Zhou, Yajing; Zheng, Yuemin; Tao, Jin; Sun, Mingwei; Sun, Qinglin; Dehmer, Matthias; Chen, Zengqiang

Servo Health Monitoring Based on Feature Learning via Deep Neural Network

Published in:
IEEE Access

DOI:
10.1109/ACCESS.2021.3132046

Published: 01/12/2021

Document Version
Publisher's PDF, also known as Version of record

Received November 6, 2021, accepted November 24, 2021, date of publication December 1, 2021,
date of current version December 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3132046

Servo Health Monitoring Based on Feature Learning via Deep Neural Network

YAJING ZHOU[1], YUEMIN ZHENG[1], JIN TAO[1,2], (Member, IEEE), MINGWEI SUN[1], QINGLIN SUN[1], MATTHIAS DEHMER[1,3], AND ZENGQIANG CHEN[1], (Member, IEEE)

[1]College of Artificial Intelligence, Nankai University, Tianjin 300350, China
[2]Department of Electrical Engineering and Automation, Aalto University, 02150 Espoo, Finland
[3]Department of Computer Science, Swiss Distance University of Applied Sciences, CH-3900 Brig, Switzerland

Corresponding author: Jin Tao (taoj@nankai.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61973172, Grant 61973175, Grant 62003175, and Grant 62003177; in part by the National Key Research and Development Project under Grant 2019YFC1510900; in part by the Key Technologies Research and Development Program of Tianjin under Grant 19JCZDJC32800; in part by the China Postdoctoral Science Foundation under Grant 2020M670633; and in part by the Academy of Finland under Grant 315660.

ABSTRACT As the core actuator of an aircraft's flight control system, the servos' reliability directly affects the safety of the flight control system and the whole aircraft. The failure of the rudder will lead to the poor control effect of aircraft, affect its flight quality and safety, and even cause major flight accidents. In order to monitor the health status of servo and determine the fault and its degree accurately, this paper presents a feature learning based health monitoring method using a deep neural network. Firstly, we combine the wavelet packet decomposition and support vector machine to synthesize the sample segment label. And then, the sliding window is employed to enlarge the sample size, and the auto-encoder is utilized to reduce the data dimension. Moreover, the Softmax classifier is used for health monitoring. At last, the numerical simulations demonstrate the effectiveness of the proposed method.

INDEX TERMS Servo health, wavelet packet decomposition, auto-encoder, softmax classifier, health monitoring.

I. INTRODUCTION

Servos are the executive mechanism of aircraft attitude control, which can drive the three main control surfaces of elevator, aileron, and rudder, and the auxiliary control surfaces of flap, slat, and spoiler [1]. Once the steering gear fails, especially the rudder, elevator, and other main control steering gears, it will pay not only economic costs but also cause serious air accidents such as engine damage and death, resulting in heavy losses [2]. Therefore, the daily health monitoring of servos is helpful to improve the reliability of the aircraft and reduce the cost of scheduled maintenance.

For aircraft, failure means that one or more performance indicators are abnormal, leading to its failure to complete the task [3]. The typical failure types of the rudder include the stuck rudder surface, the damaged rudder surface, and the loose rudder surface, which will affect the moment coefficient of the rudder [4]. Therefore, the health monitoring of steering gear has become an important issue, and it has also become the focus of academic and engineering circles. Traditional health monitoring methods include model analysis, signal feedback, signal processing, expert system, etc.

Since 1983, the development of artificial intelligence technology has provided various means and methods for health monitoring. Because it is difficult to measure and quantify the health status of bearings and gears in many cases, many vibration-based methods have been proposed to construct the health indicators of bearings and gears [5]. For example, Yan and Jia proposed a novel fault classification algorithm based on optimized support vector machine (SVM) with multi-domain feature to improve intelligent diagnostic accuracy of rolling bearing [6]. Deng *et al.* proposed a new fault diagnosis method of motor bearing based on empirical wavelet transform, fuzzy entropy, and SVM – EWTFSFD to achieve the fault diagnosis of motor bearing [7]. Liu *et al.* proposed an integrated multi-sensor fusion-based deep feature learning approach to identify the fault severity in rotating machinery processes [8]. Tang *et al.* focused on convolutional neural network (CNN)-based fault diagnosis approaches in rotating machinery [9], where several main

The associate editor coordinating the review of this manuscript and approving it for publication was Fanbiao Li.

techniques applied in CNN-based intelligent diagnosis, principally including the fast Fourier transform, wavelet transform, data augmentation, S-transform, and cyclic spectral analysis. Huang *et al.* advanced a compound fault intelligent diagnosis method based on deep decoupled convolution neural network [10]. To investigate how deep learning can be applied to infrared thermal video to automatically determine the condition of the machine, a method from the subfield of feature learning, namely deep learning, and more specifically convolutional neural networks, was researched [11]. Yang *et al.* proposed a fault diagnosis scheme combined of hierarchical symbolic analysis (HSA) and convolutional neural network (CNN), which achieved laborsaving and timesaving preliminary feature extraction, and accomplished automatically feature learning with simplified network architecture [12]. In order to reduce downtime or avoid the failure of rotating machinery, an integrated approach of an Adaptive Neuro–Fuzzy Inference System (ANFIS) and Dimensional Analysis (DA) was demonstrated to diagnose the size of the bearing faults [13]. Ajagekar and You applied Quantum computing (QC) based deep learning methods for fault diagnosis, which exploited their unique capabilities to overcome the computational challenges faced by conventional data-driven approaches performed on classical computers [14]. In order to improve network training, a residual learning algorithm was advanced [15]. Qi *et al.* proposed a novel self-decision fault diagnosis model for power transformer, which combined the characteristics of faults and adaptability of conventional deep brief network [16]. Contractive autoencoder (CAE) can easily grasp the internal factors and directly obtain the hidden robust features. So Shen *et al.* raised a method based on stacked contractive autoencoder (CAE) for automatic robust features extraction and fault diagnosis of rotating machinery [17]. When testing data in machine fault conditions are not available for training, a novel cross-domain fault diagnosis method based on deep generative neural networks can provide reliable cross-domain diagnosis result [18]. The methods mentioned in the introduction are only from a single frequency domain or abstract domain for fault diagnosis. They can't obtain failure degree of steering gear or large number of data samples must be required.

This paper considers both frequency domain and abstract domain for fault diagnosis. In the paper, a feature learning based health monitoring method using a deep neural network is proposed. For the shortage of sample size, a suitable data expansion algorithm is designed. The wavelet packet decomposition and power spectral density are used to extract the data features in the frequency domain, and then the sliding window and SVM training model is used to expand the samples and roughly integrate the labels. The automatic encoder reduces the dimension of the sample data, that is, abstractly extract data features. And the softmax classifier monitors the health condition of the actuator. By adjusting the parameters, the satisfactory performance index is finally obtained. The main contributions of this paper are summarized as follows:

TABLE 1. Example of steering gear data.

Sampling point	Position input	Position feedback	U-phase current feedback	V-phase current feedback
1	0.00°	0.00 °	0.00 A	0.00 A
2	1.00°	0.50 °	0.24 A	0.22 A
3	1.00°	0.87 °	0.16 A	0.18 A
4	1.00 °	1.00 °	0.02 A	0.01 A
...	...	...	...	...

- The sliding window and support vector machine in frequency domain are used to integrate the extended labels of samples.
- Reduce the dimension of data by automatically extracting features using an autoencoder of the abstract domain.
- Softmax classifier is used for health monitoring to get whether the failure and its degree.

The rest of the paper is arranged as follows. Section II describes the data source and data preprocessing methods. Section III discusses feature extraction of servo faults from two domains. In Section IV, two health monitoring methods are presented. Section V shows the simulation results and evaluation of servo health monitoring before Section VI concludes this paper.

II. DATA SOURCE AND PREPROCESSING

A. DATA SOURCES

The data used in this paper are the measured data of the daily maintenance of a specific type of steering gear. The original data of the steering gear has been calibrated, but some data have not been calibrated. Specifically, 17 available samples of the rudder were used, including 8 standard steering gear and 9 jitter steering gear. We can see that the sample size is relatively small. Therefore, the sample size needs to be expanded.

In the source data, there are four rudders in each data table. Several rudders are under test, and the rest of them do not operate. The interval between sampling points is 5ms. The data example is shown in Table 1. The position input, position feedback, and current feedback of four actuators are arranged horizontally. Each actuator is tested by U and V two-phase current, and each phase current has five feedback points. Therefore, each actuator has 10 current feedback, and the feedback characteristics of these 10 are similar.

B. DATA PREPROCESSING

1) DATA EXPANSION

In order to simplify the work, the unknown actuator is calibrated by Origin software. Since there are four actuator data in each data table, but not all actuators are tested, it is necessary to extract adequate information and re-integrate it. Due to the limited data of daily maintenance of steering gear, there are only 17 valid data samples of the steering gear. Even the SVM suitable for small samples cannot be used for machine learning and testing. Neural networks and other methods need a large number of samples, so it is necessary to expand the samples. There are 10 current feedbacks in the test

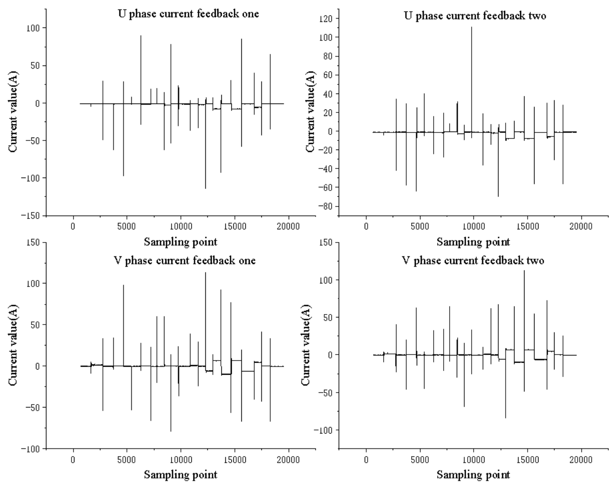


FIGURE 1. Comparison with current feedback of steering gear.

data of each actuator: 5 for U-phase current feedback and 5 for V-phase current feedback. Four of them are selected and plotted with Origin for comparison. The results are shown in Figure 1.

It is obvious to see from Figure 1 that although the multiple current feedback of the actuator is not the same in dynamic states, the performance of the actuator in steady states is the same. If each servo uses 10 groups of current waveforms to extract features and then summarize them, it will cause a lot of information redundancy. Therefore, there are only two types of data used in this paper: normal and fault. In fact, the differences between the normal steering gear and the fault steering gear are mainly the characteristics when the steering gear angle is in a steady-state rather than in a transition process of regulation. Therefore, when there is no difference in the steady-state performance of all current feedbacks, the 10 current feedbacks in each data table can be regarded as different 10 test servos, which increases the sample size of the actuator and avoids data redundancy to a certain extent. This method expands the data sample size, and 170 samples are obtained, including 80 normal samples and 90 fault samples. The sample size is enough for the SVM method.

However, deep neural networks (DNN) need more samples. Here, 170 samples are not enough to support the learning of DNN, so we need to expand the sample size further. Because the data of the servos consist of normal steering gear and faulty actuator and the fault characteristics are in steady-state stage. Therefore, the transition time of the motion control system is often short. Only a period of signal extraction can contain fault characteristics. Thus, the selected sample data of the actuator shall include position command, position feedback, U-phase current, and V-phase current for a while. Nevertheless, the duration of each rudder sample test is different, and the performance reflected in the data is that the dimensions of each sample are different, and the data dimensions of the rudder with long time tested are more. In order to unify the data dimensions, reduce the number of dimensions of a single sample, and further expand the sample size, the existing samples should be segmented.

After determining the sample standard format and sampling content, in order to further increase the sample size, the rolling window method is used to expand the sample based on the simple sample segmentation method. The basic idea of this method is to take a certain proportion of overlapped data when the original actuator is divided by every 1000 sampling points. The window size is set to 1000×4. Roll sampling in the original steering gear data table and the data overlap rate is set to 80%. The 17 sample actuator data set is successfully expanded to 4095 sample data set through these data preprocessing, which lays the data foundation for the training of DNNs.

2) ELIMINATION OF CURRENT SPIKE

As the actuator of the motion control system, the servos have the characteristics of rapid response. When the input of the control system changes, the control system adjusts quickly. Consequently, the current increases rapidly, the transient time is very short, and the sampling time is large to generate the current spike pulse. Although the control system regulation generates the current pulse, the existence of the pulse will significantly affect the feature extraction effect of subsequent wavelet packet decomposition, so filtering methods are needed to remove them.

The change of position command input causes the current pulse of actuator data, so the change of position command is related to the arrival time of the current pulse. It is easy to determine the rising edge and falling edge position when the position command is changed. Due to the pure time lag, the arrival of the current spike will delay several sampling points. If the position command information is used, the fixed-point denoising of the current spike can be completed.

The position commands in the data set used in this paper are all position inputs in the form of steps. Looking for the rising edge and falling edge is the time to find when the position command can be changed compared with the previous one.

Suppose a column of position command data is written as $A = \{X_1, X_2, X_3, X_4, \ldots, X_i\}$. The other column of position command data is arranged as $B = \{X_1, X_1, X_2, X_3, X_4, \ldots, X_{i-1}\}$ Make a difference between A and B, we obtain $C = A{-}B = \{0, X_2 - X_1, X_3 - X_2, \ldots, X_i - X_{i-1}\}$.

If B the same as the previous position command A, the corresponding position of C will be 0, Vice versa. Then the position that is not 0, where the position command changes, needs to be found. The idea of extracting the change point of position instruction is a kind of matrix calculation, which significantly reduces the calculation time compared with directly writing the loop to judge whether it is the same as the previous point. After finding out the position of the change points, the current is denoised at the fixed point. Due to a certain pure delay, the current values of the last 10 sampling points are all set as the sampling values

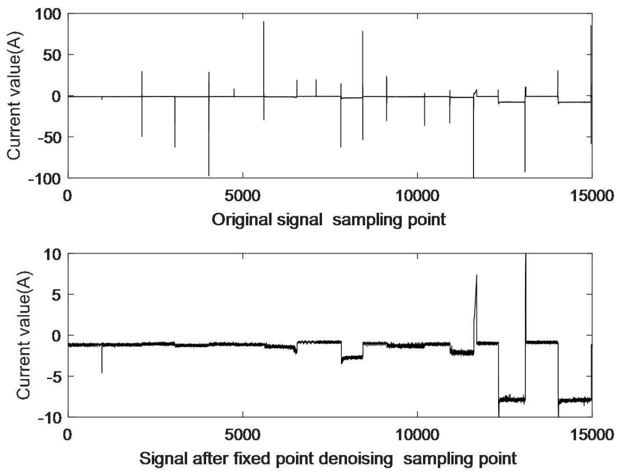


FIGURE 2. Fixed point denoising effect.

Y_j $(n+1 < j < n+11) = Y_j$. So far, the denoising at a fixed point according to the position instruction is completed.

Take a group of servo data as an example to show the effect of fixed-point denoising based on position commands. The comparison between the original signal and fixed-point denoising based on position command is shown in Figure 2. It can be seen that the effect of suppressing spike pulse is excellent.

III. FEATURE EXTRACTION OF SERVO FAULTS

A. FREQUENCY DOMAIN FEATURE EXTRACTION

The vibration signal of the daily maintenance of the servo needs to extract features in the frequency domain before the model can be established. In classical signal processing, multiscale decompositions produced by discrete wavelet transform (DWT) provide convincible solutions for many problems such as denoising and compression [19]. However, it still has some limitations as it cannot decompose the high-frequency detail signal. The wavelet packet decomposition (WPD) is a classical signal processing method, which can decompose the signal into the appropriate components and detailed components [20].

Mallat tower decomposition is a typical fast algorithm among the wavelet transforms, which can continuously carry out binary orthogonal decomposition through high pass filter and low pass filter. Figure 3 is the schematic diagram of three-layer wavelet packet decomposition.

Using the "db20" wavelet basis function to decompose the original signal data into three layers of wavelet packet, eight wavelet nodes can be obtained. Then the waveforms of each frequency band are obtained by wavelet packet reconstruction, as shown in Table 2 and Table 3. It can be seen that the proportion of energy in the low-frequency band is high, while that in the high-frequency band is low. And the proportion of low-frequency energy of normal steering gear is higher than that of dithering steering gear, so energy can be used as a feature to distinguish normal steering gear from the faulty steering gear.

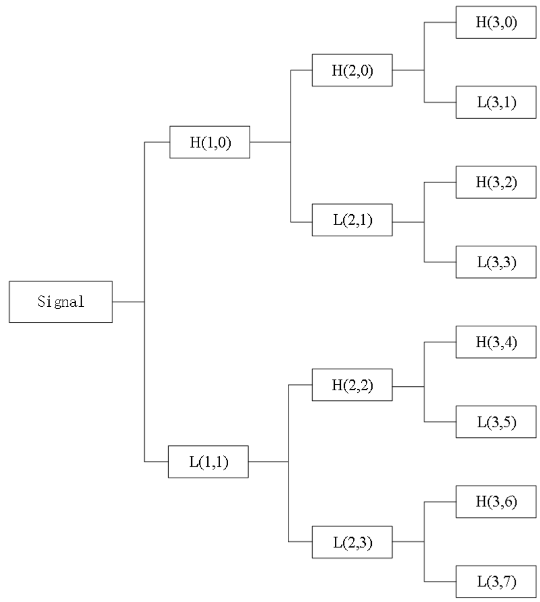


FIGURE 3. Three-layer wavelet packet decomposition.

TABLE 2. Wavelet packet feature extraction result of normal steering gear after denoising.

Number of nodes	Frequency band	Proportion of energy
1	0~25 HZ	99.7016%
2	25~50 HZ	0.1198%
3	50~75 HZ	0.0290%
4	75~100 HZ	0.0550%
5	100~125 HZ	0.0182%
6	125~150 HZ	0.0261%
7	150~175 HZ	0.0255%
8	175~200 HZ	0.0249%

TABLE 3. Wavelet packet feature extraction result of jitter steering gear after denoising.

Number of nodes	Frequency band	Proportion of energy
1	0~25 HZ	62.1666%
2	25~50 HZ	1.6669%
3	50~75 HZ	1.7312%
4	75~100 HZ	1.6060%
5	100~125 HZ	6.4412%
6	125~150 HZ	22.3764%
7	150~175 HZ	1.0844%
8	175~200 HZ	2.9273%

The power spectral density is a very important property for vibration signal and has the advantage of fast calculation [21]. Thus, the combination of wavelet packet decomposition and power spectral density is selected to extract features. The power spectral density and the proportion of energy characteristics are calculated as

$$E = \sum I^2 \tag{1}$$

$$P(i) = \frac{E(i)}{\sum E(i)}, \quad i \in [1, 8] \tag{2}$$

where I is the current. E denotes the energy. P represents the energy proportion of each frequency band, and i is the number of nodes.

The power spectral density of the reconstructed waveform of each frequency node is calculated, and then the total energy of the signal is calculated to obtain the energy proportion of

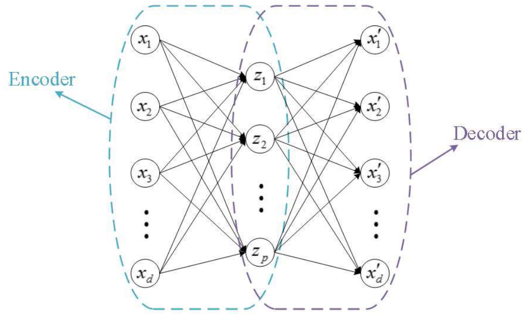


FIGURE 4. Structure of the autoencoder.

each frequency band as the feature input to the SVM. After putting these features into the SVM for training, the fault prediction model can be obtained. Then the data expansion and label sets are completed to achieve the data preprocessing.

B. ABSTRACT FEATURE EXTRACTION

In Section III-A, the method of using wavelet packet decomposition to extract vibration characteristics of the steering gear is introduced, which can effectively distinguish normal steering gear from fault. Nevertheless, this method needs professional knowledge as support. In order to reduce the demand for professional knowledge, this section continues to study the autoencoder of unsupervised learning for abstract feature extraction. Because the extracted features are abstract, we cannot know their practical significance, so we cannot verify its feasibility from feature extraction results, so we need to combine the classifiers to explain. The calibration data is input into the autoencoder for training. When the network output and input are approximately equal, the encoder is used to reduce the dimension of the data.

Autoencoder (AE) is an unsupervised learning method in machine learning. Its basic structure consists of an input layer, several hidden layers, and an output layer, a widely used deep learning model [22], and is mainly used for data dimensionality reduction and feature extraction. First, AE encodes the input features and then decodes them. The encoding process continuously reduces the dimension to extract the critical information and restores the initial data by improving the dimension during decoding [23].

Suppose that there is a set of n-dimensional samples $x^{(n)} \in R^d, 1 \leqslant n \leqslant N$, where N represents the number of samples. The encoder maps this set of data into the feature space, so that $z^{(n)} \in R^p, 1 \leqslant n \leqslant N, p$ is the dimension of the feature space. The simplest autoencoder structure is shown in Figure 4, and the layers are fully connected.

For sample x, the encoding of the active value of the autoencoder hidden layer is presented as

$$z = f(W^{(1)}x + b^{(1)}) \tag{3}$$

where $W^{(1)}$ is the connection weight matrix of the encoder input layer and the hidden layer. $b^{(1)}$ denotes the offset vector of the input hidden layer, and $f(\cdot)$ represents the activation function.

The output of the autoencoder is written as

$$x' = g(W^{(2)}z + b^{(2)}) \tag{4}$$

where similarly, $W^{(2)}$ denotes the connection weight matrix between the encoder hidden layer and the output layer. $b^{(2)}$ is the offset vector of the output hidden layer, and $g(\cdot)$ denotes the activation function.

In the autoencoder network, the activation function used is Rectified Linear Unit (ReLU). This is a left saturation function, which alleviates the gradient vanishing and speeds up the convergent speed. ReLU is defined as

$$\mathrm{ReLU}(x) = \begin{cases} x, & x \geqslant 0 \\ 0, & x \leqslant 0 \end{cases} = \max(0, x) \tag{5}$$

In machine learning, the most commonly used optimization algorithm is gradient descent method. The parameter θ_0 can be modified as

$$\theta_{t+1} = \theta_t - \alpha \cdot \frac{1}{N}\sum_{n=1}^{N} \frac{\partial L(y^{(n)}, f(x^{(n)}; \theta))}{\partial \theta} \tag{6}$$

where θ_t is the parameter value of the t-th iteration, and α is the search step, also known as the learning rate. It is a key parameter in neural network optimization.

At present, many adaptive learning rate designs have been proposed, and different learning rates are set for each parameter. In this case, the Root Mean Square Prop (RMSprop) algorithm is used. The algorithm proposed by Geoff Hinton [24] can avoid the premature attenuation problem caused by the monotonous decline of the learning rate.

IV. HEALTH MONITORING METHODS

A. CONDITION MONITORING BASED ON SVM

The feature extraction part extracted from wavelet packet decomposition is input into SVM for training, and the fault prediction model can be obtained to complete the data expansion and label set. SVM shows its unique advantages in solving small samples and nonlinear data classification problems [25]. It is a classification method based on the principle of structural risk minimization and statistical learning theory. Two principles should be followed in classification: correct classification and large enough classification interval.

The principle of traditional linear SVM is shown in Figure 5. A training data set is defined as

$$A = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)\} \tag{7}$$

where x_i is the feature vector and y_i is the data label of -1 or 1. The goal of SVM is to ensure that the sum of the geometric distance of the support vector nearest to the hyperplane is maximum.

Let the hyperplane expression satisfy the condition of $w \cdot x + b = 0$. The data set is A, then the geometric interval is $\gamma_i = \frac{1}{||w||} y_i (w \cdot x_i + b)$. Support vector means the vector with the shortest geometric interval $\gamma = \min \gamma_i (i \in [1, n])$. So the

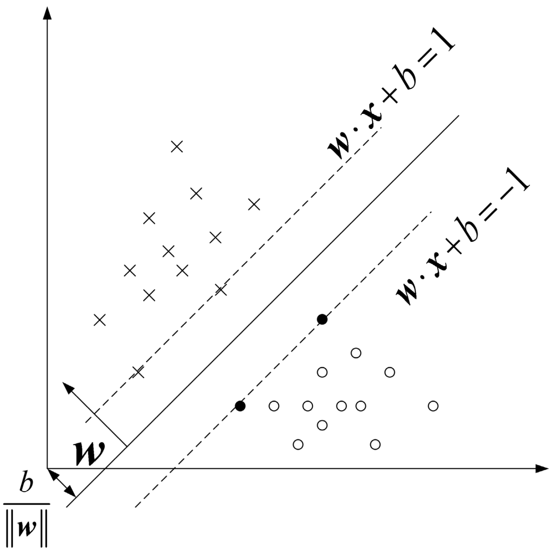


FIGURE 5. Support vector machine in linear classification.

conditions that SVM needs to meet

$$\begin{cases} \max \quad \gamma \\ s.t. \quad \frac{1}{||w||} y_i \left(w \cdot x_i + b\right) \geqslant \gamma, i \in [1, n] \end{cases} \tag{8}$$

However, in most cases, the nonlinear hyperplane is more suitable. At this point, we need to derive kernel function K, and when kernel function $K\left(x_i, x_j\right)$ meets Mercer condition, it can replace the point product in the optimal classification surface.

B. SOFTMAX CLASSIFIER MONITORING

The goal of classification is to learn a classification function or model from the manually labeled classification training samples. When there is new data, it can be categorized into a given class [26]. Classifier network structure is also composed of input layer, hidden layer, and output layer, but it is a supervised learning process, which needs one-to-one correspondence between input layer data features and discrete labels.

In the classifier network, the activation function uses ReLU and Softmax. Softmax is an activation function, which mainly realizes the function of multiple classifiers. The output is the probability distribution of each class. Multiple scalars are mapped into a probability distribution [27], and the data is classified into the category with the largest output value.

For K scalar $x_1, \cdots, x_K$, the Softmax function is defined as

$$\mathrm{Softmax}(x_k) = \frac{\exp(x_k)}{\sum\limits_{i=1}^{K} \exp(x_i)} \tag{9}$$

The cross-entropy loss function is used in the classifier network. For a random variable with the distribution of $p(x)$, the entropy $H(p)$ represents the optimal coding length. Cross entropy is the length of the real distribution p information

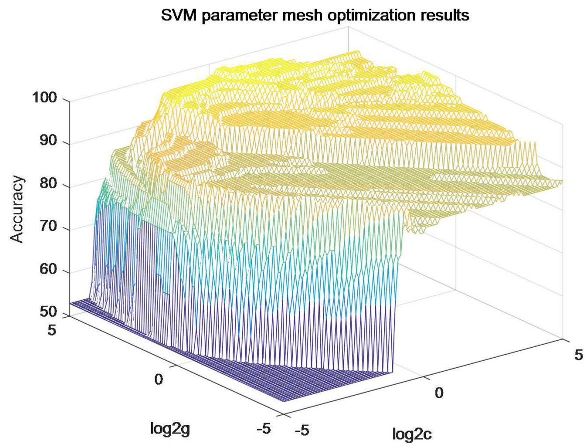


FIGURE 6. SVM mesh optimization.

encoded according to the probability distribution q optimal coding, which is defined as

$$H(p, q) = -\sum_{x} p(x) \log q(x) \tag{10}$$

Given p, the closer q is to p, the smaller the cross-entropy is. By iteratively reducing the value of cross-entropy in the network, the two are very close such that the accurate classification can be completed.

V. SIMULATION AND EVALUATION OF HEALTH MONITORING

A. SIMULATION RESULTS OF HEALTH MONITORING BASED ON SVM

SVM inputs half of the features extracted from wavelet packet decomposition to the SVM classifier for training, and then the fault prediction model can be obtained for health monitoring. Figure 6 shows the results of SVM grid optimization. It can be seen that in the grid optimization graph, large grids are above 90% accuracy, and only a few of them are in low accuracy. This shows that the feature extracted by wavelet packet decomposition has good robustness for the parameter selection of SVM. The optimal parameters are set as c=1.1487; g=2515. Figure 7 shows the SVM classification results, and the accuracy rate is as high as 98.8235% (84/85). The simulation results also show that the fault prediction time of SVM is 1.004 s, which indicates that SVM-based servo health monitoring is an accurate and efficient method.

B. SIMULATION RESULTS AND ANALYSIS OF HEALTH MONITORING BASED ON AUTOMATIC ENCODER

1) SIMULATION RESULTS

The trained model is used to expand and label the actuator data, and the coincidence rate of adjacent sliding windows is set to 0.8. Finally, 4095 × 4000-dimensional data with an 80% coincidence rate are obtained. During the training process, the training set and the test set are 70% and 30% of the total data, respectively.

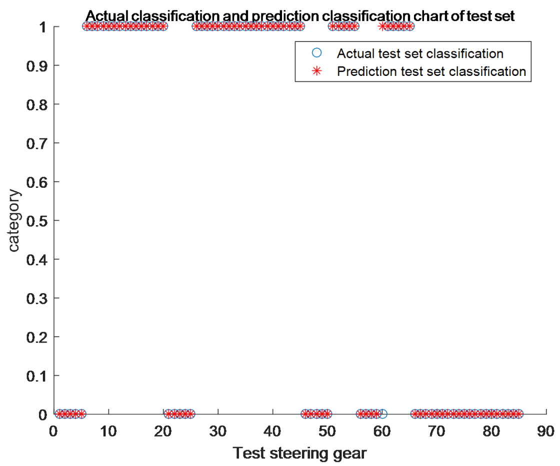


FIGURE 7. Classification results of SVM.

TABLE 4. Different parameters of encoder networks.

Network Structure	The number of layers and nodes of neural network
1	4000-3200-2400-1600-800 (0.0001)
2	4000-2400-800 (0.0001)
3	4000-3600-3200-2800-2400-2000-1600-1200-800 (0.0001)

TABLE 5. Simulation accuracy results of different neural network structures and learning rates.

Network Structure	α=0.00001	α=0.0001	α=0.001
1	0.977950±0.004361	**0.989300±0.001834**(∗)	0.976566±0.008508
2	0.976810±0.002566	**0.986493±0.002645**	0.978153±0.014197
3	0.926729±0.001637	**0.937063±0.003632**	0.926770±0.003401

TABLE 6. Loss function results for different neural network structures and learning rates.

Network Structure	α=0.00001	α=0.0001	α=0.001
1	**0.128978±0.025016**	0.147220±0.096801	1.659356±5.100259
2	**0.125428±0.035602**(∗)	0.125669±0.074836	0.337612±0.470992
3	**0.229800±0.019522**	0.183598±0.016333	0.318458±0.482950

For the optimization of the number of encoder neural network layers and the number of inter-layer nodes, the initial step size of the classifier network optimization algorithm is 0.00001, 0.0001, and 0.001, respectively, as shown in Table 4.

The average value, the covariance, the loss function and the model test time are listed in Table 5, Table 6 and Table 7, respectively.

Among Table 5 to Table 7, the bold values in the tables are the optimal result when changing the learning rate in Softmax network when the encoder network structure is fixed, while data marked with $*$ is the optimal result among all data. It can be seen from Table 5 that the accuracy of the three structures is the highest when the learning rate is 0.0001. The result of the loss function in Table 6 shows that when the network structure is relatively simple, i.e., in the case of neural network structures 1 and 2. Moreover, the smaller the learning rate is, the smaller the loss function is.

TABLE 7. Test time results of different neural network structures and learning rates.

Network Structure	α=0.00001	α=0.0001	α=0.001
1	**0.127903±0.001057**	0.128403±0.000442	0.128605±0.000793
2	**0.012335±0.001244**(∗)	0.012520±0.000558	0.012462±0.000388
3	0.174683±0.001128	**0.174814±0.000455**	0.174777±0.000450

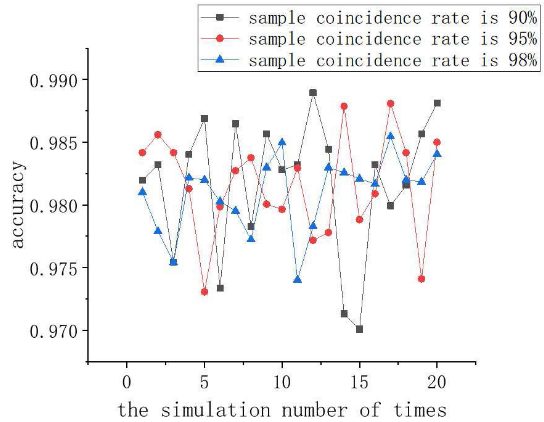


FIGURE 8. Test accuracy of extended samples.

According to the test simulation time data in Table 7, when neural network structure 2 is used, and the learning rate is 0.00001, the running time of single test data is the shortest. On the other hand, the test time of the softmax network is almost the same. For the encoder network, the model prediction time of structure 2 is short because of its simple structure, but there is little difference among the three structures. For example, the running time of encoder network structure 1 is 0.1219 s, structure 2 is 0.0061 s, and structure 3 is 0.1686 s.

For the classification model, accuracy is a crucial parameter, and there is little difference between the loss function and the test time. Therefore, we choose the network parameters with the highest accuracy as the optimal parameters. That is, the encoder uses a 4000-3200-2400-1600-800 five-layer network, and the learning rate is 0.0001. Furthermore, in order to ensure that using such network parameters can get accurate health monitoring results, we further expand the sample size and use other indicators of softmax classification to judge.

2) EXPANDED SAMPLE VALIDATION

By using SVM and sliding window to expand samples and integrate tags, the coincidence rate of samples is controlled at 80%, and 4095 samples are obtained. If we want to know whether the optimal network parameters are suitable for more samples, we need to expand the sample coincidence rate gradually. By changing the sliding window size, the sample coincidence rate is 90%, 95%, and 98%, respectively. Similarly, 20 simulations were carried out, and two evaluation indexes of accuracy and loss function were obtained, as shown in Figure 8, Figure 9, and Table 8.

We can see that the optimal network parameters are still suitable for multi-samples by expanding the samples.

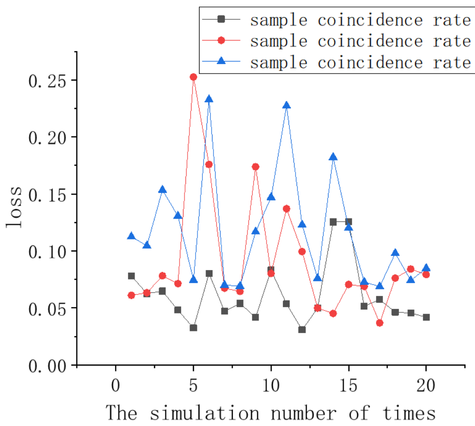


FIGURE 9. Loss function of extended samples.

Moreover, the simulation accuracy can reach 98% in both training set and test set, and there is no over-fitting phenomenon. The loss function finally achieves satisfactory results. Therefore, the optimal parameters are consistent with the case of multiple samples.

3) CLASSIFICATION PERFORMANCE EVALUATION

As the last step of classification, the performance evaluation of classifiers is indispensable. Accuracy refers to the percentage of correct classification, but it cannot indicate the potential distribution of response values and the types of classifier errors. Therefore, for the data to be classified, the classifier's performance depends on the classification effect. Based on this idea, Van Rijsbergen first proposed the evaluation indexes such as precision and recall in 1979 and obtained the recognition of many scholars. Later, receiver operating characteristic (ROC) and area under the ROC curve (AUC) were proposed successively. As for performance metrics, these two metrics have attracted more and more attention in the field of machine learning due to their superior properties [28].

For classification problems, there may be four possible classification results: TP (true positive), TN (true negative), FP (false positive), and FN (false negative). A series of evaluation indexes of classification performance can be derived from these four types of classification, such as commonly used accuracy, precision, recall, and F1 score [29]. In order to better understand these evaluation criteria, the confusion matrix shown in Table 9 is used to represent the four results of classification.

Based on Table 5, the formula of accuracy rate, recall rate and F1-score can be given as

$$P = \frac{TP}{TP + FP} \tag{11}$$

$$R = \frac{TP}{TP + FN} \tag{12}$$

$$F1\text{−}score = \frac{2P \cdot R}{P + R} \tag{13}$$

Precision rate is used to measure the accuracy of the classifier. Recall rate is mainly used to measure the recall rate of

TABLE 8. Simulation accuracy and loss function of extended samples.

Sample coincidence rate	Accuracy	Loss
90%	0.981729±0.005437	0.061013±0.026287
95%	0.981562±0.004084	0.091794±0.053285
98%	0.980925±0.003028	0.116923±0.050228

TABLE 9. Confusion matrix of classification results.

	Positive Prediction	Negative Prediction
Actually positive	TP	FN
Actually positive	FP	TN

TABLE 10. Text report of classification indicators.

	precision	recall	F1-score
Normal	0.99	1.00	0.99
Faulty	0.95	0.82	0.88
Avg/total	0.98	0.98	0.98

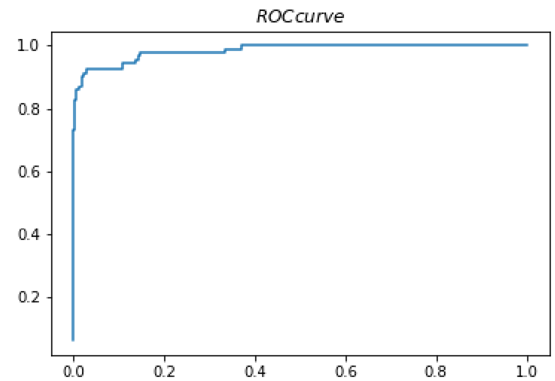


FIGURE 10. Receiver operating characteristic curve.

the classifier. F1-score is an evaluation index that considers both the accuracy and the recall rate. The larger the value is, the more effective the classifier becomes.

Another important indicator is the ROC curve, which depicts the trade-off between TP and FP. The classifier's performance varies from 0 to 1 with the threshold value, and the classifier's performance at the upper left corner is better than that at the lower right corner. AUC is the area enclosed by the ROC curve with the x-axis and $(1, 0)$ and $(1, 1)$. It is scalar data and makes it easier to compare classifier performance. The higher the AUC value is, the better the performance of the classifier becomes. It also has the advantage of the ROC curve: which can depict the overall performance of the classification algorithm, independent of the class prior distribution, the class misclassification cost, and the classification threshold. It can also depict the probability or sorting output characteristics of the classification algorithm [30].

Using the optimal parameter network structure and using the data with a sample coincidence rate of 80%, we can get the classification mentioned above evaluation index results as shown in Table 10.

It can be seen from Table 10 and Figure 10 that the precision and recall of the classifier approach to 1 for normal samples, and the result is satisfactory for fault samples. The value of AUC is 0.9837, which means that the performance of the classifier is excellent enough. It further verifies

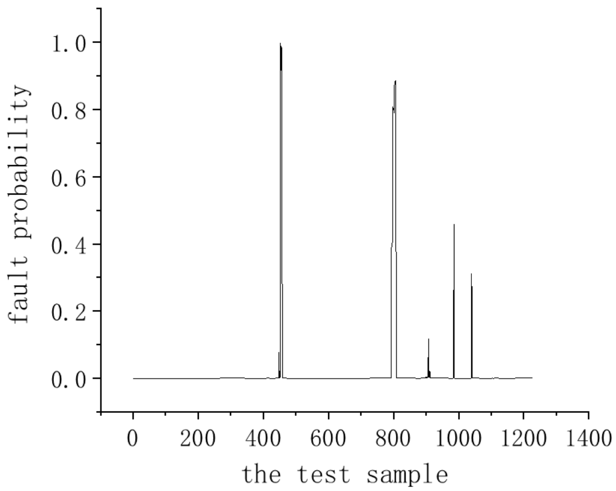


FIGURE 11. Sample health monitoring status.

that the optimal parameter network structure determined in Section V-B1 is correct.

From the health monitoring of the test set samples, we can see that the failure probability is obtained by using the softmax classifier. When the failure probability is greater than 0.5, it is determined as a failure actuator. On the contrary, it is healthy steering gear. At the same time, the probability also reflects the actuator's failure degree to complete the actuator's health monitoring.

VI. CONCLUSION

Due to the importance of the actuator in the aircraft, once the failure will cause serious consequences, it is necessary to monitor the health status of the actuator accurately. Due to the advantages of frequency domain and deep learning methods, this paper proposed a health status monitoring method based on deep neural network feature learning, which combines the two methods. Using the daily maintenance data of steering gear, the purpose of health monitoring can be achieved by optimizing the parameters. The specific contribution of this paper is summarized as follows:

(1) Because the sample data used is insufficient, and the fault degree needs to be judged, the premise of introducing a deep network is to preprocess the sample data sufficiently. In this stage, three-level wavelet packet decomposition and power spectral density are used to extract features, and then SVM and sliding window are used to expand the sample and adjust the label to complete the data processing.

(2) After obtaining the processed data, the automatic encoder is used to reduce the dimension of the data, which reduces the unnecessary calculation time for the subsequent use of the classifier, and the softmax classifier can be used to determine whether the actuator is faulty and its failure possibility. In the process of simulation, the number of layers, the number of nodes, and the initial learning rate in gradient descent of the encoder neural network are super parameters, which depend on human experience and many times of optimization.

(3) After several simulations and parameter modification times, the optimal structural parameters are obtained, and then the correctness of the selected parameters is further verified by using the results of expanded samples and classification evaluation indexes. Thus, it can provide some effective suggestions for the structure selection of neural networks and avoid the useless work caused by multiple parameter adjustments.

In the follow-up research, practical experiments will be considered to further verify the proposed method. In addition, this paper mainly uses deep neural networks to monitor the health status of servo and determine the fault degree.

REFERENCES

[1] K.-N. Areerak, T. Wu, S. V. Bozhko, G. M. Asher, and D. W. P. Thomas, "Aircraft power system stability study including effect of voltage control and actuators dynamic," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 4, pp. 2574–2589, Oct. 2011.

[2] E. Smart, D. Brown, and J. Denman, "A two-phase method of detecting abnormalities in aircraft flight data and ranking their impact on individual flights," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1253–1265, Sep. 2012.

[3] M. D. Dangut, Z. Skaf, and I. K. Jennions, "An integrated machine learning model for aircraft components rare failure prognostics with log-based dataset," *ISA Trans.*, vol. 113, pp. 127–139, Jul. 2021.

[4] R. Haque, H. Fairweather, and A. Olofinjana, "Failure analysis of boat rudder fitting fabricated from a type 304 stainless steel," *J. Failure Anal. Prevention*, vol. 18, no. 3, pp. 519–525, Feb. 2018.

[5] D. Wang, K.-L. Tsui, and Q. Miao, "Prognostics and health management: A review of vibration based bearing and gear health indicators," *IEEE Access*, vol. 6, pp. 665–676, 2018.

[6] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, Nov. 2018.

[7] W. Deng, S. Zhang, H. Zhao, and X. Yang, "A novel fault diagnosis method based on integrating empirical wavelet transform and fuzzy entropy for motor bearing," *IEEE Access*, vol. 6, pp. 35042–35056, 2018.

[8] J. Liu, Y. Hu, Y. Wang, B. Wu, J. Fan, and Z. Hu, "An integrated multi-sensor fusion-based deep feature learning approach for rotating machinery diagnosis," *Meas. Sci. Technol.*, vol. 29, no. 5, May 2018, Art. no. 055103.

[9] S. Tang, S. Yuan, and Y. Zhu, "Data preprocessing techniques in convolutional neural network based on fault diagnosis towards rotating machinery," *IEEE Access*, vol. 8, pp. 149487–149496, 2020.

[10] R. Huang, Y. Liao, S. Zhang, and W. Li, "Deep decoupling convolutional neural network for intelligent compound fault diagnosis," *IEEE Access*, vol. 7, pp. 1848–1858, 2018.

[11] O. Janssens, R. Van de Walle, M. Loccufier, and S. Van Hoecke, "Deep learning for infrared thermal image based machine health monitoring," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 151–159, Jul. 2017.

[12] Y. Yang, H. Zheng, Y. Li, M. Xu, and Y. Chen, "A fault diagnosis scheme for rotating machinery using hierarchical symbolic analysis and convolutional neural network," *ISA Trans.*, vol. 91, pp. 235–252, Aug. 2019.

[13] S. G. Kumbhar and P. E. Sudhagar, "An integrated approach of adaptive neuro-fuzzy inference system and dimension theory for diagnosis of rolling element bearing," *Measurement*, vol. 166, pp. 108266–108276, Dec. 2020.

[14] A. Ajagekar and F. You, "Quantum computing assisted deep learning for fault detection and diagnosis in industrial process systems," *Comput. Chem. Eng.*, vol. 143, pp. 107119–107135, Dec. 2020.

[15] W. Zhang, X. Li, and Q. Ding, "Deep residual learning-based fault diagnosis method for rotating machinery," *ISA Trans.*, vol. 95, pp. 295–305, Dec. 2019.

[16] B. Qi, Y. Wang, P. Zhang, C. Li, H. Wang, and J. Chen, "A novel self-decision fault diagnosis model based on state-oriented correction for power transformer," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 27, no. 6, pp. 1778–1786, Dec. 2020.

[17] C. Shen, Y. Qi, J. Wang, G. Cai, and Z. Zhu, "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder," *Eng. Appl. Artif. Intell.*, vol. 76, pp. 170–184, Nov. 2018.

[18] X. Li, W. Zhang, and Q. Ding, ''Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks,'' *IEEE Trans. Ind. Electron.*, vol. 66, no. 7, pp. 5525–5534, Jul. 2019.
[19] X. Zheng, Y. Y. Tang, and J. Zhou, ''A framework of adaptive multiscale wavelet decomposition for signals on undirected graphs,'' *IEEE Trans. Signal Process.*, vol. 67, no. 7, pp. 1696–1711, Apr. 2019.
[20] H. Liu, X. Mi, and Y. Li, ''Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network,'' *Energy Convers. Manage.*, vol. 166, pp. 120–131, Jun. 2018.
[21] S. Cui and E. Zheng, ''Generation of non-Gaussian stationary random excitations with specified cross-power spectral densities using zero memory non-linear transformation for fatigue test purposes,'' *Mech. Syst. Signal Process.*, vol. 153, pp. 107535–107557, May 2021.
[22] W. Yu and C. Zhao, ''Robust monitoring and fault isolation of nonlinear industrial processes using denoising autoencoder and elastic net,'' *IEEE Trans. Control Syst. Technol.*, vol. 28, no. 3, pp. 1083–1091, May 2020.
[23] J. Zabalza, J. Ren, J. Zheng, H. Zhao, C. Qing, Z. Yang, P. Du, and S. Marshall, ''Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging,'' *Neurocomputing*, vol. 185, no. 1, pp. 1–10, Apr. 2016.
[24] M. Ganger, E. Duryea, and W. Hu, ''Double Sarsa and double expected Sarsa with shallow and deep learning,'' *J. Data Anal. Inf. Process.*, vol. 4, no. 4, pp. 159–176, 2016.
[25] W. Fu, K. Shao, J. Tan, and K. Wang, ''Fault diagnosis for rolling bearings based on composite multiscale fine-sorted dispersion entropy and SVM with hybrid mutation SCA-HHO algorithm optimization,'' *IEEE Access*, vol. 8, pp. 13086–13104, 2020.
[26] Y. Li, X. Wang, S. Si, and S. Huang, ''Entropy based fault classification using the case western reserve University data: A benchmark study,'' *IEEE Trans. Rel.*, vol. 69, no. 2, pp. 754–767, Jun. 2020.
[27] Z.-X. Hu, Y. Wang, M.-F. Ge, and J. Liu, ''Data-driven fault diagnosis method based on compressed sensing and improved multiscale network,'' *IEEE Trans. Ind. Electron.*, vol. 67, no. 4, pp. 3216–3225, Apr. 2020.
[28] Y. Wang and S. Chen, ''A survey of evaluation and design for AUC based classifier,'' *Pattern Recognit. Artif. Intell.*, vol. 24, no. 1, pp. 64–71, Feb. 2011.
[29] E. Grøvik, D. Yi, I. Michael, E. Tong, D. Rubin, and G. Zaharchuk, ''Deep learning enables automatic detection and segmentation of brain metastases on multisequence MRI,'' *J. Magn. Reson. Imag.*, vol. 51, no. 1, pp. 175–182, Jan. 2020.
[30] M. Buda, A. Maki, and M. A. Mazurowski, ''A systematic study of the class imbalance problem in convolutional neural networks,'' *Neural Netw.*, vol. 106, pp. 249–259, Oct. 2018.

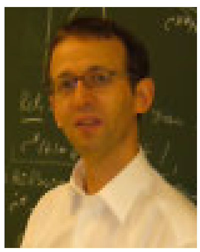
YAJING ZHOU was born in 1996. She received the B.E. degree from Chang'an University, Xian, China, in 2019. She is currently pursuing the master's degree with Nankai University, Tianjin, China. Her current research interests include fault diagnosis and data visualization.

YUEMIN ZHENG was born in 1996. She received the B.E. degree from Shijiazhuang Tiedao University, Shijiazhuang, China, in 2018. She is currently a Graduate Student at Nankai University, Tianjin, China. Her current research interests include active disturbance rejection control and reinforcement learning.

JIN TAO (Member, IEEE) received the B.Sc. degree in automation from the Qingdao University of Science and Technology, Qingdao, China, in 2008, the M.Sc. degree in control theory and control engineering from the Guangxi University of Science and Technology, Liuzhou, China, in 2011, and the Ph.D. degree in control science and engineering from Nankai University, Tianjin, China, in 2017. He is currently an Associate Professor with the College of Artificial Intelligence, Nankai University. He is also with the Department of Electrical Engineering and Automation, Aalto University. He has published more than 50 peer-reviewed articles in international journals and conferences. His research interests include intelligent control, evolutionary optimization, and multiagent systems.

MINGWEI SUN received the Ph.D. degree in control theory and control engineering from the Department of Computer and Systems Science, Nankai University, Tianjin, China, in 2000. From 2000 to 2008, he was a Flight Control Engineer at the Beijing Electro-Mechanical Engineering Research Institute, Beijing, China. Since 2009, he has been with Nankai University, where he is currently a Professor. His research interests include flight control, guidance, model predictive control, active disturbance rejection control, and nonlinear optimization.

QINGLIN SUN received the B.E. and M.E. degrees in control theory and control engineering from Tianjin University, Tianjin, China, in 1985 and 1990, respectively, and the Ph.D. degree in control science and engineering from Nankai University, Tianjin, in 2003. He is currently a Professor with the Intelligence Predictive Adaptive Control Laboratory, Nankai University, and the Associate Dean of the College of Artificial Intelligence. His research interests include self-adaptive control, modeling and control of exible spacecraft, and embedded control systems.



MATTHIAS DEHMER received the B.Sc. degree in mathematics from the University of Siegen, Germany, in 1997, and the Ph.D. degree in computer science from the Darmstadt University of Technology, Germany, in 2005. He is currently a Professor with the Department of Computer Science, Swiss Distance University of Applied Sciences. He has authored and coauthored more than 200 journal articles. His H-index equals 28 and his i10-index equals 78. His research interests include data science, cybersecurity, disaster management, complex networks, risk analysis, information systems, machine learning, information theory, bioinformatics, visual analytic, and computational statistics.

ZENGQIANG CHEN (Member, IEEE) was born in 1964. He received the B.S., M.E., and Ph.D. degrees from Nankai University, in 1987, 1990, and 1997, respectively. He is currently a Professor of control theory and engineering at Nankai University and the Deputy Director of the Institute of Robotics and Information Automation. His current research interests include intelligent predictive control, chaotic systems and complex dynamic networks, and multi-agent system control.