one of the key components in any aircraft system, a reliable aeroengine is critical for performance and flight safety [1]. Aeroengine control systems have grown increasingly safe and reliable alongside modern technological advancements; the sensors supporting these systems thus require increasingly stringent accuracy and reliability [2]. Because of its wide distribution, special installation position, and extreme working conditions, the sensor is the most vulnerable component of the aeroengine control system [3]. The digital control system works based primarily on sensor signals. Sensor faults are generated by the degradation or failure of engine functions, which represent serious economic losses. It is essential to develop timely and effective sensor diagnosis techniques as improving the safety and reliability of the aeroengine is, arguably, much more important than simply improving its performance.

Fault diagnosis involves first acquiring fault signals, then obtaining fault information from the signals through proper diagnostic techniques, and finally judging and classifying

them before making decisions accordingly [4]. The aeroengine sensor signals collected in an engineering scenario are nonlinear and nonstationary; the complex structure of the aeroengine results in a complex signal transmission path and noise coupling, which makes fault diagnosis extremely challenging [5].

Traditional model-based fault diagnosis schemes have inherent limitations such as large interference, low model accuracy, difficulty in obtaining fault information, and ineffective threshold designs [6]. It is yet necessary to secure newer, effective fault feature extraction and fault diagnosis methods. Since the era of big data, with the mining and application of massive data, data-driven fault diagnosis methods emerge and develop rapidly, becoming a research focus. The data-driven approach uses a large amount of process data, including historical data and online measurement data, for fault diagnosis. Machine learning-based fault diagnosis is one of the typical data-driven methods.

Commonly used intelligent algorithms include the BP neural network [7], extreme learning machine (ELM) [8], support vector machine (SVM) [9], and fuzzy algorithm

[10]. For example, Li et al. [9] proposed a fault diagnosis method based on the combination of SVM and principal component analysis feature information extraction, which speeds up the diagnosis. Kai et al. [7] successfully detected the faults of the aeroengine based on BP neural network. However, because the performance of SVM involves the operation of m -order matrix (M is the number of samples), especially when M is very large, this method is extremely memory consuming. BP network is a shallow model, which usually only learns one or two layers of data representation, and cannot obtain enough fault information, which limits the accuracy of the final diagnosis. In addition, for these shallow networks used in the above methods, the performance of the data-driven fault diagnosis scheme depends largely on the degree of the individual operator's professional experience or prior knowledge. This brings difficulties and errors to feature extraction [11, 12]. The performance of the data-driven fault diagnosis scheme depends largely on the degree of the individual operator's professional experience or prior knowledge [9, 10]. This brings difficulties and errors to feature extraction. Therefore, the performance of the training model is poor, and it can no longer meet the rapid and high-precision requirements of modern fault diagnosis. A faster and more reliable automated diagnostic process is still needed.

The alternative method, convolutional neural network (CNN), is regarded as the most prominent deep learning method. Compared with the shallow network, deep learning can dig deeper information [13]. It acquires feature information through a hierarchical network to resolve the dependence on artificially designed features [14]. In other words, no prior knowledge of signals is required. The CNN can solve gradient decay problems in addition to its stronger feature learning and characterization abilities than the shallow network [15, 16]. The CNN preprocessing procedure is relatively brief, so the learning time is brief—there is less data necessary to learn free parameters, which relieves the memory burden for network operation and allows for the construction of a very powerful neural network [17]. The CNN also has strong antinoise capability and preserves all necessary information during feature selection. These advantages of CNN are beneficial for the field of fault diagnosis. Therefore, it is feasible to select CNN for aeroengine fault diagnosis.

However, directly constructed CNN-based fault diagnosis method exposes the problems of large calculation and long running time because the sensor signal is a long-time series data. In addition, CNN has irreplaceable advantages in processing two-dimensional (2D) images, but the efficiency of direct processing one-dimensional (1D) data is low, especially for complex one-dimensional data such as aircraft signals, which are nonstationary, nonlinear, and low signal-to-noise ratio and have random fluctuations. Some scholars solve this problem by dividing one-dimensional data directly into two-dimensional data. For example, Li and Qu [3] reshape 256×1 data into 16×16 data as the input of the CNN network. But this method cannot solve the problem fundamentally; the effect is not satisfactory. Subsequently, some scholars proposed to convert the data into time-frequency graph first and

then take the time-frequency graph as the input of CNN. For example, Alaskar [14] processed a signal via STFT to obtain scalograms and then diagnosed bearing faults with a CNN. Gao et al. [18] used Morlet CWT to obtain the time-frequency diagram of the signal and combined with CNN for fault diagnosis. However, since STFT is used to process time-varying signals by segmenting the signal interception, STFT is still a stationary signal analysis method in nature. It is difficult to obtain good results for nonlinear and nonstationary aeroengine signals. The wavelet transform is famous for its multiresolution analysis performance [19, 20], but it needs to choose wavelet basis with the help of manual experience and lacks self-adaptability. Moreover, these time-frequency analysis methods are very complex, which is not conducive to the use of CNN for real-time diagnosis of aeroengine control system sensors. Therefore, a more effective method is needed to process aeroengine sensor signals.

Pattern gradient spectral entropy with mathematical morphology (MM) as the core idea is a good solution to the abovementioned problems of CNN for fault diagnosis. It is especially suitable for the analysis and research of all kinds of complex nonlinear signals, and has the advantages of simple calculation, fast parallel, and convenient for hardware implementation [21]. MM extracts fault features by processing the signal from a collective perspective rather than the traditional numerical modeling and analysis [17], which has obtained considerable application in many fields such as signal processing and fault diagnosis.

Pattern spectrum (PS) is a commonly used fault identification method based on MM which uses multiscale morphological analysis to process nonstationary signals to extract their hidden characteristics [22]. Li [23] characterized vibration signals based on an extended morphological pattern spectrum (MPS) calculation with a morphological erosion operator. Wang et al. [24] used the improved pattern spectrum (IPS) to process vibration signals and form effective features. However, PS can only extract unilateral fault features of the signal, which is easy to ignore the feature information hidden on the other side [25]. From this perspective, PS cannot completely and accurately describe the morphological complexity of signals. This is remedied by introducing pattern gradient spectral entropy based on Shannon entropy and gradient operator.

The advantages of this approach include the following: (1) PGSE can fully consider the dynamic characteristics on multiple time scales and the remote correlation information of signals. When different scale structural elements are used to analyze and process the signal, different characteristics of the signal can be retained. This is equivalent to an adaptive perceptual process to match the signal. (2) It is not only simple to calculate, easy to implement on-line by hardware, but also has antinoise characteristics. (3) Most importantly, it converts one-dimensional signals into spectral entropy graphs as input to CNN to refine and summarize the morphological features of faults at multiple levels while retaining details. This greatly reduces the calculation and training time of CNN, and makes accurate CNN online diagnosis possible.

PGSE still has several limitations for fault diagnosis. Some limitations are in fault diagnosis based on PGSE. The

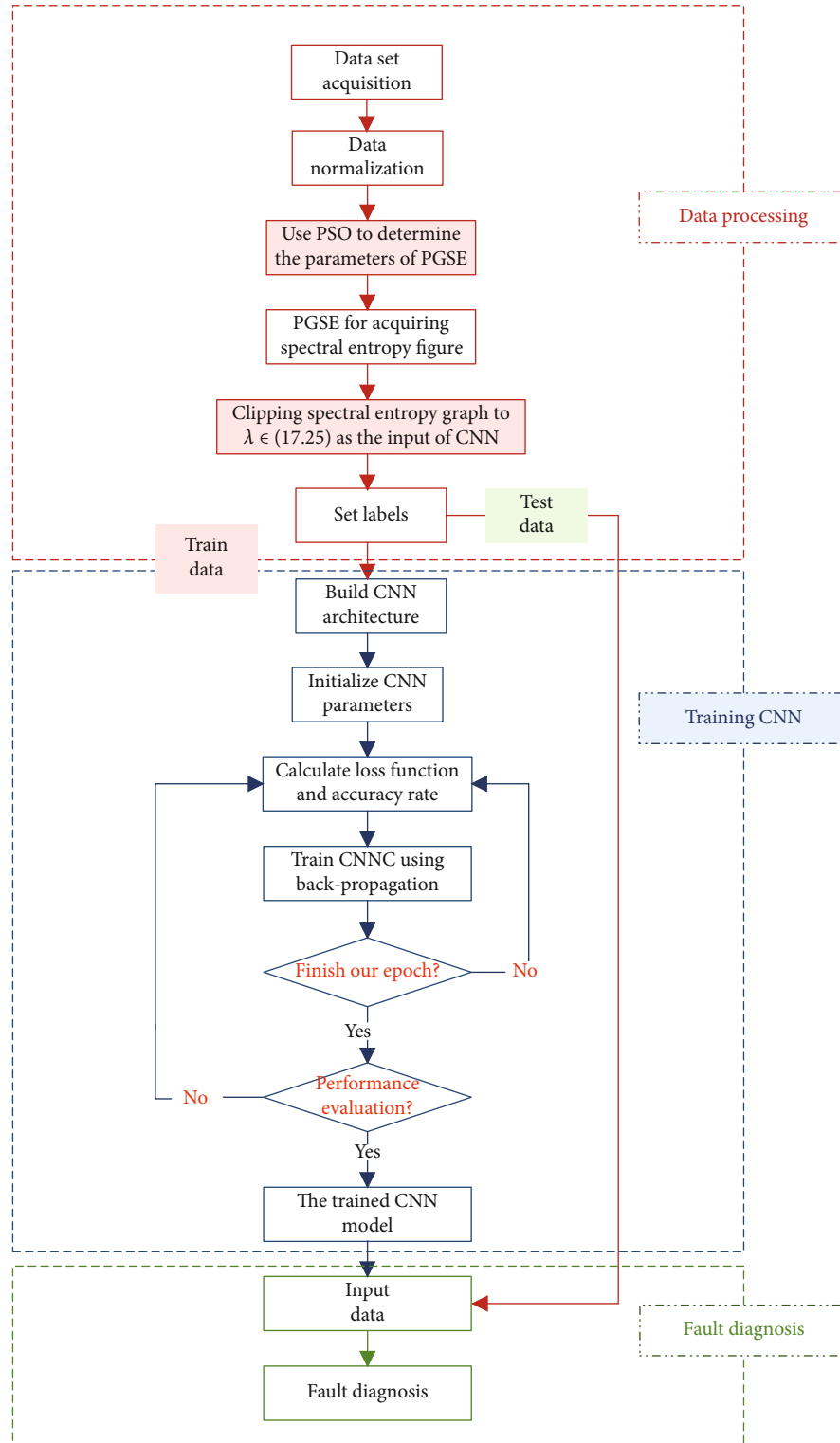


FIGURE 1: Flowchart of proposed method.

key parameter (scale factor λ) value is usually determined in PGSE according to the empirical parameters and by manual intervention, which does not typically provide effective treatment. In view of this, particle swarm optimization (PSO) algorithm is proposed in this paper to improve the pattern

gradient spectral entropy (IPGSE), which can be used to determine the scale factor λ adaptively to better match CNN's pattern classification. Compared with the random or empirical selection of the scale factor, the optimal parameter λ_{op} selected via PSO increases the differentiation degree

between different faults. This facilitates the accurate identification of subsequent classifiers and improves the fault diagnosis accuracy. In a word, IPGSE can extract fault features more effectively and quickly and is conducive to accurate classification.

This paper proposes a hybrid intelligent fault diagnosis scheme which combines IPGSE and CNN for aeroengine control system sensors and introduces PSO algorithm to determine the scaling factor adaptively. The IPGSE first obtains the multiscale fault information of the preprocessed sensor signal and generates a spectral entropy graph. The CNN then classifies the labeled spectral entropy graph. Experimental results, as discussed in detail below, show that the proposed method is applicable for the online automatic fault diagnosis of aeroengine control system sensors.

The major contributions and innovations of this paper are summarized as follows.

- (i) A new data-driven diagnosis method is developed which combines the IPGSE and CNN to extract fault features effectively and quickly to accurately classify faults
- (ii) In order to solve the shortage and real-time problem of CNN in processing one-dimensional aeroengine signals, PGSE was proposed as a good supplement to CNN. In this method, one-dimensional signals are converted into two-dimensional spectral entropy diagrams as the input of CNN, which maximizes the advantages of CNN and makes accurate CNN online diagnosis possible
- (iii) To solve the problem of insufficient fitness of the algorithm, PSO algorithm is proposed to determine scale factor of structural element adaptivity, and PGSE is improved to better match the CNN pattern classification

The structure of this paper is as follows. Section 2 presents the framework of the intelligent fault diagnosis method based on IPGSE and CNN for aeroengine control system sensors. Section 3 discusses the experiments conducted to evaluate the effectiveness of the proposed method. Section 4 gives concluding remarks.

2. Fault Diagnosis Method Framework

For the signals collected on the sensors of the aeroengine control system, the fault characteristics are weak owing to noise coupling and the complex transmission channel. The mixed waveforms of different fault signals are difficult to distinguish, so it is difficult to obtain ideal fault diagnosis results. Pattern gradient spectrum entropy analysis is a workable supplement to traditional fault diagnosis techniques. The IPGSE based on MM processes nonstationary signals, extracting the characteristic information hidden in the signal efficiently. The CNN was selected in this study for its strong feature classification effects. Joining IPGSE and CNN to establish an intelligent fault diagnosis scheme was the focus of this work.

The framework of the proposed fault diagnosis algorithm is shown in Figure 1. It is mainly a four-step process.

- (1) Fault signal marking rules and preprocessing: fault marking rule is defined to highlight the differences between different fault states. The collected sensor signal is preprocessed to minimize the error of the measurement signal generated in the signal collection process and reduce the computational complexity
- (2) Fault feature extraction: IPGSE processes the preprocessed sensor signal to relieve the limitation of parameter selection based on experience. The PSO calculates the PGSE after obtaining the optimal SE scale and then outputs a spectral entropy diagram
- (3) Fault diagnosis model training: a CNN model is constructed with the spectral entropy map within the interception range $\lambda \in (17, 25)$ as the input to train the model. CNN depends on supervised learning, so the spectral entropy map generated in the previous step is labeled according to the fault marking rule before it is input to the network
- (4) Performance evaluation: after the CNN model is trained, a performance evaluation function is established to judge its quality. If the performance meets the requirements, the trained model is used to identify and isolate the faults

2.1. Fault Signal Marking Rules and Preprocessing

2.1.1. Fault Signal Marking Rules. Sensor faults are generally divided into two categories according to their performance: hard faults and soft faults. A hard fault is manifested as a sudden change in the output value of the sensor due to structural damage or other reasons (e.g., open-circuit faults, short-circuit faults, and bias faults). Soft faults are manifested as a slow change in the output value of the sensor due to device aging or other reasons (e.g., spike faults, drift faults, and periodic disturbances). Seven health conditions of aeroengine control system sensors were investigated in this study: the abovementioned six fault conditions and the normal working condition of the sensor.

In the supervised machine learning process, the expected output is already known and serves as the category of sample data before the training process begins. Fault marking rules were stipulated in this study accordingly as shown in Table 1.

2.1.2. Fault Signal Marking Rules and Preprocessing. In practice, aeroengine variables usually use different measurement units. Taking into account the importance of eliminating errors in signal acquisition, the measurement data should be standardized. Normalizing a signal generally allows the signals to be processed at the mean level (Equation (1)). This can reduce computation complexity and processing time in subsequent steps.

$$\begin{aligned}
y(k) &= \frac{x(k) - \bar{x}(k)}{\sigma(x(k))}, \\
\bar{x}(k) &= \frac{1}{n} \sum_{i=1}^n xi(k), \\
\sigma(x(k)) &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (xi(k) - \bar{x}(k))^2},
\end{aligned} \tag{1}$$

where $y(k)$ is the processed signal and $x(k)$ is the sensor signal to be processed.

2.2. Fault Feature Extraction Based on IPGSE

2.2.1. Introduction to Pattern Gradient Spectral Entropy. The core idea of pattern gradient spectral entropy is MM, which is a nonlinear filtering method that creatively utilizes SEs with certain morphologies to interact with signals. Its main objective is to extract the shape information of signals to grasp the signal characteristics. Researchers have introduced multiscale SEs into MM-based analyses to extract fault feature information, referring to the entropy and gradient operators, and ultimately generate pattern gradient spectral entropy.

Take the expansion operation as an example. For a given signal $f(n)$, suppose g is a unit structural element (USE) and $g(m)$ is an SE, λ is the scale factor, and λg is the SE under scale factor λ . The domain is $G = (0, 1, 2, \dots)$.

$$\begin{aligned}
\lambda g &= \frac{\lambda-1}{g \oplus g \oplus \dots \oplus g} = \frac{\lambda-1}{((g \oplus \dots \oplus g) \oplus g) \oplus g} \\
(f \oplus \lambda g)(n) &= f \oplus \frac{\lambda-1}{(g \oplus g \oplus \dots \oplus g)}
\end{aligned} \tag{2}$$

where \oplus denotes the dilation operator and \ominus is the erosion operator.

The pattern spectrum (PS) can be defined as follows:

$$\text{PS}(f, \lambda, g) = \begin{cases} -\frac{dA(f \circ \lambda g)}{d\lambda}, & \lambda \geq 0, \\ -\frac{dA(f \cdot (-\lambda)g)}{d\lambda}, & \lambda < 0, \end{cases} \tag{3}$$

where \circ is the open operator and \bullet is the close operator.

Entropy maps signal complexity and uncertainty. According to Shannon entropy, the pattern spectrum entropy (PSE) is defined as follows:

$$\text{PSE}(f, \lambda) = - \sum_{\lambda=\min}^{\lambda_{\max}} q(\lambda) \ln q(\lambda), \tag{4}$$

where $q(\lambda) = \text{PS}(\lambda, g) / \sum \text{PS}(\lambda, g)$.

After introducing the gradient operator, the pattern gradient spectrum (PGS) and the pattern gradient spectral

TABLE 1: Fault labels.

Fault	Short-circuit	Open-circuit	Spike	Bias
Label	0	1	2	3
Fault	Drift	Normal	Periodic disturbance	—
Label	4	5	6	—

entropy (PGSE) can be defined as follows:

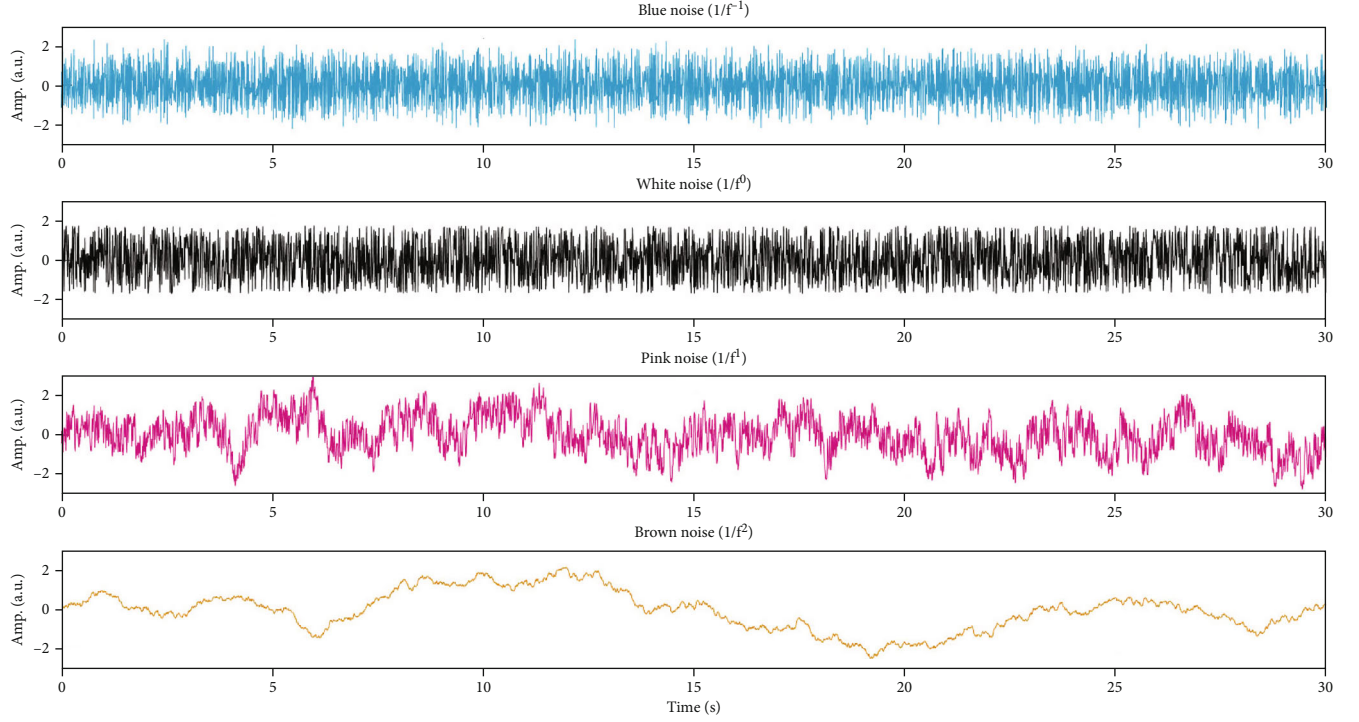
$$\begin{aligned}
\text{PGS}(f, \lambda, g) &= \begin{cases} -\frac{dA(f \oplus \lambda g - f \ominus \lambda g)}{d\lambda}, & \lambda \geq 0, \\ -\frac{dA(f \oplus (-\lambda)g - f \ominus (-\lambda)g)}{d\lambda}, & \lambda < 0, \end{cases} \\
\text{PGSE}(f, \lambda) &= - \sum_{\lambda=1}^{\lambda_{\max}} q_1(\lambda) \ln q_1(\lambda),
\end{aligned} \tag{5}$$

where $q_1(\lambda) = \text{PGS}(\lambda, g) / \sum \text{PGS}(\lambda, g)$.

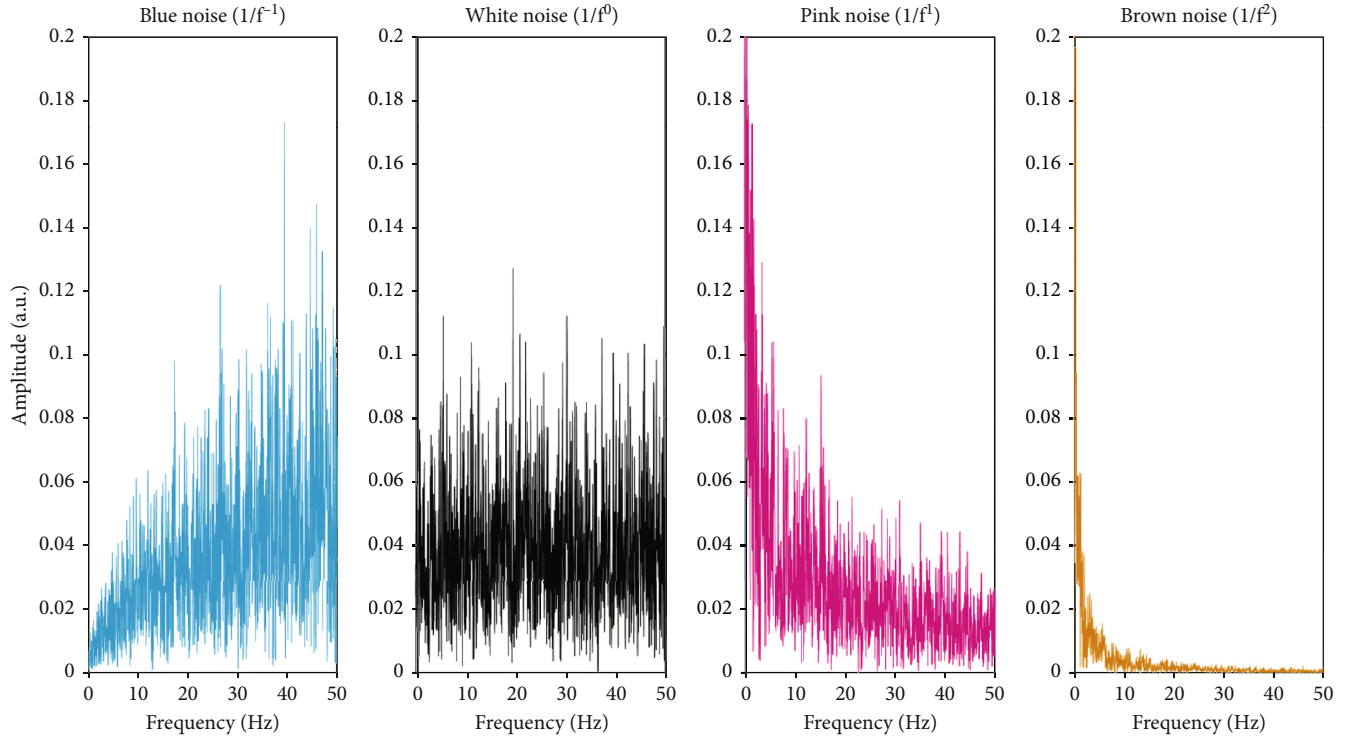
Four types of random signal (blue, white, pink, and brown noise) were taken as the research objects without loss of generality to test the PGSE in this study. The temporal waveforms and frequency-domain waveforms of these random signals are shown in Figure 2. Figure 3 shows the PS, PSE, PGS, and PGSE waveforms of the processed random signals. The abscissa in Figure 3 represents SE scale, and the ordinate represents the calculation result. It is clear from Figure 3 that PGSE is more separable and stable than the other three methods.

2.3. Improved Pattern Gradient Spectral Entropy. Morphological operations use SEs at different scales to flatten or strengthen the required parts of the signal. Figure 4 shows some common SE shapes at different scales. The flat SE is the focus of this paper because of its simplicity. The height H of the flat SE is zero, so it is necessary to determine only its length L in this case. This is equivalent to using a set to operate on the signal, which is widely applicable.

The geometric meaning of the PGSE indicates that the scale of SEs affects the calculation results. The size selection of SEs is important in this regard, but there is no standard selection principle for SE parameters. Parameters are typically selected according to experience and by manual intervention, which makes it difficult to achieve good treatment effects. Nikolaou and Antoniadis [26] suggested that the length of the flat SE be between $0.6T$ and $0.7T$ to obtain acceptable filtering results. Dong et al. [27] proposed the maximum kurtosis criterion and selected the optimal SE length from $0.1T, 0.2t, \dots, T$. Others have suggested that the maximum analysis scale is the ratio between the sampling frequency and fault characteristic frequency, where the scale increment is 1 [28]. The maximum length of SE L_{\max} is usually chosen as $\lfloor f_s / f_g \rfloor$ [29], where f_s represents the sampling frequency of the signal, f_g is the fault characteristic frequency, and $\lfloor \cdot \rfloor$ is the downward integration operation. However, these approaches increase the interference over a series of



(a) Temporal waveform of random signals



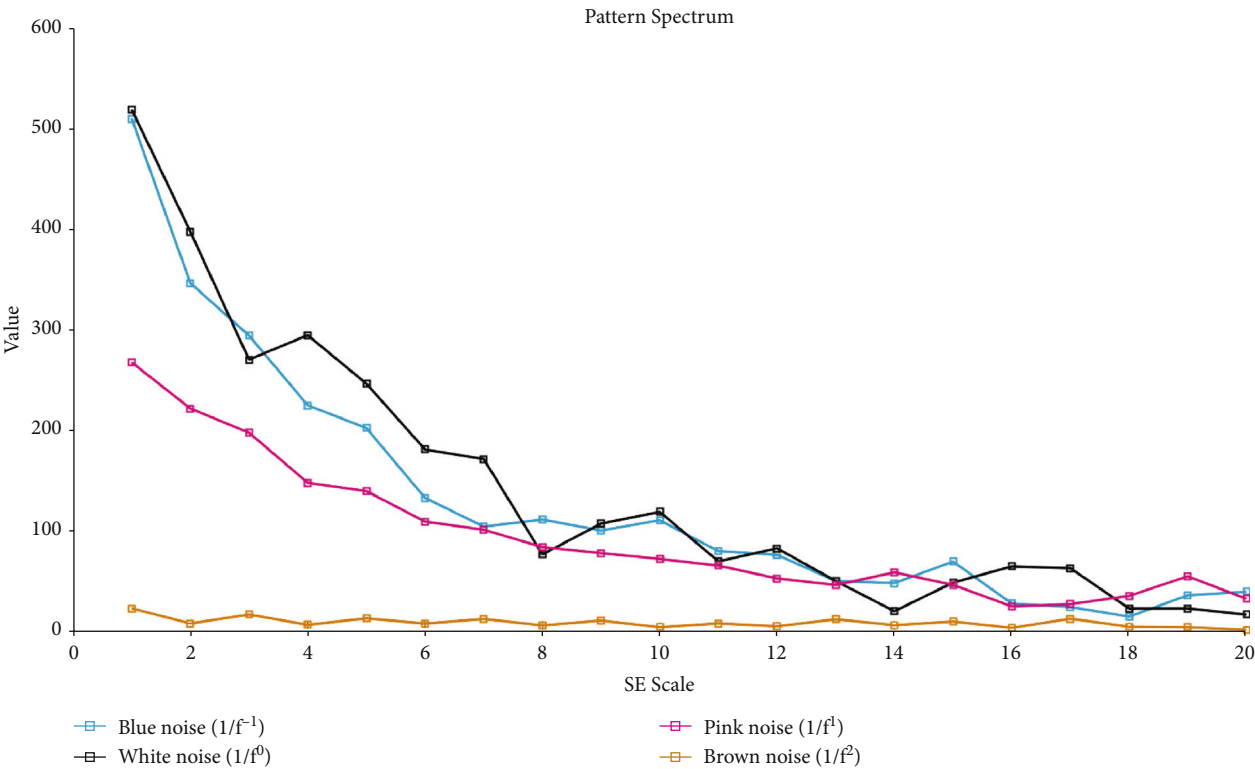
(b) Frequency-domain waveform of random signals

FIGURE 2: Time-domain waveform and frequency-domain waveform of random signals.

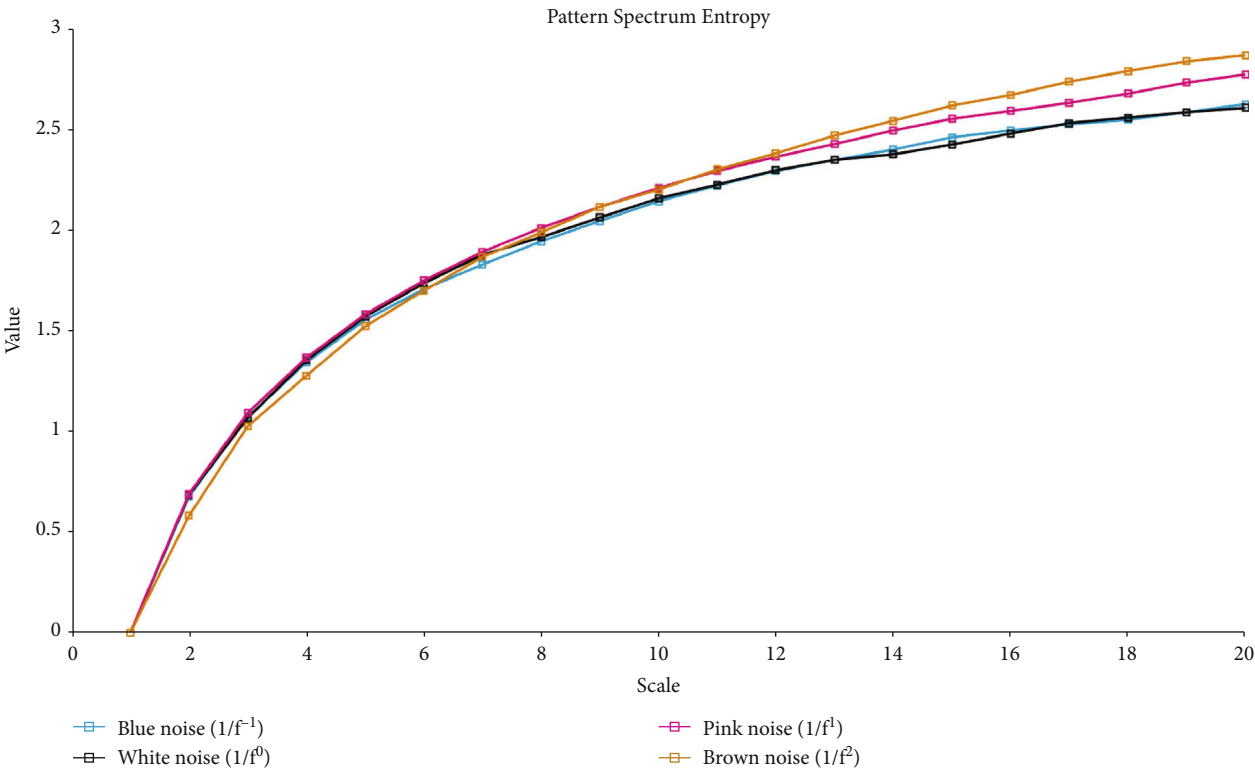
human factors and obscure the optimal SE scale parameters, which can degrade the final diagnosis results considerably.

We drew spectral entropy diagrams at different scales for sensor signals under seven different health conditions

to illustrate the effects of the SE scale factor λ on the fault feature extraction effects of aeroengine control system sensors. The minimum scale we set here is 1 and the increment is 1, so the parameter to be determined is actually the SE maximum scale λ_{\max} . We explored the results

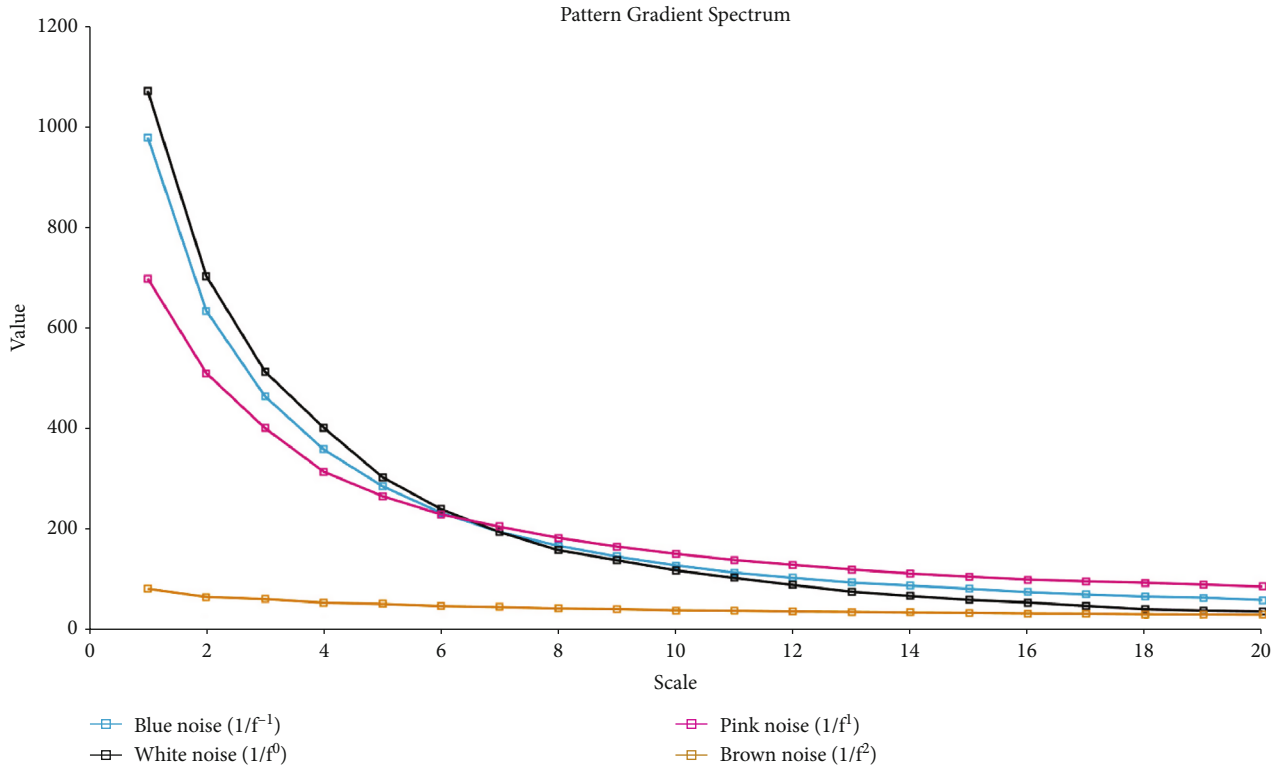


(a) PS graph of random signals

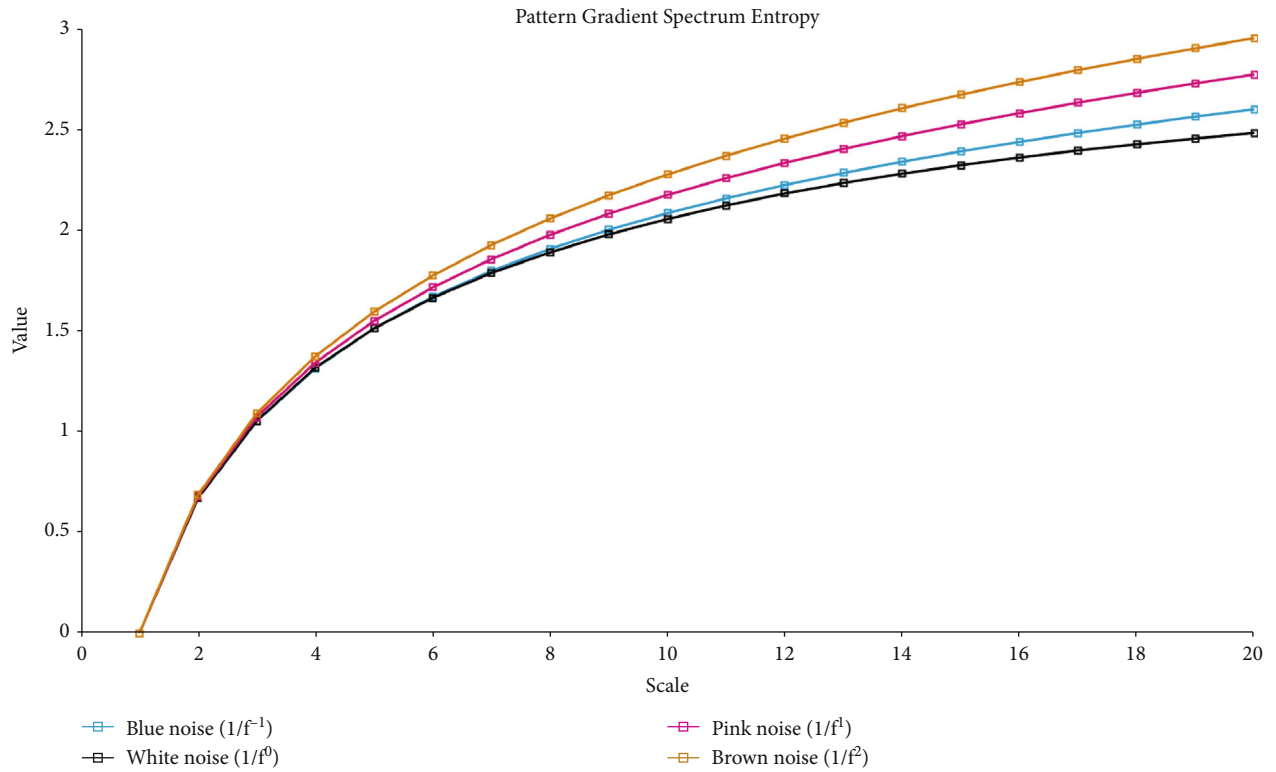


(b) PSE graph of random signals

FIGURE 3: Continued.



(c) PGS graph of random signals



(d) PGSE curve of random signals

FIGURE 3: Curves of simulation signals processed by different methods.

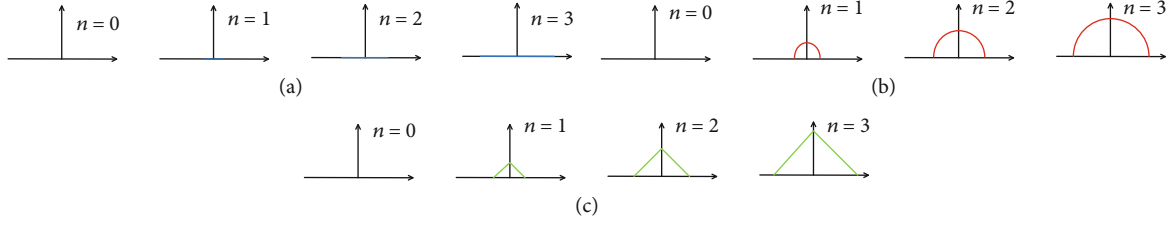


FIGURE 4: SE with varied scale: (a) flat SE, (b) semicircular SE, and (c) triangular SE.

with λ_{\max} of 10, 25, and 30, respectively, as shown in Figure 5. The actual processing of a sensor signal under certain health conditions only produces a curve. The feature extraction effects differed in this case over different SE scale ranges and different SE maximum scale λ_{\max} values. When λ_{\max} was too small, the noise suppression ability was poor and the characteristics of each fault signal could not be extracted effectively. When λ_{\max} was too large, the signal detail retention was poor.

The black dotted line box in Figure 5(c) is a local enlarged image. The solid red line box in Figure 5(c) shows where the curve of the periodic interference fault was mixed with other curves, making it difficult for subsequent classifiers to distinguish various fault signals. We also carefully considered the increase in computational burden introduced by adding scales, which in practice would reduce the real-time fault diagnosis performance. Our goal in conducting this study is to establish a method for effectively optimizing SE parameters while accurately calculating the size of SEs.

The PSO algorithm has strong parameter global optimization capability. We used it here to adaptively optimize the SE maximum scale λ_{\max} of the PGSE to improve the feature extraction effects of our proposed method. Compared with key parameters selected randomly or empirically, the optimal parameters selected via PSO according to the error classification rate of the training samples increase the differentiation degree between different faults, which facilitates effective pattern recognition and improves the fault diagnosis accuracy. Derived from the concepts of “population” and “evolution”, the PSO searches an optimal solution in complex space through the cooperation and competition among individuals. The stepwise PSO process is as follows (Figure 6).

- (1) Input dataset and initialize particle swarm $X_{\lambda i}$, $i = 1, 2, \dots, m$, where m is the dimension of the parameter to be optimized. The parameters to be set include population size $N = 100$, maximum iteration number $T = 200$, learning factor $c_1 = 1.5$, $c_2 = 1.5$, maximum inertia weight $w_{\max} = 0.9$, and minimum inertia weight $w_{\min} = 0.8$. The random factor θ is a random number between $[0, 1]$. Initialize the position x_i and velocity v_i of each particle
- (2) According to the error classification rate of the training samples, the fitness value is calculated and compared to determine the $\text{fit}[i]$ of each particle

- (3) Update the optimal particle. A larger fitness function value means the corresponding particle position is closer to the global optimal position. For each particle, the fitness value $\text{fit}[i]$ is compared with its individual extremum P_{id} . If $\text{fit}[i] < P_{id}$, then $\text{fit}[i]$ is replaced with P_{id} . For each particle, the $\text{fit}[i]$ is compared with the global extremum P_{gd} . If $\text{fit}[i] < P_{gd}$, then $\text{fit}[i]$ is replaced with P_{gd}
- (4) Iterate to update the velocity x_i and velocity v_i of the particle

$$v_{id}(t+1) = v_{id}(t) + c_1 r_1 (P_{id} - x_{id}) + c_2 r_2 (P_{gd} - x_{id}),$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1),$$
(6)

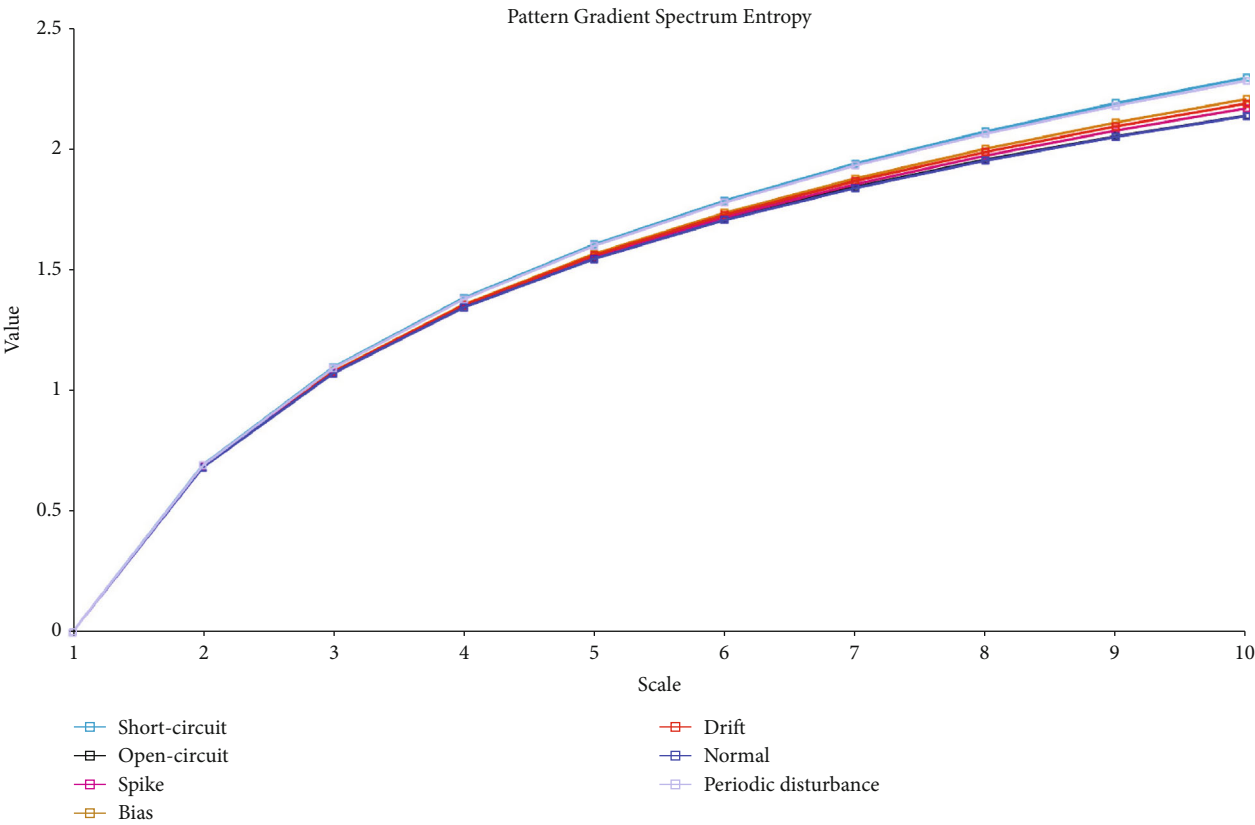
where d represents the dimension of the system.

- (5) Conduct boundary condition processing
- (6) Judge whether the algorithm termination condition is satisfied. If the preset error or the number of iterations is satisfied, the optimization operation ends and the optimization result is output. Otherwise, return to Step 2

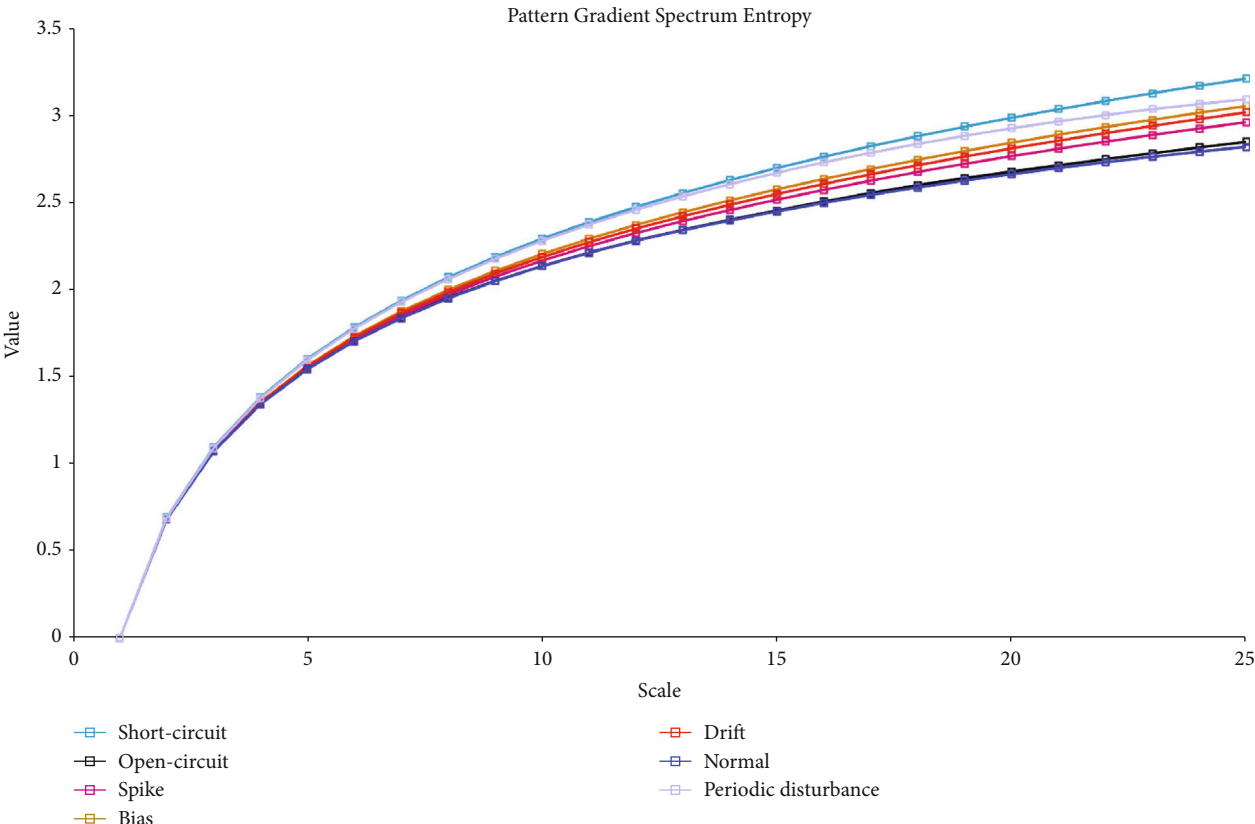
The optimal parameter $\lambda_{\max, \text{op}} = 25$ was obtained here after PSO intelligent optimization. We randomly selected 350 sets of sensor data including 50 sets for each health condition. The method proposed above was used to extract features of fault signal clusters. The results are shown in Figure 7. Figure 7 demonstrates that each health condition was roughly separable, but unrecognizable to the naked eye—a powerful intelligent algorithm is still needed to classify the fault type. We accomplished this with the CNN, which is adept at learning features and representations.

2.4. Convolutional Neural Network Theory. The CNN is an excellent deep learning algorithm [30] that can be used to obtain hierarchical feature information rather than relying on the manual design of features. As a feedforward neural network, the CNN is primarily made of a convolutional layer, pooling layer, and full connection layer [31]. The architecture of the CNN used in the paper is shown in Figure 8.

On the convolutional layer, after the convolution of the input image is carried out through multiple convolution

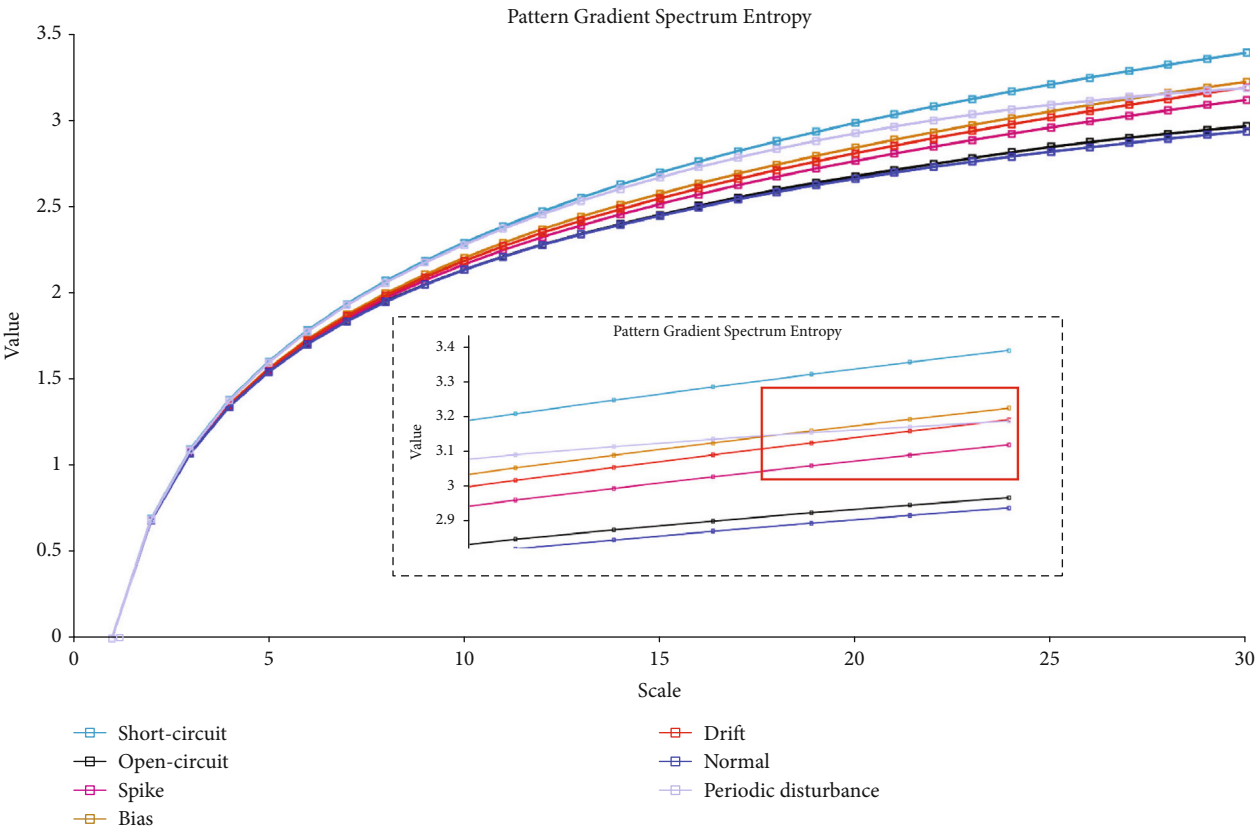


(a) $\lambda_{\max} = 10$



(b) $\lambda_{\max} = 25$

FIGURE 5: Continued.



(c) $\lambda_{\max} = 30$

FIGURE 5: Maximum SE scales with different values.

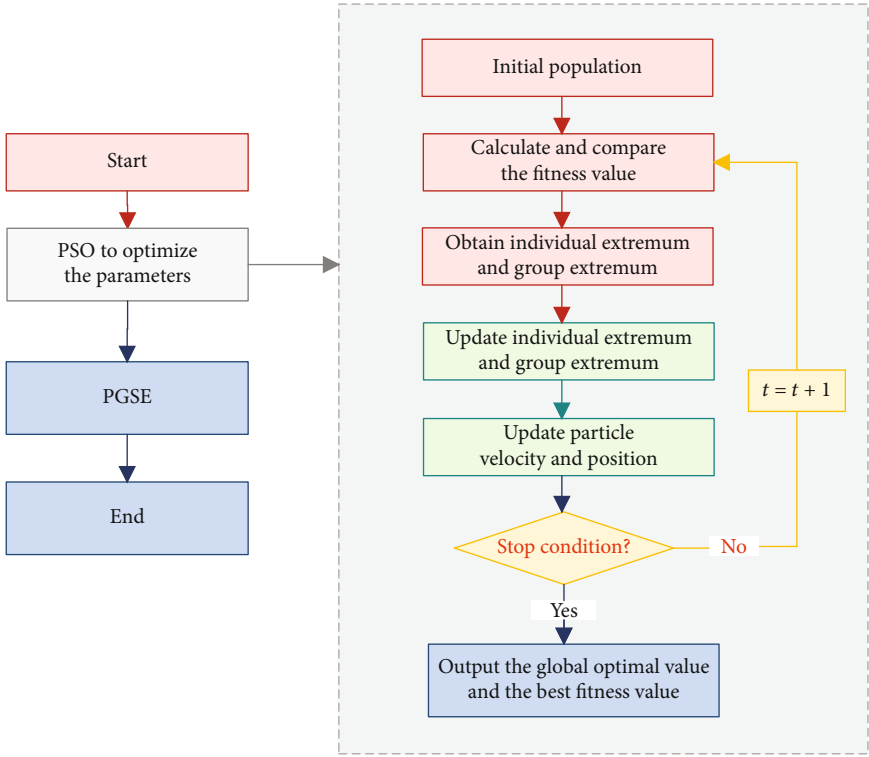


FIGURE 6: Flowchart of IPGSE.

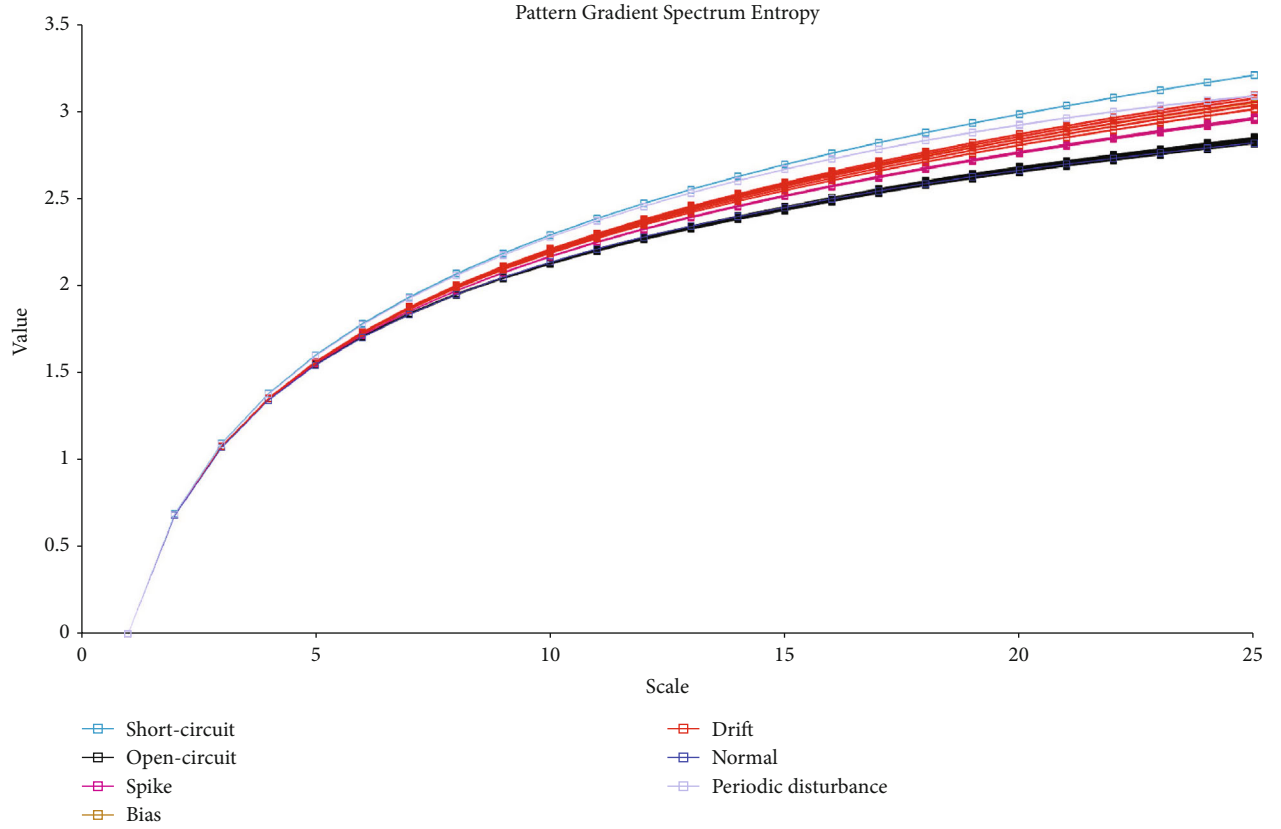


FIGURE 7: Feature extraction results of fault signal clusters.

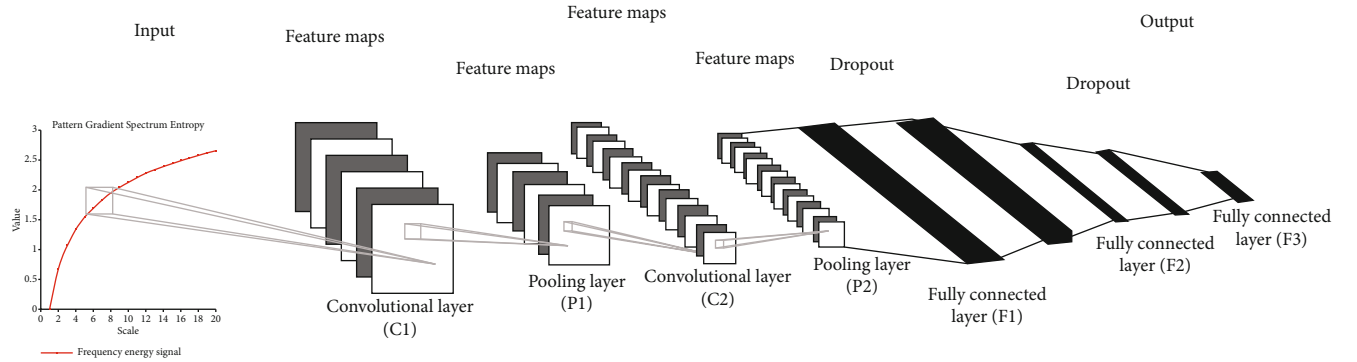


FIGURE 8: CNN architecture.

kernels and the bias term is added, a series of characteristic graphs can be obtained through the nonlinear transformation of the activation function. The convolution is denoted as follows:

$$x_k^n = \phi \left(\sum_{i \in M_k} x_i^{n-1} * \beta_{ik}^n + b_k^n \right), \quad (7)$$

where x_k^n is the j^{th} feature graph of the i^{th} convolution layer, $\phi(\cdot)$ is the activation function, M_k is the input graph set, β is the convolution kernel, and b_k^n is the corresponding bias.

After the convolutional layer comes a pooling layer. The dimensionality reduction of feature map is the major purpose

of the pooling layer. This operation extracts essential features, reduces the data complexity, and improves the network's tolerance to environmental changes. The pooling layer can be expressed as follows:

$$x_k^n = \phi(w_k^n \text{down}(x_k^{n-1}) + b_k^n), \quad (8)$$

where $\text{down}(\cdot)$ is the subsampling function, w_k^n denotes the weight matrix, and the meaning of other parameters is the same as in the above equation.

After multiple alternate propagations between the convolutional layer and the pooled layer, the input image is

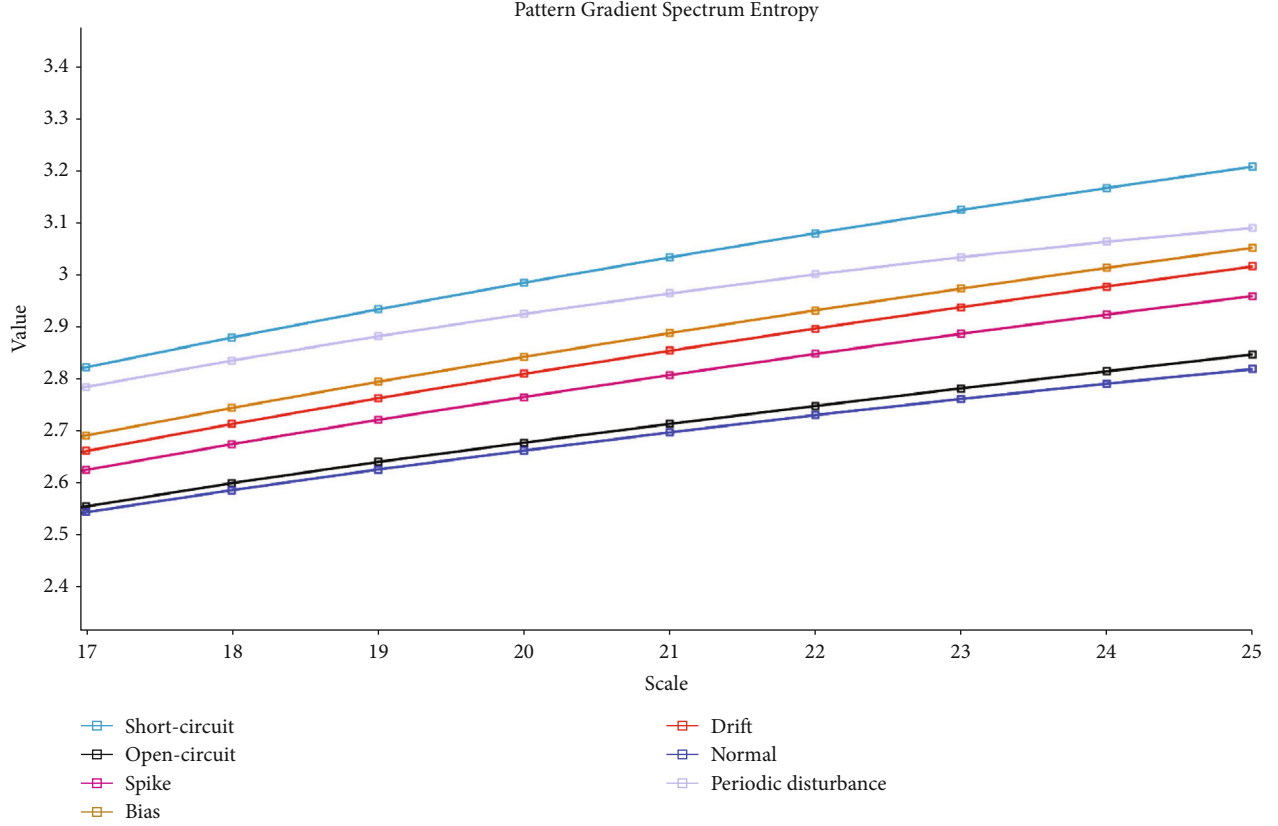


FIGURE 9: Locally enlarged spectrum as CNN input.

classified using the full connection layer:

$$y_k^n = \phi(w_k^n x_k^{n-1} + b_k^n), \quad (9)$$

where w_k^n denotes the weight matrix, y_k^n is the output of the full connection layer, and the meaning of other parameters is the same as in the above equations.

Figure 6 shows the architecture of the CNN implemented in this study. To prevent overfitting and improve the generalization ability and robustness of the network, a portion of neurons and their connections were randomly discarded during training by dropout technology [32]. We used the PyTorch deep learning framework, BP algorithm, and adaptive moment estimation (Adam) algorithm to optimize the model. During the optimization process, the deviation degree between the actual output value and target output prediction was calculated, and then the internal parameters of the model (e.g., weights and bias) were rapidly updated and fine-tuned until the training error was minimal.

The cross-entropy loss function is often used in classification problems, especially in neural networks. The sigmoid function (Softmax) combined with cross-entropy loss depends on an interclass competition mechanism and effectively learns interclass information. The purpose of the cross-loss function is to measure training error. The cross-loss function $F(\theta)$ is calculated as follows:

$$F(y_c, y_l) = -\frac{1}{n} \sum_{\text{class}} [y_c \ln y_l + (1 - y_c) \ln (1 - y_l)], \quad (10)$$

where n indicates the dimension of training data. Class is the tag category into which signals need to be classified, y_c is the output of the sensor fault signal of the neural network, and y_l is the mark result of the sensor fault signal.

The spectral entropy diagram is regarded as the input of the CNN. Figure 5 shows the spectral entropy diagram of aeroengine sensors under different health conditions. For open-circuit faults and normal conditions, though they look similar in Figure 5(b), we found when we enlarged the image locally that there were notable differences between them that were recognized by the CNN. In other words, each health condition has a distinguishable degree that the CNN can recognize. As shown in Figure 5(b), when $\lambda_{\max, \text{op}} = 25$, there was considerable discrepancy between signals within the range of $\lambda \in (17, 25)$; the signals within the range of $\lambda \in (1, 16)$ showed a small degree of differentiation.

The spectral entropy diagrams were processed so as to enhance the classification accuracy. We intercepted the spectral entropy diagram within the range of $\lambda \in (17, 25)$ as the input of the CNN, as shown in Figure 9. The output of the CNN in this case is the classification result. We selected optimal parameters via PSO according to the misclassification rate of the CNN, which increases the distinction between different faults and is conducive to feature-

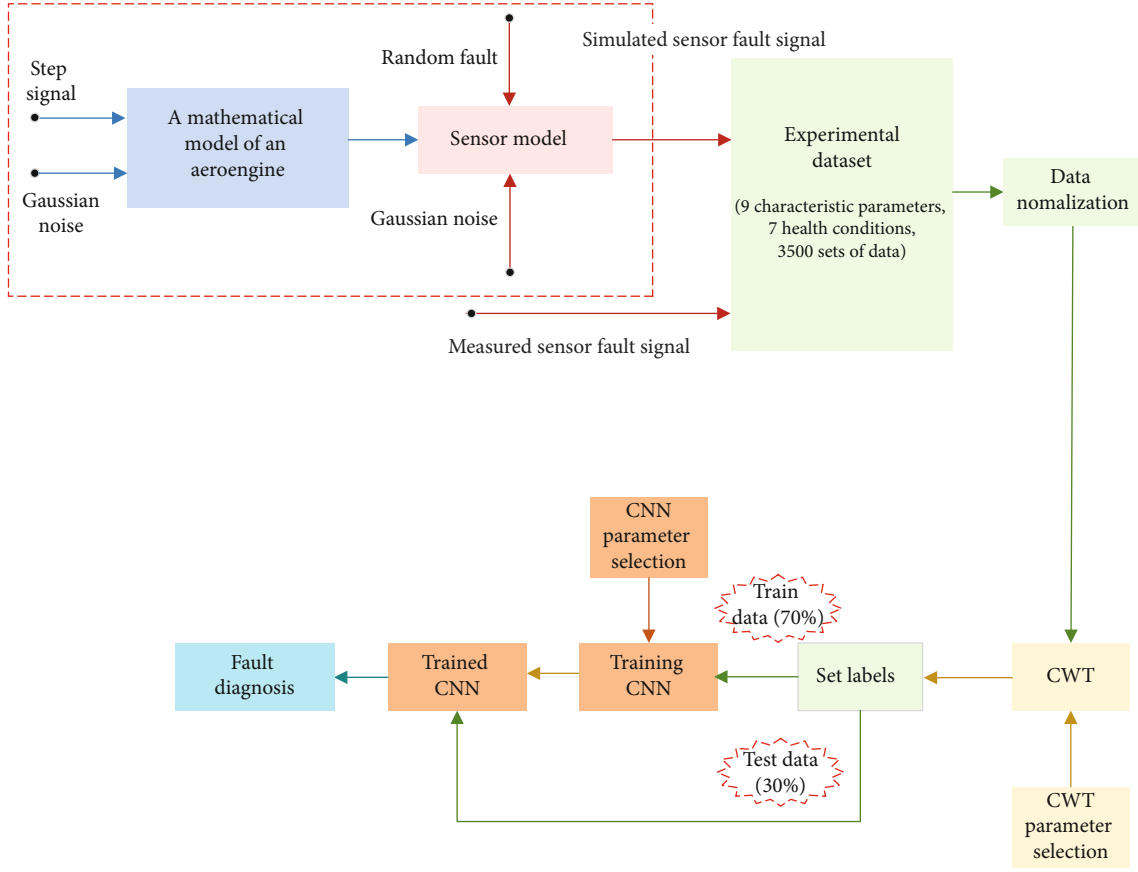


FIGURE 10: A flowchart of experiment.

TABLE 2: Common sensor types, fault causes, and simulation methods.

Fault	Reason	Simulation method
Short-circuit	Pollution caused by bridge road erosion line short connection	Signal is close to 0.1
Open-circuit	Signal line is broken; chip pin is not connected	Signal approaches maximum
Spike	Random disturbance in power supply and ground wire, surge	Add a pulse signal to original signal
Bias	Bias current or bias voltage	Add a small constant or random signal to original signal
Drift	Temperature drift	Signal is offset by a certain rate
Periodic disturbance	50 Hz interference from power supply	Signal of a certain frequency superimposed on original signal

learning and pattern recognition, thus improving the fault diagnosis accuracy.

The CNN is a supervised machine learning technique that needs tagged inputs. The spectral entropy diagram should be labeled according to the marking rules given in Section 2.1. Considering the complexity of the network, we compressed the image so that the size of the processed time-frequency graph was the (224×224) pixel size best managed by the CNN.

2.5. Performance Evaluation Function. It is generally acknowledged that the quality evaluation of the CNN model after training is an important issue. The classification accu-

racy rate is a powerful tool to assess the performance of any CNN model. When training the deep network here, we calculated the classification accuracy of the training dataset after each iteration and then conducted reliability tests to determine the accuracy of the test dataset. "Classification accuracy" refers here to the ratio of the correctly identified sample number to the total sample number.

$$AC(y_c, y_1) = \frac{\text{sum}(y_c == y_1)}{N}, \quad (11)$$

where $AC(y_c, y_1)$ is the accuracy rate, the summing function

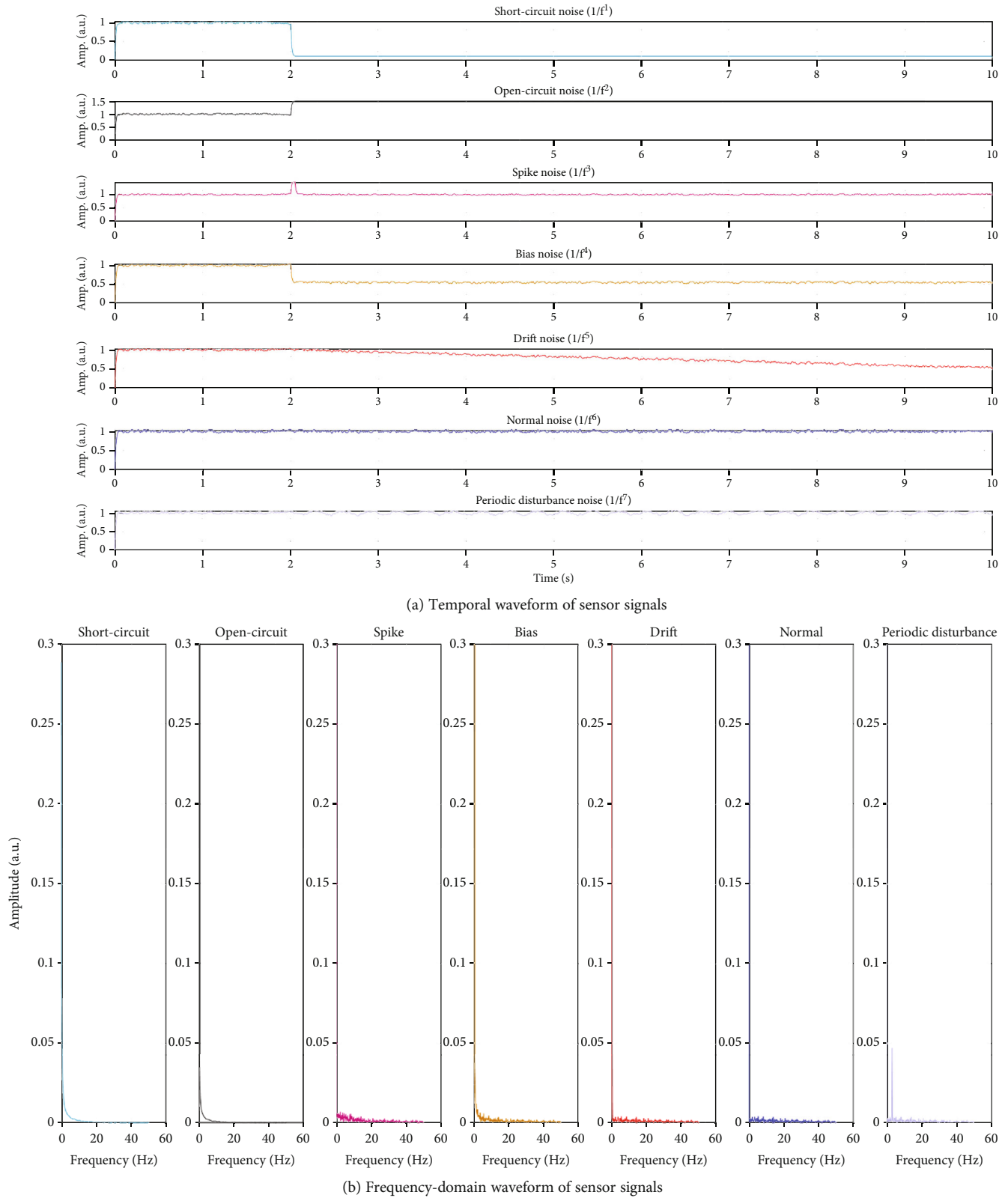


FIGURE 11: Simulated sensor signals for 7 health conditions.

is the fault classification result of the neural network output, y_c and sensor fault signal y_l have the same value, and N is the dimension of y_c .

The sensor signals under seven healthy states of aeroengine control system were standardized, and then the features of the processed signals were extracted by IPGSE to generate

TABLE 3: Network parameters.

Name	Parameter	Output feature size
Input layer	Input data	$3 \times 224 \times 224$
Convolutional layer (C1)	6 filters, size 5×5 , stride 1	$6 \times 220 \times 220$
Pooling layer (P1)	Filter size 2×2 , stride 1	$6 \times 110 \times 110$
Convolutional layer (C2)	16 filters, size 5×5 , stride 1	$16 \times 106 \times 106$
Pooling layer (P2)	Filter size 2×2 , stride 1	$16 \times 53 \times 53$
Fully connected layer (F1)	120 nodes, dropout = 0.5	44944×120
Fully connected layer (F2)	84 nodes, dropout = 0.5	120×84
Fully connected layer (F3)	7 nodes	84×7
Output layer	Output data	7×1
Learning rate	0.00004	None
Training dataset	2450	None
Test dataset	1050	None
Iteration times	35000	None
Scale factor $\lambda_{\max, \text{op}}$	25	None

the spectral entropy diagram. Next, we trained the CNN model. The input of the CNN was the marked spectral entropy diagram and the output was the health of the sensor. We evaluated the quality of the trained CNN model to find that it met the relevant requirements for fault detection and isolation.

3. Experiment and Discussion

Model validations such as the one we conducted in this study are reasonable evidence of the diagnostic capability of a given method. We tested the effectiveness of our method on the sensor faults of an aeroengine control system as well as comparing it against other state-of-the-art methods (IPGSN+BP network). A flow chart of this experiment is given in Figure 10.

3.1. Acquisition of Experimental Dataset. We established an experimental dataset to test the proposed method. Deep learning requires numerous datasets with different instances. A low data quantity tends to result in underfitting. Increasing the amount of data can remedy this. In addition, most aeroengine control systems are under normal operation conditions most of the time—fault conditions are rare—so fault data and normal data tend to be asymmetrical, leaving few actual fault data to train the model efficiently. A model with poor generalization ability can result in incorrect diagnosis results. So we need to obtain a sufficient quantity of analog sensor signals under various faults.

In other words, part of the experimental dataset is the simulated sensor signal. The remainder of the experimental data was collected from flight measurement recordings of an aeroengine under different working conditions and fault patterns.

We first establish a simulation model of the sensor. The second-order inertial link [6] was adopted to establish the sensor simulation model by referring to previous studies. Its

transfer function is as follows:

$$G(s) = \frac{w_n^2}{s^2 + 2\xi w_n s + w_n^2} \cdot e^{-\tau s}, \quad (12)$$

where $\xi = 1.25$, $w_n = 9$, $\tau = 0.12$.

The input signal of the sensor simulation model comes from the C-MAPSS aeroengine simulation model. The engine model operates at a ground design point and high altitude off-design point, respectively. The output signals of the sensor model were processed in accordance with the method in Table 2 for fault simulation. Environmental noise (Gaussian white noise) was then injected randomly into the input and output signals to make the experimental data as realistic as possible and to realistically reflect the robustness of the proposed method.

We randomly inject faults to generate various sensor fault signals, as our data was stochastic and initially asymmetrical. We changed the fault occurrence time, fault pattern, fault degree, fault period, and other parameters at random in the experiment. We did this because the resulting bias fault is affected, for example, when the value of the set bias is different; similarly, the peak fault differs when the impulse response is different. A random slope setting can result in a variety of drift faults and a variety of periodic disturbance faults can be obtained by randomly setting the period and amplitude of the disturbance.

Figure 11 shows the temporal waveform and frequency-domain waveform of the fault analog signal. The time-domain waveform of complex signals shows nonlinear and nonstationary characteristics. When the sensor faults, the time-domain waveform presents obvious impact characteristics. We could not accurately identify fault types based on the time waveform and spectrum alone, so we extracted the fault characteristics from sensor signals and subjected them to pattern recognition to distinguish the different fault types.

According to the sensor measurements and degree of sensitivity to aeroengine control system fault, we conducted an independence analysis between the measured value and influence degree of the measured values of engine work. We selected an aeroengine control system with nine key measurable sensors: the throttle lever angle PLA, inlet temperature T_1 , low-pressure rotor speed N_1 , high-pressure rotor speed N_2 , compressor outlet pressure P_3 , low-pressure turbine outlet pressure P_5 , compressor inlet temperature T_{25} , high-pressure turbine outlet temperature T_{45} , and low-pressure turbine outlet temperature T_5 . These parameters are considered appropriate indicators of engine performance changes.

We sampled 1000 data points to maintain the original precision of the signal. The time interval between any two sample points was 10 ms. This implies that the signal size of each sensor is $1 * 1000$. Each fault dataset was constructed with 500 sample data resulting in a total of 3500 sets of data. The dataset we used in this experiment includes actual parameter recordings and analog signals of the engine under various working and flight conditions. As described above, this experimental dataset contains a large number of simulated fault signals obtained through the above operation. The remainder of the experimental dataset was collected from flight measurement recordings of an aeroengine under different working conditions and fault patterns. A usable sample dataset was obtained after mixing these two groups of data randomly. And then we adopted a cross-validation strategy to train the CNN network as effectively as possible. We also arbitrarily partitioned the original sample dataset into training and test sets accounting for 70% and 30% of the total, respectively.

3.2. Experimental Settings. Before training the CNN, we set the initial parameters and important feature parameters including the initial data segment size, filter size, number of convolutional layer filters, and learning rate. Our CNN structure was designed with an input 3-channel RGB image of 224×224 , 2 convolutional layers, 2 pooling layers, and 3 full connection layers plus a dropout layer in the latter 2 full connection layers. The learning rate of the training process was 0.000015, and the dropout layer was 0.5. We set the iteration termination condition to 35,000 times. The parameters to be set are listed in Table 3.

3.3. Results and Discussion. Under normal circumstances, an accuracy curve can be drawn to represent the performance of a trained CNN model. Accuracy rate and mean computation time are used as indicators of evaluating method performance. Note that the mean computation time consists of training time and testing time. Here, we calculated the classification accuracy rate of each iteration, then adjusted the bias after each training until the iteration was terminated, and then drew the accuracy curve of IPGSE+CNN shown in Figure 12. The training iteration times and accuracy rate are the x -coordinate and the y -coordinate in Figure 12, respectively. The results shown in Figure 12 indicate that our method can accurately diagnose faults.

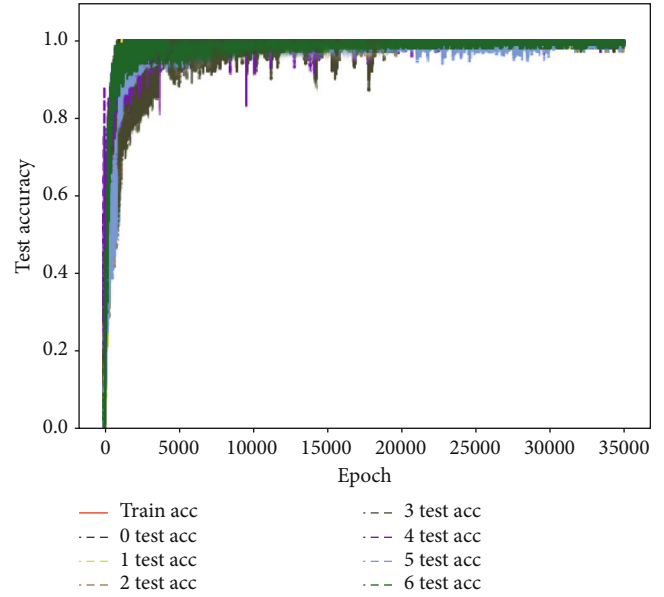


FIGURE 12: Accuracy rate curve of IPGSE+CNN.

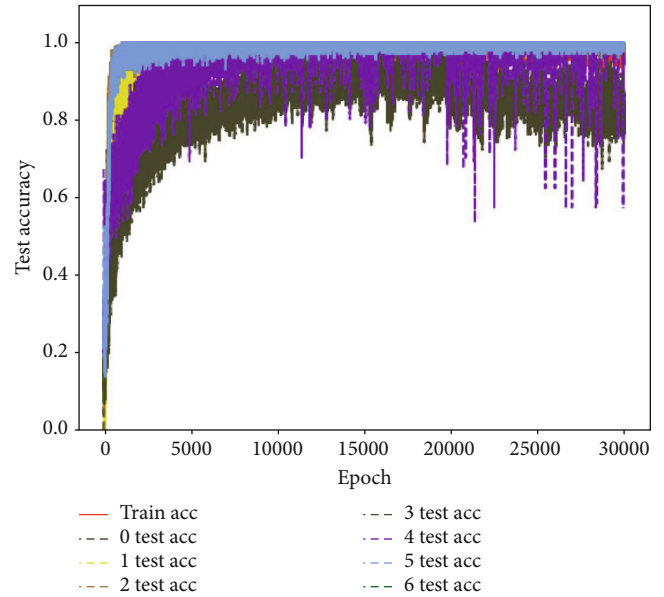


FIGURE 13: Accuracy rate curve of CNN.

First, in order to verify the necessity of IPGSE, we compare the method proposed in this paper with a fault diagnosis scheme based on CNN that directly takes raw data as input. The accuracy curve of CNN that directly takes raw data as input is shown in Figure 13. It can be seen from the comparison between Figures 13 and 12 that the accuracy of directly constructing CNN is obviously not as high as the method proposed in this article. From Table 3, the computation time of CNN is 313 s, and the computation time of IPGSE+CNN is 256 s. This shows that IPGSE that we introduce successfully reduces the computational load and computation time of CNN. This is because IPGSE converts one-dimensional signals into spectral entropy graphs as input to CNN to refine

TABLE 4: Network parameters.

Sensor data	Accuracy rate	Mean computation time
IPGSE+CNN		
Training data	97.3%	256
Short-circuit fault test set	97.5%	
Open-circuit fault test set	96.7%	
Spike fault test set	97.09%	
Bias fault test set	98.02%	
Drift fault test set	98.58%	
Normal fault test set	96.86%	
Periodic disturbance fault test set	99.6%	
IPGSE+BP		
Training data	93.3%	247
Short-circuit fault test set	93.5%	
Open-circuit fault test set	95.65%	
Spike fault test set	93.07%	
Bias fault test set	95.2%	
Drift fault test set	94.56%	
Normal fault test set	94.7%	
Periodic disturbance fault test set	93.8%	
CNN		
Training data	90.5%	313
Short-circuit fault test set	90.7%	
Open-circuit fault test set	93.6%	
Spike fault test set	91.97%	
Bias fault test set	87.02%	
Drift fault test set	85.9%	
Normal fault test set	93.4%	
Periodic disturbance fault test set	85.2%	

and summarize the morphological features of faults at multiple levels while retaining details, and the PSO algorithm further matches the CNN classification and makes accurate CNN diagnosis possible. Data-driven fault diagnosis schemes are generally offline training and online running. The greatly reduced training time is conducive to improving the ability of the scheme to be used for onboard fault diagnosis.

We also compared our method against other classification methods to further test its effectiveness. We applied a BP network composed of an input layer, a hidden layer, and an output layer to the same training samples and test samples, with a sigmoid activation function in the hidden layer. The input of the BP neural network in this case was the entropy value of the signal processed via IPGSE, and the output layer size was set to 7 neurons. We used classification accuracy as a metric of the diagnostic performance of the different methods. Table 4 shows the results. We found that the recognition accuracy of PGSE and BP fault diagnosis

algorithms was lower than that of the joint IPGSE and CNN fault diagnosis method we established in this study. The accuracy of the former algorithms was between 93.3% and 95.65%, while ours reached 96.7%-99.6%. On the other hand, the computation time of BP seems to be shorter than that of CNN. The application of IPGSE reduces the training time of CNN and makes up for this shortcoming. Compared to the small tolerable difference in time, the larger difference in accuracy is more important.

In a sense, fault diagnosis is a pattern recognition problem. The CNN we used in this study is a valuable classifier with good pattern recognition effects. The traditional BP network's diagnostic performance is not as good as that of the CNN. The traditional BP network is limited in its ability to express complex functions and has limitations in the generalization of model results. This defect is passed down to the optimization algorithm for parameter selection and the statistical measure for model selection in the neural network. Our experimental results show that the proposed method outperforms IPGSE and BP in terms of accuracy as well as efficiency. The classification accuracy of the combined IPGSE and CNN scheme was found to be over 96%.

Our experimental results show that this method is better than directly constructed CNN in accuracy and efficiency. This shows that our proposed scheme is useful IPGSE as a supplement to CNN. In addition, we also compared the method in this article with other classification methods (BP), which proved the superiority of the CNN we selected. Our approach remits the advantages of both IPGSE and CNN methods, making it a good choice for aeroengine fault diagnosis.

4. Conclusion

This paper proposed an intelligent scheme for fault diagnosis called joint IPGSE and CNN for aeroengine control system sensors. Our approach remits the advantages of both IPGSE and CNN methods. It does not need to build an engine model, but makes full use of process data to mine fault information for fault diagnosis. CNN provides automatic selection of representative features and automatic pattern recognition without threshold design. Pattern gradient spectral entropy is a good complement to CNN. It is not only simple to calculate, easy to implement on-line by hardware, but also has antinoise characteristics. It also converts one-dimensional signals into spectral entropy graphs as input to CNN to refine and summarize the morphological features of faults at multiple levels while retaining details. This greatly reduces the calculation and training time of CNN and makes accurate CNN online diagnosis possible. Besides, with the aim to solve the problem of traditional PGSE selecting parameters through experience, we improved PGSE to use particle swarm optimization algorithm to adaptively optimize the influencing parameters to better match CNN's pattern classification, and provide more reliable fault diagnosis. The fault diagnosis scheme was further improved to enhance its comprehensiveness as well. Experimental results show that the proposed intelligent fault diagnosis method is applicable to the online

automatic fault diagnosis of aeroengine control system sensors.

Though preliminary results reflect the practicability of the proposed method, it does show room for further improvement. For example, its performance in multiple fault diagnosis is unknown. The fault signal may be rebuilt by a deep learning algorithm, because obtaining information such as fault occurrence time and fault size is necessary to recover engine performance. These problems will be addressed in the future.

It does not need to build an engine model, but makes full use of process data to mine fault information for fault diagnosis. CNN provides automatic selection of representative features and automatic pattern recognition without threshold design.

Data Availability

The aeroengine control system sensor fault data used in this study are included within the supplementary information files.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work is supported by the National Science and Technology Major Project (2017-V-0011-0062) and the Fundamental Research Funds for the Central Universities (Grant No. 31020190MS702).

Supplementary Materials

The supplementary material file contains the experimental sensor signal set used to produce the dataset upon which we validated the proposed method (Section 2). The datasets were sourced from the records of a certain type of aircraft engine under various fault types and the simulation experiments of MATLAB. The data is given in .mat format including 70% training data and 30% test data, respectively. The dataset can be divided into seven types: normal data and six types of fault data. More details can be found in the data_description.docx in the supplementary material. (*Supplementary Materials*)

References

- [1] L. Gou, Z. Liu, D. Fan, and H. Zheng, "Aeroengine robust gain-scheduling control based on performance degradation," *IEEE Access*, vol. 8, pp. 104857–104869, 2020.
- [2] L. Gou, Y. W. Shen, H. Zheng, and X. Zeng, "Multi-fault diagnosis of an aero-engine control system using joint sliding mode observers," *IEEE Access*, vol. 8, pp. 10186–10197, 2020.
- [3] J. Li and W. Qu, "Aero-engine sensor fault diagnosis based on convolutional neural network," in *2018 37th Chinese Control Conference (CCC)*, pp. 6049–6054, Wuhan, China, July 2018.
- [4] P. M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy," *Automatica*, vol. 26, no. 3, pp. 459–474, 1990.
- [5] H. M. Zhao, D. Y. Li, W. Deng, and X. H. Yang, "Research on vibration suppression method of alternating current motor based on fractional order control strategy," *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, vol. 231, no. 4, pp. 786–799, 2016.
- [6] L. Gou, H. Li, H. Zheng, H. Li, and X. Pei, "Aeroengine control system sensor fault diagnosis based on CWT and CNN," *Mathematical Problems in Engineering*, vol. 2020, Article ID 5357146, 12 pages, 2020.
- [7] Y. Kai, Z. Shisheng, N. Yuan, and L. Zhen, "Research on aero-engine fault detection technology based on BP neural network," *AeroEngine*, vol. 43, no. 1, pp. 53–57, 2017.
- [8] J. Lu, J. Huang, and F. Lu, "Sensor fault diagnosis for aero engine based on online sequential extreme learning machine with memory principle," *Energies*, vol. 10, no. 1, p. 39, 2017.
- [9] Y. Li, C. Wu, and X. Xu, "A fault sensor method for vibration sensor based on PCA and SVM," *Automation Instrumentation*, vol. 40, no. 10, pp. 48–52, 2019.
- [10] R. Chen, S.-K. Guo, X.-Z. Wang, and T.-L. Zhang, "Fusion of multi-RSMOTE with fuzzy integral to classify bug reports with an imbalanced distribution," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 12, pp. 2406–2420, 2019.
- [11] X. Wang, S. Huang, L. Cao, D. Shi, and P. Shu, "LSSVM with fuzzy pre-processing model based aero engine data mining technology," in *Advanced Data Mining and Applications. ADMA 2007. Lecture Notes in Computer Science*, vol. 4632, R. Alhajj, H. Gao, J. Li, X. Li, and O. R. Zaiane, Eds., Springer, Berlin, Heidelberg, 2007.
- [12] Z. Feng, M. Liang, and F. Chu, "Recent advances in time-frequency analysis methods for machinery fault diagnosis: a review with application examples," *Mechanical Systems and Signal Processing*, vol. 38, no. 1, pp. 165–205, 2013.
- [13] W. Deng, H. Liu, J. Xu, H. Zhao, and Y. Song, "An improved quantum-inspired differential evolution algorithm for deep belief network," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 7319–7327, 2020.
- [14] H. Alaskar, "Deep learning-based model architecture for time-frequency images analysis," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 12, 2018.
- [15] Z. Chen, J. Cen, and J. Xiong, "Rolling bearing fault diagnosis using time-frequency analysis and deep transfer convolutional neural network," *IEEE Access*, vol. 8, pp. 150248–150261, 2020.
- [16] H. Shao, M. Xia, G. Han, Y. Zhang, and J. Wan, "Intelligent fault diagnosis of rotor-bearing system under varying working conditions with modified transfer CNN and thermal images," *IEEE Transactions on Industrial Informatics*, p. 1, 2020.
- [17] S. Duan, H. Zheng, and J. Liu, "A novel classification method for flutter signals based on the CNN and STFT," *International Journal of Aerospace Engineering*, vol. 2019, 8 pages, 2019.
- [18] D. Gao, Y. Zhu, X. Wang, K. Yan, and J. Hong, "A fault diagnosis method of rolling bearing based on complex morlet CWT and CNN," in *2018 Prognostics and System Health Management Conference*, pp. 1101–1105, Chongqing, China, October 2018.
- [19] W. Deng, J. Xu, Y. Song, and H. Zhao, "Differential evolution algorithm with wavelet basis function and optimal mutation strategy for complex optimization problem," *Applied Soft Computing*, no. article 106724, 2020.
- [20] H. Zhao, S. Zuo, M. Hou et al., "A novel adaptive signal processing method based on enhanced empirical wavelet transform technology," *Sensors*, vol. 18, no. 10, p. 3323, 2018.

- [21] L. Yang, *New Morphological Filtering Method and Its Application in Fault Diagnosis of Rotating Machinery*, Yanshan University, 2013.
- [22] B. Li, M. Gao, Q. Guo, and F. Liu, "Characterizing vibration signals utilizing morphological pattern spectrum for gear fault diagnosis," in *3rd International Conference on Mechatronics, Robotics and Automation*, Shenyang, China, April 2015.
- [23] B. Li, *Mathematical Morphological Analysis and Intelligent Classification of Mechanical Fault Signals*, National Defense Industry Press, Beijing, 2011.
- [24] B. Wang, X. Hu, W. Wang, and D. Sun, "Fault diagnosis using improved pattern spectrum and fruit fly optimization algorithm-support vector machine," *Advances in Mechanical Engineering*, vol. 10, no. 11, Article ID 1687814018810935, 2018.
- [25] P. Liu, H. Li, and B. Xu, "A performance degradation feature extraction method and its application based on mathematical morphological gradient spectrum entropy," *Vibration and Impact*, vol. 35, no. 16, pp. 86–90, 2016.
- [26] N. G. Nikolaou and I. A. Antoniadis, "Application of morphological operators as envelope extractors for impulsive-type periodic signals," *Mechanical Systems and Signal Processing*, vol. 17, no. 6, pp. 1147–1162, 2003.
- [27] Y. Dong, M. Liao, X. Zhang, and F. Wang, "Faults diagnosis of rolling element bearings based on modified morphological method," *Mechanical Systems and Signal Processing*, vol. 25, no. 4, pp. 1276–1286, 2011.
- [28] H. Ruijiang, L. Wenxiu, and C. Fulei, "Mathematical morphology extraction method of rolling bearing fault signal," *Chinese Journal of Electrical Engineering*, vol. 28, no. 26, pp. 65–70, 2008.
- [29] Q. Chen, Z. Chen, W. Sun, G. Yang, A. Palazoglu, and Z. Ren, "A new structuring element for multiscale morphology analysis and its application in rolling element bearing fault diagnosis," *Journal of Vibration and Control*, vol. 21, no. 4, pp. 765–789, 2015.
- [30] X. Kun, "Portability analysis of multi-layer convolutional neural network deep learning algorithm," *Journal of Harbin Engineering University*, pp. 1–6, 2020.
- [31] N. Passalis, J. Raitoharju, A. Tefas, and M. Gabbouj, "Efficient adaptive inference for deep convolutional neural networks using hierarchical early exits," *Pattern Recognition*, vol. 105, p. 107346, 2020.
- [32] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.