

Received November 27, 2019, accepted December 17, 2019, date of publication December 30, 2019, date of current version June 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2963193

Fault Diagnosis Framework of Rolling Bearing Using Adaptive Sparse Contrative Auto-Encoder With Optimized Unsupervised Extreme Learning Machine

XIAOLI ZHAO^{1,2}, (Student Member, IEEE), MINPING JIA¹, (Member, IEEE),
AND ZHENG LIU², (Senior Member, IEEE)

¹School of Mechanical Engineering, Southeast University, Nanjing 211189, China

²Faculty of Applied Science, School of Engineering, The University of British Columbia, Kelowna, BC V1V 1V7, Canada

Corresponding author: Minping Jia (mpjia@seu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51675098, in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province, China, under Grant SJKY19_0064, in part by the Youth Program of Jiangsu Natural Science Foundation under Grant BK20190911, and in part by the China Scholarship Council (CSC).

ABSTRACT Nowadays, the intelligent fault diagnosis based on deep learning have achieved remarkable results in the fields of the industrial equipment health monitoring and management. To implement the adaptive feature extraction and fault isolation for key components of rotating machinery (rolling bearings, etc.), two new algorithms, Adaptive Sparse Contrative Auto-encoder (ASCAE) algorithm and Optimized Unsupervised Extreme Learning Machine (OUSELM) classifier by Cuckoo Search Algorithm (CSA), can be firstly designed in this paper, respectively. Furthermore, a new rolling bearing fault diagnosis framework based on ASCAE combined with OUSELM is first of all proposed in this paper. Accordingly, this designed fault diagnosis framework can be divided into three main steps: i). Firstly, the vibration signals of rolling bearings can be collected and processed on the key components of rotating machinery, and then the collected vibration signals can be accordingly converted into frequency signals; ii). Secondly, the transformed spectral signals can be entered into the constructed ASCAE for feature learning to exploit the multi-layer sensitive features from the hidden raw data; iii). Thirdly, the extracted multi-layer sensitive features can be flowed into the trained OUSELM classifier for unsupervised fault state separation and diagnosis. More specifically, our designed fault diagnosis framework (ASCAE-OUSELM) can employ the homotopy regularization theory, sparse theory, intelligent optimization algorithm and other tools to optimize the parameters and improve the performance of the original Contrative Auto-encoder (CAE) algorithm and Unsupervised Extreme Learning Machine (USLEM) algorithm, respectively. At the same time, the proposed fault diagnosis framework can achieve effective sparse and sensitive feature information extraction in the feature extraction stage (ASCAE) to avoid over-fitting. In fault isolation stage, the issue of the supervised and low training efficiency caused by traditional deep learning model can be perfectly addressed by OUSELM. Eventually, the experimental data of rolling bearings validated the effectiveness of the proposed fault diagnosis framework and two deigned algorithms.

INDEX TERMS Rolling bearings, fault diagnosis framework, homotopy regularization, Adaptive Sparse Contrative Auto-encoder (ASCAE), Optimized Unsupervised extreme learning machine (OUSELM), Cuckoo Search Algorithm (CSA).

I. INTRODUCTION

At present, rotating machinery has been playing an indispensable and irreplaceable role in the fields of industrial

production and intelligent manufacturing [1]–[3]. At the same time, the health maintenance and management of mechanical system has attracted more and more attention of the enterprises and research staff [4], [5]. That is to say, an effective and appropriate mechanical equipment condition monitoring and fault diagnosis framework can not only ensure the safe

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

operation of rotating machinery, but also can reduce unnecessary faults, increase the service life of mechanical equipment, and improve the economic benefits of the entire industrial system [6], [7].

To implement the effective fault diagnosis of the core components of mechanical equipment such as rolling bearings, the fault diagnosis framework based on vibration signal analysis has gradually matured and is widely applied in many industrial system [8]–[10]. As we well-known that mechanical equipment occurs failure, the energy distribution of the vibration signals will be changed accordingly. These changes will be reflected in the vibration signals collected by the different sensors and measurements [10]–[12]. Nowadays, the conventional fault diagnosis methods based on vibration signal analysis are mainly employed to process and analyze the collected signals and extract the effective fault characteristic frequency from the original vibration signals, and it also can provide the research basis for the follow-up fault diagnosis and maintenance decision [12]–[15], [18]. More specifically, [16] analyzed the vibration signals from the actual multi-fault wind turbine gearbox with catastrophic failure, and then a multi-scale envelope spectrum map (MuSEnS) of complex wavelet transform was designed for simultaneous decomposition and solution. Liu *et al.* [17] designed a new fault diagnosis method combining Least Squares Support Vector Machine (LS-SVM) with empirical mode decomposition (EMD) to improve the performance of the conventional EMD.

However, most of the above-mentioned fault diagnosis methods based on signal analysis are difficult to quantify the fault diagnosis results, and the vibration signals collected by the sensing and measuring devices are mostly unmarked and unknown. Therefore, fault diagnosis based on vibration signal analysis is limited in the process of effectively identifying and analyzing fault diagnosis results. At present, data-driven intelligent diagnosis methods have emerged in the field of fault diagnosis [19], [20]. That is to say, the data-driven intelligent fault diagnosis method is mostly based on the machine learning, i.e. the “feature extraction + intelligent classifier” based-intelligent fault diagnosis mode [20], [21]. This diagnosis mode can effectively avoid the defects that the vibration signal based-fault diagnosis is inaccurate and relying on a large amount of expert knowledge, which can implement the intelligence and automation degree of the traditional fault diagnosis model. At present, the related fault diagnosis mode based on “feature extraction + intelligent classifier” is widely applied in the field of fault diagnosis [19], [22]. The main task of fault diagnosis is to extract more fault characteristic information from the measured signals to make its diagnosis more accurate and reliable. To obtain more inherent global and local fault information from mechanical running data, Zhao *et al.* [19] firstly designed a new fault diagnosis method of rolling bearings based on feature reduction, namely, a new manifold learning using Global Local Fisher analysis (GLMFA). Su *et al.* [22] proposed a fault diagnosis method based on incremental enhanced supervised

local linear embedding (I-ESLLE) and adaptive nearest neighbor classifier (ANNC). However, the above-mentioned data-driven intelligent fault diagnosis methods still have the following two shortcomings: i) the above feature extraction relies on expert knowledge and artificial experience, it is impossible to mine the effective fault features from the original vibration signals; ii) the above-mentioned data-driven intelligent fault diagnosis methods are mostly shallow learning.

As a newcomer in the field of intelligent fault diagnosis, deep learning has received great attention in recent years [2], [21], [24], [27]. The purpose of deep learning is to autonomously mine valuable information hidden in massive measurement and monitoring data through multiple layers of repeated nested feature transformation and feature learning, and it is used to establish the accurate mapping relationship between the device and its operating state through data and models. That is to say, deep learning is the whole process of the unified feature learning and classification, which can realize the transformation of the traditional fault diagnosis mode of “original signal + deep learning” [2], [25], [26]. Therefore, the application of deep learning in intelligent fault diagnosis has an important positive significance for ensuring the safe operation of industrial equipment.

Since Tamilselvan and Wang [25] firstly applied the Deep Belief Network (DBN) into the health assessment of aircraft engines in 2013, this direction of intelligent fault diagnosis based on deep learning has attracted more and more attention. In [26], a new multiscale convolutional neural network (MSCNN) was designed to perform multi-scale feature extraction and classification simultaneously. In recent years, many scholars at home and abroad have developed the modified models such as Sparse Autoencoder (SAE) [28], Denoising Autoencoder (DAE) [29], and Contrastive AE (CAE) [30] on the basis of standard AE. In [31], a multi-step progressive fault diagnosis method based on energy entropy (EE) theory and hybrid integrated automatic encoder (HEAE) was firstly proposed. At the same time, Contrastive Auto-encoder (CAE) [30], [32]–[33] is an extension of the basic AE, which can add a more powerful feature representation by adding shrink penalty terms. In other words, CAE explicitly encourages the robustness of feature expression from the raw data by introducing the F-norm of Jacobian matrix as a constraint in AE. In summary, CAE algorithm primarily suppresses the perturbation of the input data in all directions. It can be analyzed that CAE has stronger data compression capability than traditional AE variant model, and its stability can be more stable. In [30], a stack-based CAE was applied to automatic robust feature extraction and fault diagnosis method for rotating machinery. However, the current CAE based-fault diagnosis is mainly composed of some simple and identical basic models. The above-mentioned fault diagnosis methods based on CAE still have some defects: i).the sparse performance of the general CAE model should be improved, and the sparsity of the data can be conducive to remove the redundancy of the raw data

and improve its generalization performance; ii). it is difficult to adaptively determine parameters such as the penalty coefficient of deep learning model. iii). usually, the supervised Soft-max can be employed in classification stage of CAE, but the classification category label of the measuring samples is scarce in the actual fault diagnosis, which makes the diagnosis more difficult in the absence of the sample category labels.

To solve the above-mentioned three problems, we can analyze and discuss from the following three aspects: Firstly, too few training samples will lead to over-fitting problems [34], [35]. More importantly, the sparse deep learning model can effectively solve the above problem by reducing the dimensional dimension of the variable. Sparse group [34], [35] has good interpretability, facilitating data visualization, reducing computation and transferring storage. Therefore, the designed sparse CAE (SCAE) algorithm with sparse group can learn the sparse representation of the input signals. Secondly, the sparse coefficient and contractive coefficient have an influence on the optimization result of the designed model simultaneously. And the number of parameters can be reduced by homotopy regularization technique [36]. Thirdly, one of the problems faced by data-driven intelligent fault diagnosis is how to achieve unsupervised diagnosis of fault diagnosis in the case of scarce sample label information. And the traditional classification stage is mostly based on supervised Soft-max, it is not conducive to the actual industrial intelligent fault diagnosis. Fortunately, unsupervised learning provides an effective solution for intelligent clustering and partitioning of deep learning [8].

Recently, Huang *et al.* [37], [38] first designed an extreme learning machine (ELM) for training single hidden layer feedforward neural network (SLFN). Compared to most existing methods, ELM only updates the output weight between the hidden layer and the output layer, and the parameters (i.e., the input weight and bias of the hidden layer) are randomly generated. ELM has been widely used in face recognition, motion recognition, EEG signal processing, image processing, and fault diagnosis [37], [38], owing to its effectiveness and rapid learning process compared with gradient-based optimization. In many real-world applications, the labeled data is often expensive. On the basis of the multivariate regularization, Huang *et al.* [39] proposed two ELM variants, semi-supervised ELM and unsupervised ELM (USELM). Compared with the existing algorithms, USELM can provide effective support and expansion for unsupervised learning of deep learning. According to [39], the relevant parameters affecting the dimension reduction effect of USELM algorithm are divided into two parts, one is related to the Laplacian matrix's penalty coefficient (λ), and another is related to the selection of the neighboring number (NN) in the USELM embedding process. To implement USELM's parameter adaptive optimization is very meaningful for unsupervised intelligent diagnosis. In 2013, Yang *et al.* [40] proposed a new type of bionic intelligent search algorithm with global convergence (Cuckoo search

algorithm, CSA). Compared with traditional evolutionary optimization algorithms, such as genetic algorithm (GA), particle swarm optimization (PSO), etc., CSA can combine cuckoo nest parasitism and Levy flight mode with fewer control variables and simpler implementation process and excellent parameter optimization performance. Many researchers had employed CSA algorithms to solve complex parameter optimization problems in different fields. Yan *et al.* [11] firstly designed CSA-VMD (Variational mode decomposition) and optimal scale morphological slice bispectrum to enhancing outer race fault detection of rolling element bearings.

In summary, a new fault diagnosis framework of rolling bearing based on Adaptive Sparse Contractive Auto-encoder (ASCAE) combined with optimized unsupervised extreme learning machine (USELM) by CSA is firstly designed in this study. The proposed framework first of all inputs the extracted vibration signal into the constructed ASCAE model for feature extraction, and the extracted sensitive features can be entered into CSA-optimized USELM (OUSELM) unsupervised classifier for fault isolation. The core of this new fault diagnosis framework can be displayed as below: i) In the feature extraction stage, the traditional CAE model is optimized by sparse learning theory and homotopy regularization. A more powerful adaptive multi-layer feature extractor (so-called ASCAEs) is firstly designed in this paper; ii) In the fault separation stage, the parameters of USELM classifier are optimized by CSA intelligent optimization algorithm, so that the designed OUSELM algorithm can realize adaptive unsupervised fault separation and intelligent recognition.

Furthermore, the rest of this paper can be organized as follows: in the second section, the existing theories of CAE, CSA and USELM are briefly reviewed, respectively; afterwards, the new feature extractor (ASCAE) algorithm and fault classifier (OUSELM) are designed, and an intelligent rolling bearing fault diagnosis framework based on two designed algorithms can be presented in the third section. And the experimental results and analysis are described in the fourth section. Finally, some research conclusions are given in the fifth section.

II. BACKGROUND THEORIES

In this section, the basic theory and implementation process of the related CAE, CSA, USELM and other algorithms in this paper are mainly introduced in this section, respectively. It is able to make sufficient preparations for the algorithm design and fault diagnosis framework proposed.

A. THE BASIC PRINCIPLE OF CONTRACTIVE AUTO-ENCODER (CAE) ALGORITHM

Contractive Auto-encoder (CAE) is a variant of Auto-encoder (AE). A constraint can be added into the loss function of the original Auto-encoder model. In general, the objective function of AE with weight decay can be

generally described as

$$J = \left\{ \frac{1}{N} \sum_{i=1}^N \|z_i - x_i\| + \lambda \sum_{i,j=1} W_{ij}^2 \right\} \quad (1)$$

Compared with traditional AE, a Jacobian matrix is added in the objective function of CAE to extract the compression characteristics of the data. Therefore, CAE-based objective function can be defined as

$$J_{CAE} = \frac{1}{N} \sum_{i=1}^N \|z_i - x_i\| + \lambda \|J_f(x)\|_F^2 \quad (2)$$

the $\|J_f(x)\|_F^2$ value is the Jacobian matrix of the hidden layer output value. The λ value of the weight loss function between the weight loss function and the constraint term is determined experimentally. It is the square of the \mathbf{F} -norm of the Jacobian matrix $\|J_f(x)\|_F^2$, that is, the square of each element in the Jacobian matrix is summed. Accordingly, the specific Jacobian matrix can be defined as follows.

$$\|J_f(x)\|_F^2 = \sum_{i,j} \left(\frac{\partial h_j(x)}{\partial x_i} \right)^2 \quad (3)$$

the sum of the squares of the \mathbf{F} norm for Jacobian matrix can be rewritten as a more specific mathematical expression.

$$\|J_f(x)\|_F^2 = \sum_{i,j} \left(\frac{\partial h_j(x)}{\partial x_i} \right)^2 = \sum_{i=1}^{d_h} [h_i(1 - h_i)]^2 \sum_{j=1}^{d_x} W_{ij}^2 \quad (4)$$

where h_i is the output of the hidden layer, and W_{ij} is the connection weight of the input layer and the hidden layer. The Jacobian matrix contains the information of the data in all directions, which makes the extracted features invariant to the disturbance of the inputting data to a certain extent. However, the traditional CAE cannot extract the sparsity characteristics of data, and its computational complexity is high.

B. THE BASIC PRINCIPLE OF CUCKOO SEARCH ALGORITHM (CSA)

Yang *et al.* [40] first proposed a Cuckoo search algorithm (CSA) with a global convergence in 2013. Compared with traditional evolutionary optimization algorithms, CSA can combine cuckoo nest parasitism and Levy flight mode with fewer control variables and simpler implementation process. And it is show excellent parameter optimization performance. In nature, cuckoos increase reproduction rates in the form of parasitic brooding chicks and look for nests in a random manner. Based on this behavior, the CSA is proposed to find the optimal solution by simulating the behavior of the cuckoo to find the parasitic bird eggs in the nest and combining the Levy flight characteristics of the birds. In this algorithm, the bird's egg in the nest is regarded as a solution, and the cuckoo's egg is regarded as a new solution. The goal is to make the cuckoo's egg be replaced by the host's egg, that is, the cuckoo is good. To summarize the cuckoo

search algorithm, three idealized conditions can be firstly set as below: 1) Each cuckoo produces only one egg at a time, and chooses the nest to place the eggs in a random manner; 2) The best bird nest (solution) should be retained to the next generation; 3) The number of available nests is a fixed value. The probability that the owner of the nest can find the foreign bird is pa , where $pa \in [0, 1]$. In this case, the owner of the nest can choose to abandon the bird's egg, or directly abandon the bird's nest and build a brand-new nest in a new location. It can be approximated that the pa value of n nests is replaced, and accordingly a new random solution is found at the new position.

The algorithm can enhance global search capabilities through Levy flight:

1) LEVY FLIGHT

Levy flight is a random walk pattern with a step-and-length distribution that increases the diversity of the population and avoids falling into local optimum. Generating a random number with Levy flight should consist of two steps: i). the selection of the random direction and the generation of the step size should be obeyed the Levy distribution. ii). The generation of directions obeys uniform distribution. The most effective and direct method for generating step size is to implement symmetric Levy distribution with Mantegna algorithm. In the Mantegna algorithm, the step size s can be calculated using two variables U and V obeying the Gaussian distribution, namely:

$$s = \frac{U}{|V|^{\frac{1}{\lambda}}}, U \sim N(0, \sigma^2), V \sim N(0, 1) \quad (5)$$

where the calculation of variance can be expressed as follows:

$$\sigma^2 = \left[\frac{\Gamma(1 + \lambda)}{\lambda \Gamma((1 + \lambda)/2)} \cdot \frac{\sin(\pi \lambda / 2)}{2^{(\lambda - 1)/2}} \right]^{1/\lambda} \quad (6)$$

2) CUCKOO SEARCH

The cuckoo search algorithm is a balanced combination of local random walks controlled by the switch parameter pa and globally explored random walks. Accordingly, the local random walks can be expressed as

$$x_i^{t+1} = x_i^t + \alpha s \otimes H(pa - \varepsilon) \otimes (x_j^t - x_k^t) \quad (7)$$

where \otimes —point multiplication; H (*)—unit step function; ε —uniformly distributed random number; α —step size scaling factor; pa —switching parameter; s —step size; x_j^t, x_k^t —random two different solutions. And global exploration of random walks using Levy flight is displayed as below.

$$x_i^{t+1} = x_i^t + \alpha L(s, \lambda) \quad (8)$$

$$L(s, \lambda) = \frac{\lambda \Gamma(\lambda) \sin(\pi \lambda / 2)}{\pi} \cdot \frac{1}{s^{1+\lambda}} \quad (9)$$

where: α —step size scaling factor; s —step size; $\Gamma(\lambda)$ — is a constant for a given λ . Through local random walks and

global exploration random walks, it is possible to effectively avoid falling into local optimum.

Step 1: The initial population of the optimization algorithm consists of nests of n random positions. Find the current global optimal gt_0 position by calculating the objective function value of each nest.

Step 2: Generate a new solution $x_i^t + 1$ by

$$x_i^{t+1} = x_i^t + \alpha \times L(\lambda), \quad i = 1, 2, 3, \dots, n \quad (10)$$

Step 3: Select the random number r from the uniform distribution $[0, 1]$, compare r with the host discovery probability pa , and update x_{i+1}^t to obtain a new global optimal solution gt^* (c, g);

Step 4: If stop requirement is met, then gt^* (c, g) is the best global solution found so far. Otherwise, return to Step 2.

In summary, the cuckoo search algorithm has the characteristics of elitism similar to the genetic algorithm, that is, the optimal solution is retained to the next generation to prevent it from being evicted and the population. In addition, CSA has few parameters, and it is simple and easy to performed. It does not need to re-match a large number of parameters when dealing with complex optimization problems. Compared with other optimization methods, mCSA is more effective than other heuristic algorithms. At present, the cuckoo algorithm is rarely used in the field of fault diagnosis.

C. THE BASIC PRINCIPLE OF UNSUPERVISED EXTREME LEARNING MACHINE (USELM) ALGORITHM

From the perspective of neural network, USELM can be regarded as an unsupervised learning method of single hidden layer neural network. On the basis of ELM, the Laplacian eigenmaps (LE) method is combined to change the calculation method of output weight so that it is no longer related to data samples. Label correlation is related to the adjacency relationship of the data samples, thereby unsupervised learning of the input data can be achieved by using USELM. In the setting of USELM, the training data $X = \{x_i\}^N, i = 1$ is unmarked (N is the number of training modes). Our goal is to find the infrastructure of the raw data. If there is no data labeled structure, its objective function can be defined as below.

$$\min_{\beta \in R^{nh \times n_0}} \|\beta\|^2 + lamTr(\beta^T H^T LH \beta) \quad (11)$$

here λ is the adjustment parameter of the penalty item, which β is the output weight of the hidden layer and the output layer. $H = [\eta(x_1)^T, \eta(x_2)^T, \dots, \eta(x_N)^T] \in R^{N \times nh}$, $\eta(x_N)$, the output vector $\eta(x_N)$ represents the hidden layer relative to x . And $L \in R^{(l+u) \times (l+u)}$ is a Laplacian constructed with both labeled and unlabeled data. Denoted that the above $\beta = 0$ always reaches the minimum value. Accordingly, the additional constraints must be introduced to avoid degenerate solutions. Specifically, the manifold regularization term is added into the objective function of USELM. The Equation of

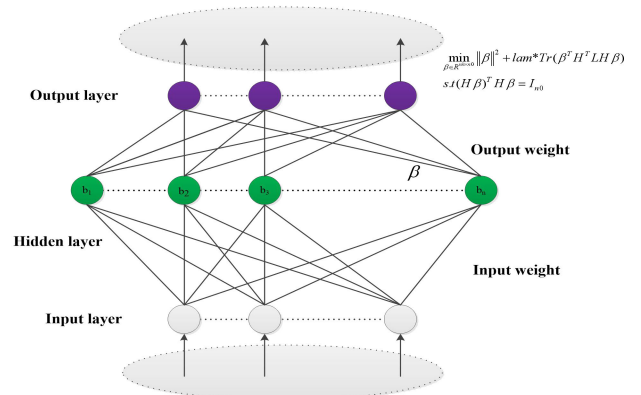


FIGURE 1. Schematic diagram of USELM.

the unsupervised ELM can be described as follows:

$$\min_{\beta \in R^{nh \times n_0}} \|\beta\|^2 + lamTr(\beta^T H^T LH \beta) \quad (12)$$

$$s.t. (H\beta)^T H \beta = I_{n_0}$$

the solution of the above Equations can be obtained by solving the above solution with constraints. The column of the matrix β is the eigenvector corresponding to the first minimum eigenvalue problem

$$(I_{nh} + lamH^T LH)v = \gamma H^T H v \quad (13)$$

where $\gamma_1, \gamma_2, \dots, \gamma_{n_0+1}$ is $N_0 + 1$ minimum eigenvalues, $v_1, v_2, \dots, v_{n_0+1}$ is its corresponding feature vector, the equation (12) can rewritten as

$$\min_{\beta \in R^{nh \times n_0}, \beta \beta^T = I_{n_0}} Tr(\beta^T A \beta) \quad (14)$$

where $A = I_{nh} + \lambda H^T LH$ and $B = H^T H$, the output weights are displayed as

$$\beta^* = [\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_{n_0+1}] \quad (15)$$

where $\tilde{v} = v_i / \|Hv_i\|, i = 2, \dots, n_0 + 1$, it is a normalized feature quantity.

If the number of marked data is less than the number of hidden layer neurons, then Eq. (13) cannot be determined. In this case, Eq. (13) can be replaced with

$$(I_u + lamHLH^T)u = \gamma HH^T u \quad (16)$$

therefore, we can assume that u is the minimum feature of the corresponding generalized eigenvector, then the final solution can be described as below.

$$\beta^* = H^T [\tilde{u}_2, \tilde{u}_3, \dots, \tilde{u}_{n_0+1}] \quad (17)$$

where $\tilde{u}_i = u_i / \|HH^T u_i\|, i = 2, \dots, n_0 + 1$ is a normalized feature vector.

It can be known from the above equations that the relevant parameters affecting the dimensionality reduction effect of the USELM algorithm are divided into two parts, one is related to the Laplacian matrix's penalty coefficient (lam),

and the other is related to the selection of the neighboring number (NN) in the USELM embedded process. If the number is too small, the local characteristic range retained by the matrix L is too small, and the data points have similar characteristics. On the contrary, if the number is too large, the local characteristic range of the reservation is too large, and the data points that are far apart are close in the low-dimensional space.

III. DESIGNED ALGORITHMS AND FAULT DIAGNOSIS FRAMEWORK FOR ROLLING BEARINGS

In recent years, the structural design of deep learning model and its application to fault diagnosis have been widely recognized by more and more scholars. CAE is one of the extensions of the classical AE. The essence is that the introduction of the Jacobian matrix is equivalent to doing a similar dimensioning operation for the inputting data, and then the high value in the original input space can be obtained after feature encoder. The Jacobian matrix contains the information of the data in all directions, which makes the extracted features invariant to the disturbance of the inputting data to a certain extent. However, the traditional CAE model still has the following shortcomings:

- The generalization performance of the designed model should be enhanced, and the sparseness of data is beneficial to remove the redundancy of data.
- Parameters of designed mode such as sparse coefficient and contrative coefficient of the designed deep learning algorithm are difficult to determined adaptively, resulting in poor stability and automation.
- The general CAE model is supervised at the classification stage, but the sample label of the data in the actual fault diagnosis process is scarce. Therefore, it is difficult to diagnose the faults in the case of less sample category labels.

To solve the above-mentioned three types of problem: firstly, the sparse group combining L1 norm and L2 norm can be added into CAE model for improving the sparse feature performance of the model, and homotopy regularization technique can be applied to improve the parameter adaptive capability. Finally, the optimized USELM classifier by CSA is designed to the classification layer for fault isolation and diagnosis.

A. DESIGNED ADAPTIVE SPARSE CONTRATIVE AUTO-ENCODER (ASCAE) ALGORITHM FOR FEATURE EXTRACTION

This paper we proposes a new feature extraction(ASCAE) algorithm with the optimized USELM, theories of sparse learning, homotopy regularization, unsupervised learning, and intelligent optimization of CSA are efficiently utilized to enhanced feature extraction and unsupervised classifier fault diagnosis performance, respectively. The two designed algorithms (ASCAE and OUSELM) and the rolling bearing

fault diagnosis framework using ASCAE with OUSELM are described in detail as follows:

1) DESIGNED SPARSE GROUP FOR THE NEW OBJECTIVE FUNCTION

Inspired by sparse learning such as compressed sensing, the sparse constraint based on L1 norm has received unprecedented attention [34], [35]. As a basic regularization condition, it has rapidly become a frontier research topic. The solution of the sparse regularization method based on the L1 norm is usually sparse, and the solution of the L2 norm regularization method by controlling the energy of the variable to be sought is often non-sparse. The emergence of sparse theory provides an effective solution for the extraction of dilution features, which can improve the generalization performance of the model. In practical applications, we find that the traditional CAE algorithm is not robust to signal noise, and the complexity of fault feature extraction is high. Most fortunately, in the modeling of high-dimensional problems, there is often a strong statistical correlation between features. It is not feasible to directly learn the statistical correlation between all hidden units. Therefore, the sparse group method to CAE and proposed Sparse CAE (SCAE). That is to say, if the advantages of the L1 and L2 norm are combined, the sparse characteristics of the data and the generalization performance of the deep neural network can be sufficiently adjusted. To comprehensively utilize the above two types of regularization techniques, a combination of sparse group norms ($L1/L2$) that combines the L1 and L2 norms to improve the generalization ability of the general CAE algorithm can be designed.

By constraining the weight matrix w , a sufficiently smooth projection is obtained in the low dimensional space to maintain the structural information of the data original space. Commonly used sparse group methods include the L1 norm, the L2 norm, and the norm combination. That is to say, the L1 norm is expressed as the sum of the absolute values of each element of the w vector, and it can be written as follows.

$$l_1 = \|w\|_1 = \sum_{i=1}^n |w_i| \quad (18)$$

The solution of the L1 norm is usually sparse and tends to select a smaller number of feature vectors. The L2 norm is a 1/2 power of the sum of the squares of each element of the w vector. Therefore, the L2 norm is also so-called the Euclidean norm (Euclidean distance).

$$l_2 = \|W\|_2 = \sqrt{\sum_{i=1}^n w_i^2} \quad (19)$$

The smaller the L2 norm, the smaller each element of w can be close to 0, but unlike the L1 norm, it does not let it equal 0 but is close to zero. According to [34], [35], [41], the

L1 / L2 sparse group norm can be written as follows:

$$l_1 / l_2 = \frac{\|w\|_1}{\|w\|_2} = \frac{\sum_{i=1}^n |w_i|}{(\sum_{i=1}^n w_i^2)^{\frac{1}{2}}} \quad (20)$$

To the best of our knowledge, the regularization term of the above combination sparse group norm can make the learned structural information sparser and more generalized. Owing to it can make irrelevant weights zero, from the perspective of Bayesian learning, the regularization term of the combined sparse group norm has such a function that if the neural unit of the neural network is related to the unit, it is prominent, otherwise it is zero. Therefore, the combination sparse group norm is punished with prior knowledge in this case, which can make the model's weight synthesis ability stronger. In other words, this constraint makes the extracted feature classes more compact and the class distance is more dispersed, which makes the general performance of the extracted features better.

In summary, the regularization term of the above combination sparse group norm can be embedded in objective function of the original CAE. Therefore, the new objective function with a combined sparse group norm is defined as:

$$J_{ASCAE} = \frac{1}{N} \sum_{i=1}^N \|z_i - x_i\| + \vartheta \|J_f(x)\|_F^2 + \beta \frac{\|w\|_1}{\|w\|_2} \quad (21)$$

2) DESIGNED HOMOTOPY REGULARIZATION FOR THE NEW OBJECTIVE FUNCTION

This section is based on the above-mentioned designed Sparse Contrative Auto-encoder model. It is difficult to optimize the adjustment parameters of two regularization terms (Sparse coefficient ϑ and Contrative coefficient β) in its objective function. In view of the above-mentioned problems, the idea of homotopy regularization in the inverse problem of mathematical equations can be introduced into the model to improve the range of the above two regularization parameters from the infinite interval $(0, +\infty)$ to a finite interval $(0, 1)$ is used to establish a new Adaptive Sparse Contrative Auto-encoder (ASCAE) algorithm for feature extraction. Compared with the original model, the designed ASCAE model with homotopy regularization parameters are easier to be optimized. In the model of the design, the homotopy regularization idea in mathematics is applied, but the range of regularization parameters in the model is an infinite interval $(0, +\infty)$, which is difficult to optimize. The homotopy regularization method is a new regularization method in recent years. It has been widely applied in the inverse problem of mathematical equations and has achieved very ideal effects. The regularization parameters of the model are easy to optimize, and the parameter adaptive ability is stronger.

The so-called homotopy, in a broad sense, is to introduce a regularization parameter λ for any two functions $F(x)$ and $G(x)$, and a completely new function: $H(x) = (1 - \lambda)F(x) + \lambda G(x)$ can be obtained, where the parameter λ has

a value range of $(0, 1)$. When $\lambda = 0$, $H(x) = F(x)$; when $\lambda = 1$, $H(x) = G(x)$. When λ changes continuously from 0 to 1, the function $H(x)$ changes continuously from $F(x)$ to $G(x)$, so the parameter λ associates $F(x)$ with $G(x)$. In the above-mentioned formulas, $F(x)$ and $G(x)$ are called homotops. In mathematics, the homotopy methods are employed to prove the existence and uniqueness of the solution of the equation. The homotopy regularization idea is often applied to solve the inverse problem of mathematical equations. The advantage is that the regularization parameters are transformed from the original infinite interval $(0, +\infty)$ into a finite interval $(0, 1)$, so the experimental design becomes very easy, the main thing is that the regularization parameters are easy to optimize and improve the recognition rate of the model. Based on the advantages of homotopy regularization in the inverse problem of mathematical equations, this idea can be introduced into the above-mentioned Sparse Contrative Auto-encoder (SCAE) algorithm. More specifically, the objective function of the improved Adaptive Sparse Contrative Auto-encoder (ASCAE) can be defined as

$$J_{CAE} = \frac{1}{N} \sum_{i=1}^N \|z_i - x_i\| + \lambda \|J_f(x)\|_F^2 + (1 - \lambda) \frac{\|w\|_1}{\|w\|_2} \quad (22)$$

The biggest change for this new algorithm is that the range of homotopy regularization parameters λ in it changes from infinite to finite $(0,1)$. In this way, in the process of iteration, the optimization problem of two parameters will be transformed into a parameter problem. Since the value range of λ is small and fixed between $(0, 1)$, the objective function is minimized, and the actual operation is performed. And each iterative process, due to the narrowing of the parameter value range, the actual operation is relatively easy.

Adaptive Sparse Contrative Auto-encoder (ASCAE) algorithm can effectively extract the low-dimensional features of the original data in an unsupervised mode, and it has powerful feature expression capabilities. Similar to DBN and DAE, the construction of the DASCAE model also consists of two phases: an unsupervised layer-by-layer pre-training phase and a supervised global fine-tuning phase.

In summary, ASCAE algorithm primarily suppresses the perturbation of input data in all directions. In general, a deep ASCAE is a deep neural network with composed of the multi-layer ASCAEs. The previous hidden layer of CAE is regarded as the input layer of the subsequent layer of ASCAE. All parameters of the multi-layer ASCAE are obtained by using the layer-by-layer greedy training method. The weight matrix of the layer and the offset vector of the multi-layer ASCAEs can be obtained. During the training phase, the stacked multilayer ASCAEs can be treated as a whole and fine-tuned by back-propagation. Through multiple forward propagation and backpropagation processes, the parameters between neurons can be optimized. When the error between the output result and the actual result meets the requirement or reaches the maximum number of iterations,

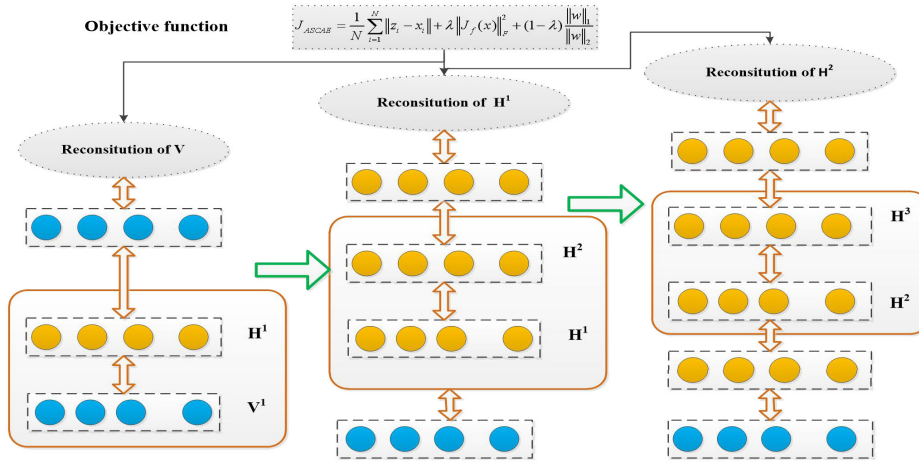


FIGURE 2. Structure of the multi-layer Adaptive Sparse Contrastive Auto-encoders (ASCAEs).

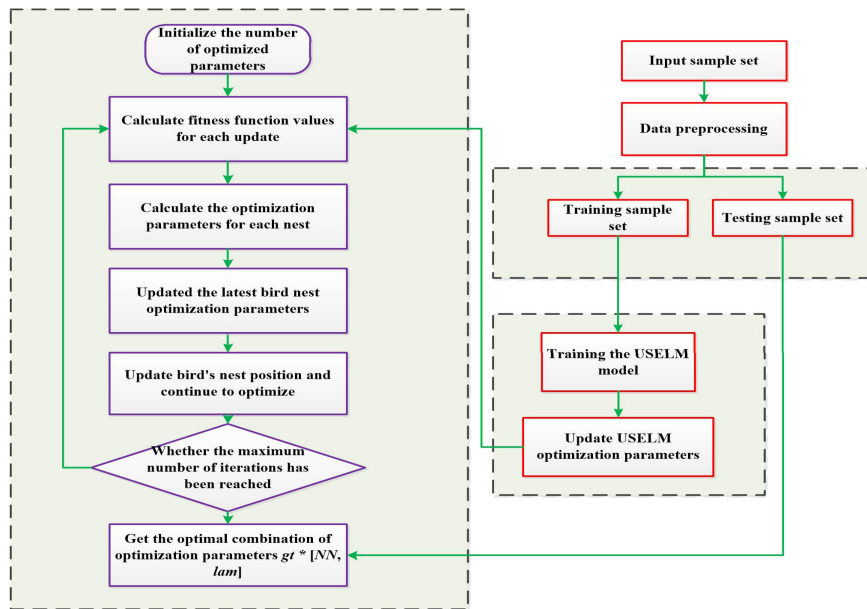


FIGURE 3. Structure of the intelligent optimization algorithm based on CSA-USELM.

this learning is stopped. Finally, the classification accuracy of the model is tested by testing samples.

B. OPTIMIZED USELM (OUSELM) FOR FAULT ISOLATION

In this subsection, USELM is optimized through CSA algorithm to train the nearest neighbor NN and Laplace penalty coefficient lam in USELM, and the training set to predict the error rate of USELM is employed as the fitness function (the lower the error rate, the corresponding NN and lam), to adaptively find an optimal NN and lam combination.

The specific steps for optimized USELM by CSA are described as follows:

- (1) Set the range of USELM parameters NN and lam ;
- (2) Set up the number of iterations of the CSA, the number of nests n , the probability of being discovered pa , the number of the parameters to be optimized;

- (3) For each nest, randomly initialize the values of NN and lam as the parameters of USELM, and the predicted error rate are regarded as the fitness value to find the optimal nest (i.e. the current optimal NN and lam).

$$f_{new} = 1 - \frac{fitness}{100} \quad (23)$$

where fitness—the accuracy of the USELM classification of parameters by NN and lam ; f_{new} —the fitness at this time (the purpose is to find the NN and lam under the minimum f_{new}).

- (4) Update the nest using Levy flight, and regard the updated values of NN and lam as parameters of USELM, and repeat step (3) to obtain a new set of nests;
- (5) Randomly eliminate some of the nests with the probability of pa , and update the nest according to Eq. (8), repeat step (3) to obtain a new set of nests, and obtain

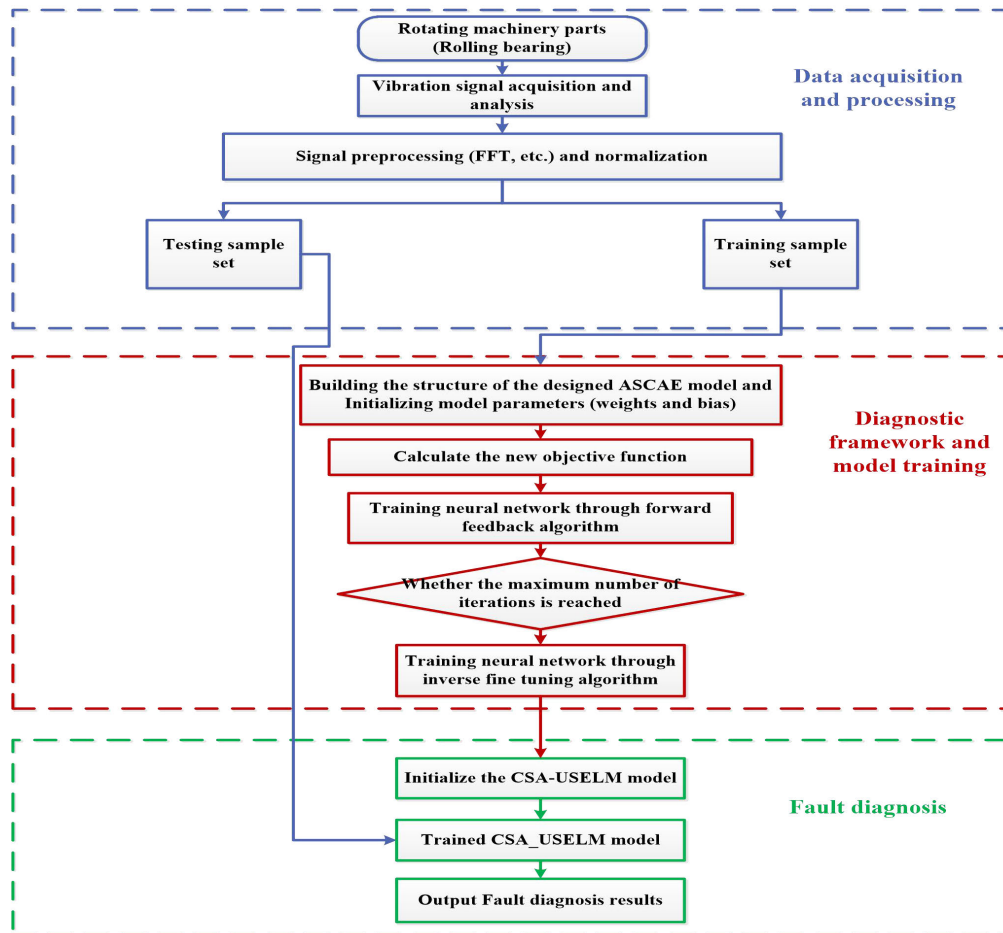


FIGURE 4. Fault diagnosis framework of rolling bearings based on ASCAE-OUSELM.

the new fitness value and the fitness value obtained in step (4). Compare and get the best nest;

- (6) Find the best nest from step (5). If the optimization objective function value meets the end condition, then output is the best nest and fitness value, otherwise return to step (4) to continue optimization.

Because the CSA algorithm is a combination of local optimal search and global optimal search, it is not easy to fall into the local optimal situation, and the classification accuracy will be higher than the traditional method.

C. THE PROCEDURE OF THE DESIGNED ROLLING BEARING FAULT DIAGNOSIS FRAMEWORK

Fault diagnosis can be divided into three-step: signal acquisition, feature extraction and pattern recognition. Pattern recognition is mainly divided into supervised classification and unsupervised clustering. To implement the adaptive feature extraction and fault isolation for key components of rolling machinery (rolling bearings, etc.), based on the designed new algorithms, Adaptive Sparse Contrative Auto-encoding and Unsupervised extreme learning machine (USELM)

optimization by Cuckoo Search Algorithm (CSA); furthermore, a new rolling bearing fault diagnosis framework based on Adaptive Sparse Contrative Auto-encoding combined with CSA optimized unsupervised extreme learning machine (OUSELM) is proposed in this paper. Accordingly, the designed fault diagnosis framework can be divided into three main steps: i). Firstly, the vibration signals can be collected and processed on the key parts of the rotating machine (rolling bearings, etc.), and the collected vibration signal can be converted into the frequency signal; ii). Secondly, the transformed spectral signal is entered into the constructed ASCAE for feature learning to exploit the multi-layer sensitive features from the hidden inside the data; iii). Thirdly, the extracted multi-layer sensitive features are input to the trained Optimized Unsupervised Extreme Learning Machine (OUSELM) classifier for unsupervised fault state separation and diagnosis. Among them, our proposed fault diagnosis framework (ASCAE-OUSELM) can employ the homotopy regularization theory, sparse theory, CSA intelligent optimization algorithm and other tools to optimize the parameters and improve the performance of the original Contrative Auto-encoder (CSA) model and USLEM. At the same

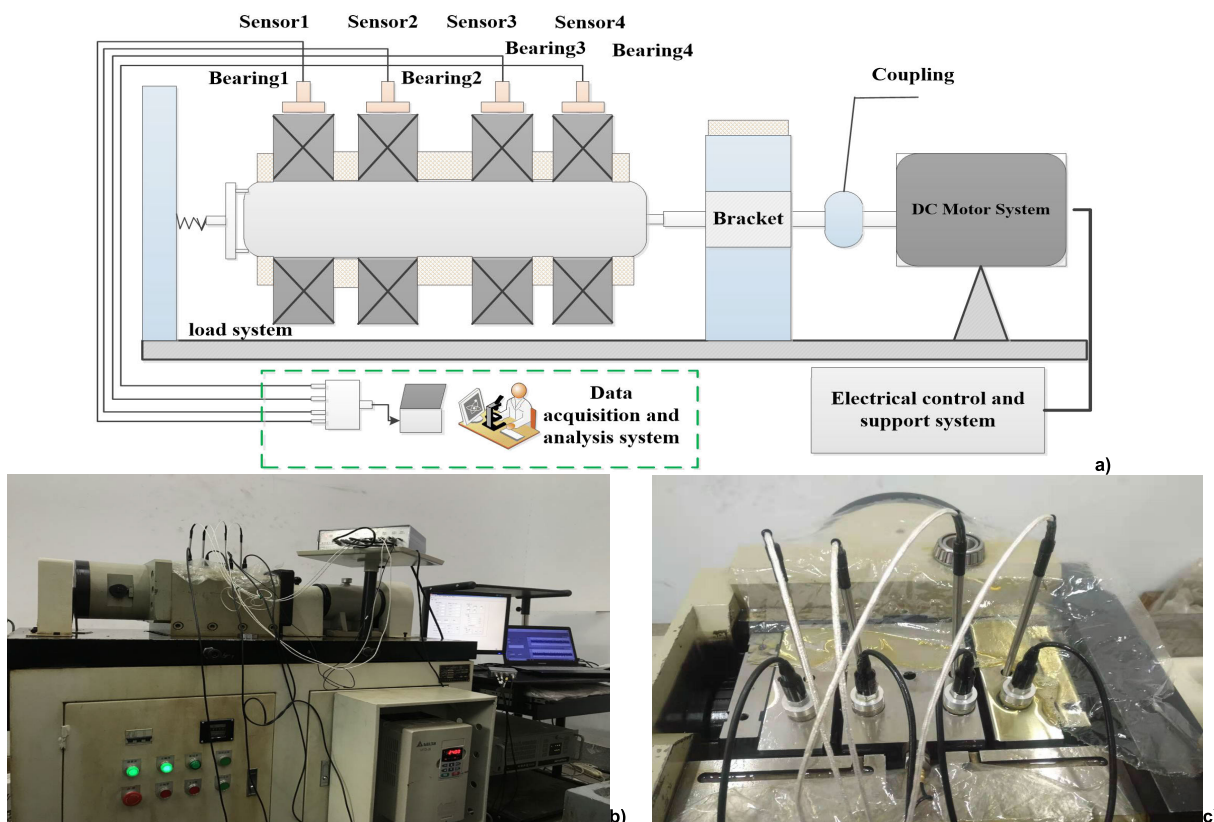


FIGURE 5. Motor-bearing system test bench (ABLT-1A): a). structure schematic diagram of test rig; b). real-life diagram of test rig; c). the schematic diagram of the sensor arrangement.

time, the proposed fault diagnosis framework can achieve effective sparse and sensitive feature information extraction in the feature extraction stage to avoid over-fitting. In the fault isolation phase, the traditional deep learning model can be rid of the supervised and low training efficiency by OUELM. Eventually, the experimental data of rolling bearings validated the effectiveness of the proposed method. The specific steps of the ASCAE-OUSELM based rolling bearing fault diagnosis framework are shown as follows:

- At the key components (rolling bearings) of the rotating machine, the vibration signals of mechanical equipment are collected by arranging the corresponding different sensors.
- The collected vibration signals can be truncated to obtain a plurality of sets of samples, and the pre-processed signals are transformed by FFT or the like, and randomly divided into a training sample set and a testing sample set;
- Initialized the parameters of ASCAE and OUSELM models to generate an initial population, the number of populations is set to N , and the evolution algebra is set to M ;
- Each ASCAE model should be gradually trained and learned, and the extracted features were entered into the OUSELM model, and the error rate of the training samples was regarded as the fitness function.

Afterwards, the fault diagnosis framework based on ASCAE-OUSELM has trained;

- The testing sample set will be input into the optimized fault diagnosis model (ASCAE-OUSELM) to obtain fault separation and diagnosis results.

IV. EXPERIMENTS AND ANALYSIS FOR FAULT DIAGNOSIS FRAMEWORK OF ROLLING BEARINGS

To illustrate the availability of the proposed fault diagnosis framework, this section we can validate the superiority of the proposed framework through the cases of rolling bearing fault experimental dataset (i.e. our laboratorial rolling bearing fault data set of Accelerated Bearing Life Tester (ABLT-1A) in Southeast University (SEU).

A. DATA ACQUISITION FROM ACCELERATED BEARING LIFE TESTER (ABLT-1A)

The ABLT-1A in this experiment case is suitable for the fatigue life strengthening test of rolling bearings with an inner diameter of $\varphi 10-60\text{mm}$. Fig.5 a) and Fig. 5 b) displayed the structure and real-life diagram of Accelerated Bearing Life Tester (ABLT-1A). The test machine mainly consists of test head, test head base, transmission system, loading system, lubrication system, electrical control system, computer monitoring system and other components, respectively. The test head can be installed in the experimental head block.

TABLE 1. The experimental parameters and main technical indicators of motor-bearing system (ABLT-1A).

Parameter/device	Value /type
Driving speed	1050 r/min
Sample frequency	12800Hz
Sensor	PCB eddy current transducer
Data acquisition system	NI 9234
Collection procedures	LabVIEW, MATLAB
PC environment	Intel(R)core (TM)i7-9750H CPU@2.60GH
Bearing designation	Single row deep groove ball bearing 6308
Number of bearing tests	4 sets
Loading	0kN
Data saving interval	30s
Length of vibration signal data	25600
Whole machine size	1500*720*1300mm

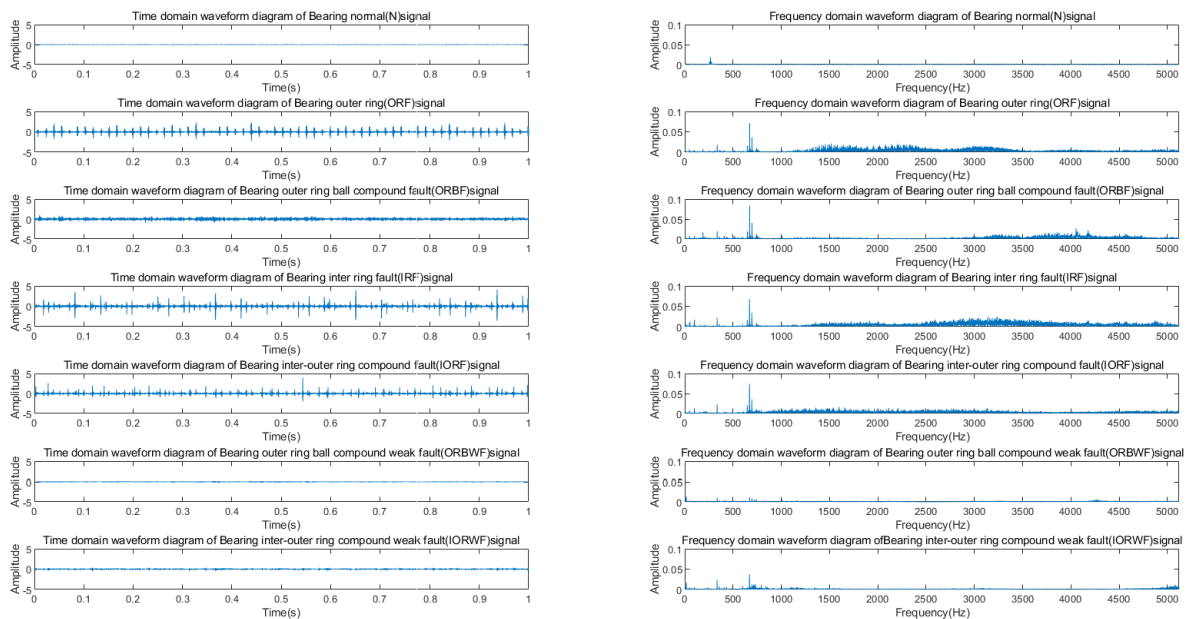


FIGURE 6. Time domain and frequency domain waveform of the motor-bearing system.

The traditional system transmits the movement of the motor, and its test shaft is rotated according to a certain speed through the coupling. And the loading system provides the load required for the tester, and the lubrication system makes the test shaft fully lubricated under normal conditions.

In this experiment, the object to be tested is a single row deep groove ball bearing with the model number 6205. A fault bearing was installed at the sensor 1st channel, and the other 3 normal bearings were installed in the 2nd, 3rd, and 4th channel sensors, respectively. Accordingly, the schematic diagram of the sensor arrangement was displayed in Fig. 5 c). And normal (N), outer ring fault (ORF), outer ring ball compound fault (ORBF), inner ring fault (IRF), and inner and outer ring compound fault (IORF), outer ring ball compound weak fault (ORBWF), inner and outer ring compound weak fault (IORWF) and other seven categories of health condition were simulated under zero load condition. When the speed is 17.5HZ, and the sampling frequency is 10240HZ, the data

is collected for 5s every 1min interval. At last, the vibration signal can be picked up by the eddy current sensor, and the electrical signal is converted into digital signal to the PC through the data acquisition card. Data acquisition and signal analysis are carried out by the software platforms such as LabVIEW and MATLAB. The specific parameters of the experiment are described in Tab 1.

In this paper, the experimental data can be selected with 1024 vibration points to intercept the length, and each state intercepts 200 groups of samples, which can be divided into training samples and testing samples, respectively. To characterize mechanical behavior from multiple angles, the vibration signals of rolling bearing can be converted into frequency domain signals. In the rolling bearing fault data, the time domain and unilateral spectrum frequency domain wave-forms of the vibration signals under the seven types of rolling bearings of the rolling bearing are displayed in Fig. 6, respectively.

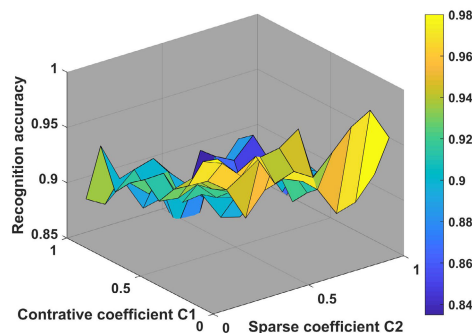


FIGURE 7. The overall recognition rate change with the regularization adjustment parameters C1 and C2.

B. PARAMETER SETTING AND OPTIMIZATION OF THE PROPOSED DIAGNOSTIC FRAMEWORK

To enhance the reliability of the diagnosis results and reduce the influence of data, the training samples and testing samples of each type for different fault diagnosis methods will be repeated 10 times, and the average accuracy of 10 time for testing samples will be taken as the final result to evaluate the performance of the diagnostic framework. According to the proposed fault diagnosis framework in this paper, the multi-layer ASCAEs algorithm can be firstly constructed, including input and output layers and two hidden layers. The number of neurons in the inputting layer can be determined by the dimension of the compressed data, and the number of neurons in the output layer can be determined by the number of bearing health categories. Taking the inputting of 1024 dimensions as an example, according to the data compression theory, the principle of halving should be adopted, that is, the number of input neurons of the designed multi-layer ASCAEs is 1024; that is, the number of the first hidden layer and the second hidden layer of the multi-layer ASCAEs are respectively set to 500 and 200, the number of output neurons is 7, that is, the structure of the designed DNN model constructed based on ASCAE-OUSELM is 1024-500-200-1024-2000-7. The other main parameters of the designed ASCAE-OUSELM algorithm are weight penalty coefficient, learning rate, and so forth.

Furthermore, according to the proposed fault diagnosis method, the testing sample is the same as the training samples, and USELM can be employed as classifier to separately study the variation rules of the Sparse coefficient (C1) and Contrastive coefficient (C2) are described in Fig. 7. It can be seen that the simultaneous optimization of the two types of parameters is difficult to determine the optimal value of each of the two parameters, and only a part of the area is better.

λ is the homotopy coefficient of the degree of contribution of the regularization term after adding the regularization of homotopy, but the two parameters can be normalized into the optimization of a parameter target in the designed ASCAE algorithm, and the parameters' range of homotopy can be reduced to [0, 1]. The homotopy coefficient (λ) can be obtained in Fig. 8 by the grid search method. It can be seen that the homotopy coefficient $\lambda = 0.4$ is more suitable.

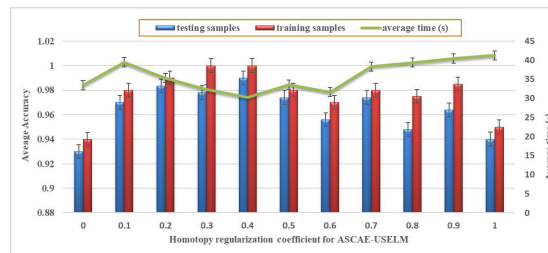


FIGURE 8. Influence of the homotopy coefficient for ASCAE feature extraction.

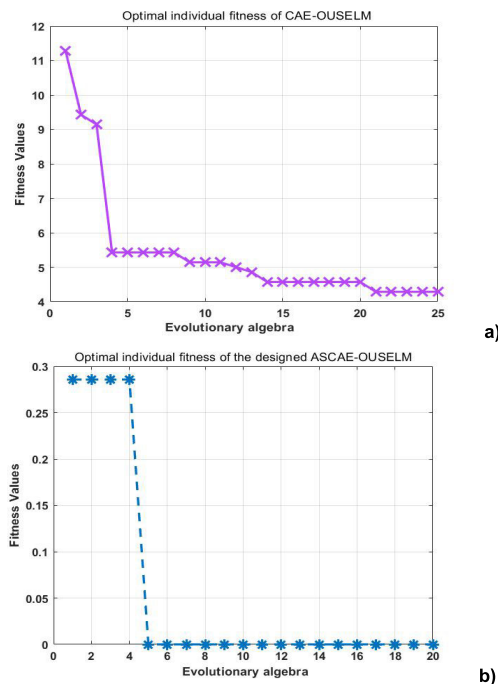


FIGURE 9. Adaptive fitness value for optimized USELM by CSA: a). Optimal individual fitness of CAE-OUSELM; b). Optimal individual fitness of ASCAE-OUSELM.

More importantly, as can be known from the above-mentioned equations that the relevant parameters affecting the dimensionality reduction effect of the USELM algorithm are divided into two parts, one is related to the Laplacian matrix's penalty coefficient (λ), and the other is related to the selection of the neighboring number (NN) in the USELM embedded process. In the calculation process of the Laplacian matrix L , the most important thing is to select the number of nearest neighbors when constructing the neighbor graph. If the number is too small, the local characteristic range retained by the matrix L is too small, and the data points have similar characteristics. On the contrary, if the number is too large, the local characteristic range of the reservation is too large, and the data points that are far apart are close in the low-dimensional space. In both cases, the data after the dimension reduction is difficult to cluster. Regarding the setting of these two key parameters, since there are no good rules to refer to, the traditional method can only be determined through repeated experimental tests.

TABLE 2. Parameters setting up of the designed ASCAE-OUSELM algorithms.

Parameter name	Value	Parameter name	Value
Learning rate of ASCAE	0.1	nn.output of ASCAE	'sigm'
The number of hidden layers of ASCAE	2	Number of iterations of ASCAE	100
The number of units in the input layer of ASCAE	1024	nn. weight Penalty L2 of ASCAE	0
The number of units in the first hidden layer of ASCAE	500	homotopy coefficient of ASCAE	0.4
The number of units in second hidden layer of ASCAE	200	Batchsize of ASCAE	25
The input units of OUSELM	1024	Number of nests of OUSELM	n=20
The hidden units of OUSELM	2000	Iteration of OUSELM	25
The output units of OUSELM	17	Discovery probability of OUSELM	$P_{\alpha}=0.25$
Momentum of ASCAE	0.5	$gt^*(NN, lam)$ of OUSELM	$= [9, 0.12]$

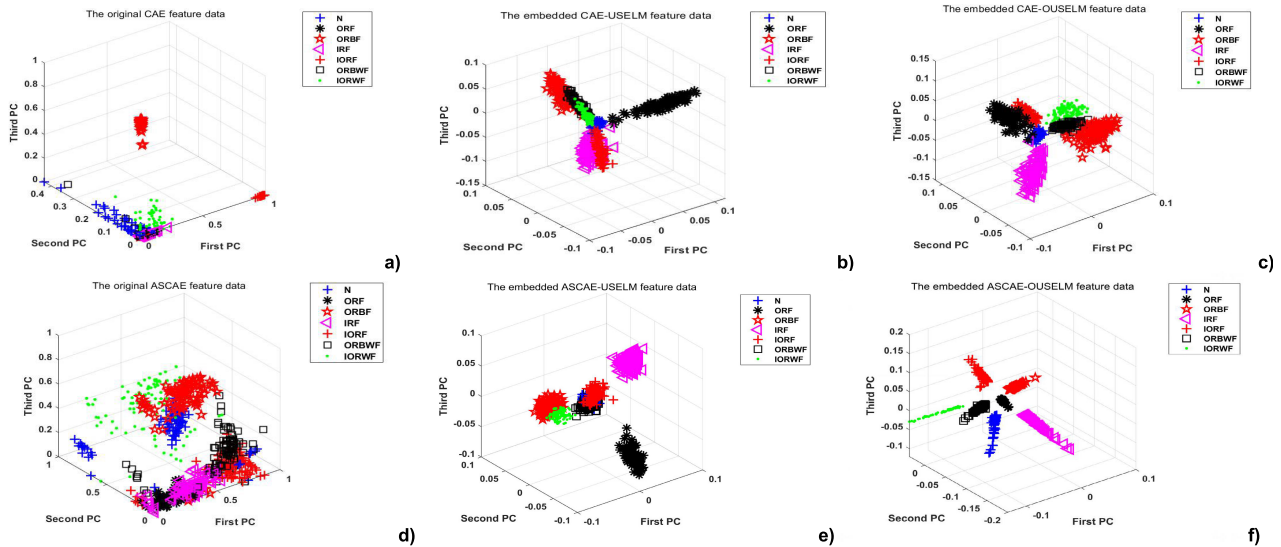


FIGURE 10. Scatter plot based on the different fault diagnosis framework of feature embedding: a). CAE; b). CAE-USELML; c). CAE-OUSELM; d). ASCAE; e). ASCAE-USELML; f). ASCAE-OUSELM.

According to the flow chart described in Fig. 3 of OUSELM-based intelligent optimization algorithm, the two types of parameters are optimized by CSA, and the fitness curve of the CSA optimized USELML for feature extraction by CAE and ASCAE can be displayed in Fig. 9, respectively. Correspondingly, the optimized parameter combination of ASCAE-OUSELM is $gt^*(NN, lam) = [9, 0.12]$.

To sum up, parametric performance analysis and optimization based on ASCAE-OUSELM, the parameter settings of the designed fault diagnosis framework based on ASCAE-OUSELM are more specifically described in Tab 2.

C. FAULT ISOLATION OF ROLLING BEARINGS BY THE PROPOSED FAULT DIAGNOSIS FRAMEWORK

To verify the validity of the proposed ASCAE-OUSELM based fault diagnosis framework of rolling bearings, the above-mentioned constructed rolling bearing fault data set is applied and entered into the proposed fault diagnosis framework (the parameter settings are shown in Tab. 2 and the fault diagnosis flow chart is shown in Fig 4). At the same time, the number of training sample and testing sample are equal in this section, which is 100:100. And the traditional CAE-based fault diagnosis method as the benchmark algorithm for this part of the experiment, and the

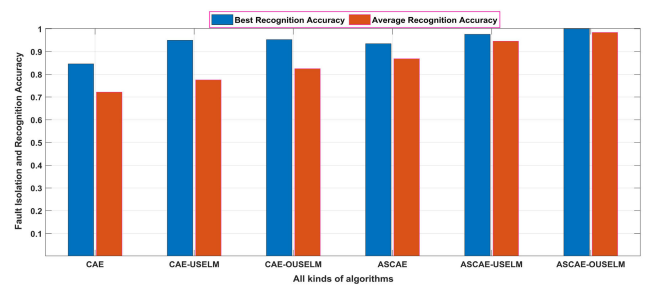


FIGURE 11. The best diagnosis and average diagnosis result under different fault diagnosis framework.

rolling bearing fault diagnosis can be performed according to the above fault diagnosis process. Afterwards, the features extracted by the combination fault diagnosis framework of CAE-USELML, CAE-OUSELM, ASCAE, ASCAE-USELML, ASCAE-OUSELM are input into K-means respectively to obtain the three-dimensional feature distribution before and after embedding, which can be specifically displayed in Fig.10.

The essence of the above-mentioned feature embedding methods are employed to construct a Laplacian matrix instead of a category label to measure the relationship between data, thereby unsupervised feature embedding can

be implemented. At the same time, the embedding features extracted from the above-mentioned different optimization combination models, such as CAE; CAE + USELM; CAE + OUSELM; ASCAE; ASCAE + USELM; ASCAE + OUSELM were entered into FCM cluster learning for fault diagnosis, respectively. Furthermore, the best diagnosis results with 100 times fault diagnosis results and its average fault diagnosis results can be displayed in Fig. 11.

It can be seen from the Fig. 11 that the proposed method (ASCAE-OUSELM) achieves the best fault separation effect, and the average recognition rate is also the best.

D. COMPARED WITH OTHER FAULT DIAGNOSIS FRAMEWORKS

To further validate the effectiveness of the proposed framework, this section we compare the proposed ASCAE-OUSELM framework with other standard deep learning model fault diagnosis frameworks. Among them, the

comparison algorithm ASCAE_OUSELM ASCAE_USELM CAE_USELM SAE-USELM CAE_OUSELM DBN-USELM are selected as the contrast frameworks. It should be noted that the data utilization situation here is set as below: the training sample randomly selects 50 samples for each type of health condition, and the testing sample is the remaining 150 groups of samples, the total training sample is $50 \times 7 = 350$ groups, and the testing samples are $150 \times 7 = 1050$ groups. Compared to the he above sample settings, this part of the training sample is less, and its generalization performance requirements are higher. According to the above-mentioned fault diagnostic flowchart, the sample diagnosis results of the above-mentioned six fault diagnosis methods are described in Fig. 12 a)-f) by using Fuzzy C-Means (FCM) [42], respectively. It can be seen from the different clustering diagram that only the accuracy of ASCAE_OUSELM is close to 100%, and the seven health states of the bearing can be accurately identified.

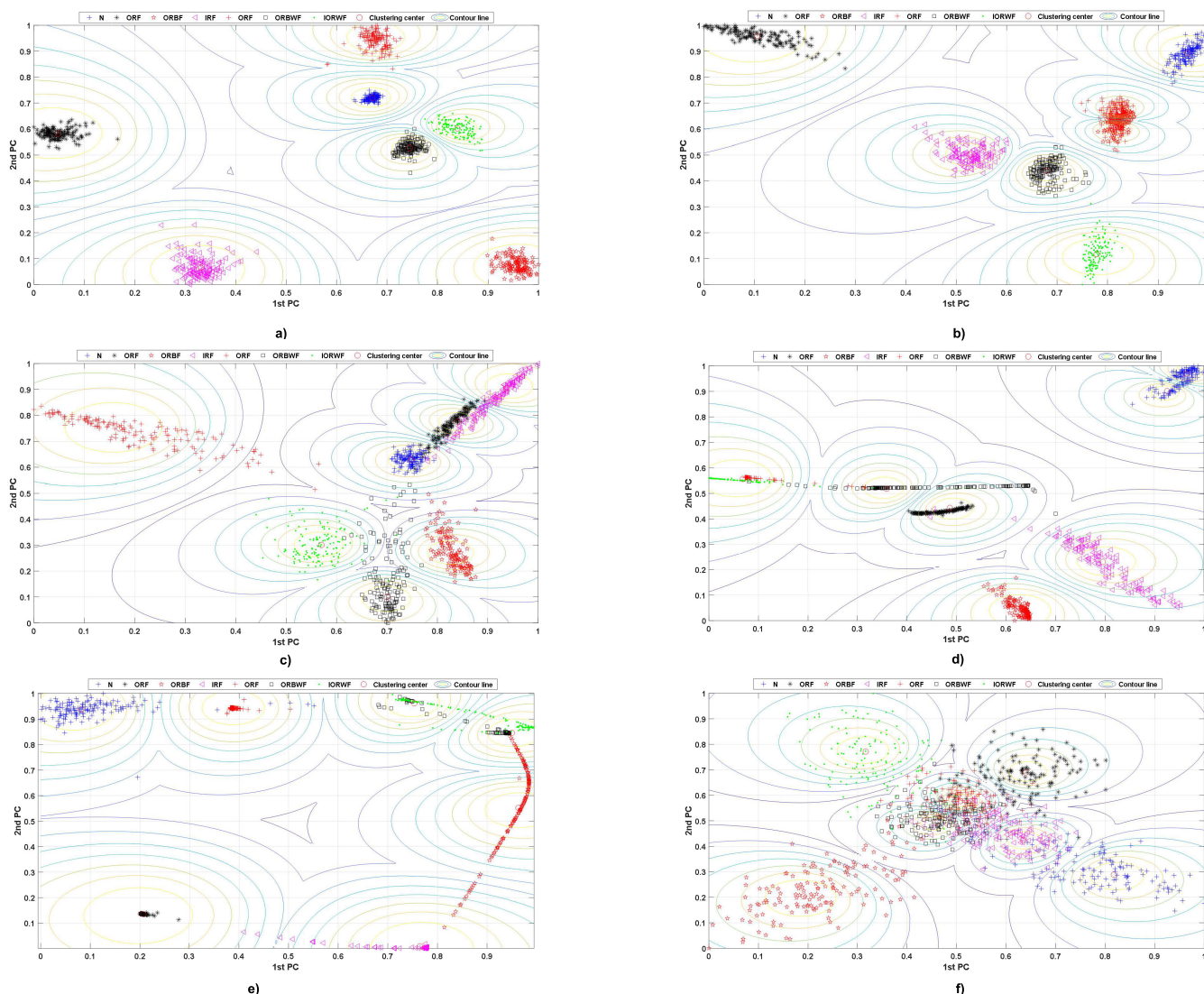


FIGURE 12. Clustering contour maps for different fault diagnosis frameworks: a). ASCAE_OUSELM; b). ASCAE_USELM; c). CAE_USELM d). SAE-USELM; e). CAE_OUSELM; f). DBN-USELM.

TABLE 3. Evaluation indicators of different fault diagnosis frameworks.

Frameworks	DNN-OUSELM	DBN-OUSELM	SAE-OUSELM	CAE-OUSELM	ASCAE-USELM	ASCAE-OUSELM
Best_ACC	0.918095	92.4762	0.939048	0.935238	0.971429	1
Ave_ACC	78.7838	0.825714	0.75559	80.7171	0.879048	0.976733
Standard deviation	2.534	1.576	2.358	0.6875	0.4580	0.3347
PC	0.5508	0.8981	0.8243	0.7403	0.8417	0.9201
CE	0.9711	0.2280	0.3823	0.5750	0.4005	0.1996

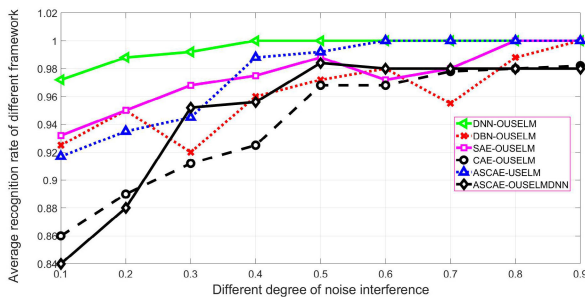


FIGURE 13. Effects of the various clustering methods in response to changes in interference coefficients.

At the same time, the clustering results of the above-mentioned fault diagnosis frameworks are calculated by the membership degree and clustering evaluation indexes such as Partition Coefficient (PC) and Classification entropy (CE) [8], the diagnostic results can be obtained in Table 3 according to the diagnostic results and clustering evaluation indexes.

From the Tab. 3, the proposed ASCAE-OUSELM framework can achieve the best recognition effect, owing to ASCAE has strong feature extraction ability and its optimized OUSELM can quickly and stably separate the seven types of faults.

To validate the anti-noise performance of the ASCAE-OUSELM based-diagnostic framework of rolling bearing, the raw data can be mixed with noise according to different interference coefficients g [19], [43], and other parameters (such as the ratio of training samples to test samples) remain unchanged. Random noise of $g = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8,$ and 0.9 were added to the experimental data set in this section, respectively. Therefore, the data set X becomes $X_{new} = x + a \cdot \text{rand}(\text{size}(x))$, where $\text{size}(x)$ represents the size of the signal, and the $\text{rand} (*)$ function is a random function in Matlab generates a random number in the function. Calculating the average recognition rate of the test samples in the above six fault diagnosis frameworks under different noise interference conditions, as shown in Fig. 13.

It can be seen from the Fig.13 that the recognition rate of the proposed fault diagnosis framework is relatively concentrated and stable, and its accuracy is always higher than other fault diagnosis frameworks. It is demonstrated that the fault diagnosis framework based on ASCAE-OUSELM has strong generalization ability and good stability.

E. DISCUSSION AND FUTURE WORK

- 1). In the designed ASCAE-OUSELM based-fault diagnosis framework, the parameter adjustment of the neural network’s structural parameters are just set through experience. This will cause a series of problems. We will optimize the network structure in detail in the future.
- 2). In this experiment, only rolling bearing fault data set of ABLT-1A was applied to validate the effectiveness of the proposed fault diagnosis framework. Later, we will apply the fault diagnosis framework to other bearing fault data sets for verification.

V. CONCLUSIONS

More recently, data-driven fault diagnosis has gradually become one of the mainstream trends in intelligent fault diagnosis of rolling bearings. Aiming at the traditional deep neural network cannot implement adaptive feature extraction and fault separation, this paper proposes a new rolling bearing fault diagnosis framework based on the designed ASCAE combined with OUSELM. This designed fault diagnosis framework can be divided into three main steps: i). Firstly, the vibration signals of rolling bearings can be collected and processed on the key parts of the rotating machine, and the collected vibration signal can be converted into the frequency signals; ii). Secondly, the transformed spectral signals can be entered into the constructed ASCAE for feature learning to exploit the multi-layer sensitive features from the hidden inside the data; iii). Thirdly, the extracted multi-layer sensitive features are input to the trained OUSELM classifier for unsupervised fault state separation and diagnosis. Among them, our proposed fault diagnosis framework (ASCAE-OUSELM) can employ the homotopy regularization theory, sparse theory, CSA intelligent optimization algorithm and other tools to optimize the parameters and improve the performance of the original Contrative Auto-encoder (CSA) model and USLEM, respectively. At the same time, the experimental case of rolling bearing validated the superiority of the designed fault diagnosis framework and two new algorithms.

ACKNOWLEDGMENT

Thanks to Peng Ding (Southeast University), Cheng Yang (Southeast University), Lin Zhu (Yangzhou University) and others for their help and contribution. Meanwhile, the authors would like to appreciate the anonymous reviewers and the editor for their valuable comments.

REFERENCES

- [1] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, Feb. 2018.
- [2] X. Zhao and M. Jia, "A new local-global deep neural network and its application in rotating machinery fault diagnosis," *Neurocomputing*, vol. 366, pp. 215–233, Nov. 2019.
- [3] M. Xia, T. Li, L. Xu, L. Liu, and C. W. De Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME Trans. Mechatron.*, vol. 23, no. 1, pp. 101–110, Feb. 2018.
- [4] X. Li, W. Zhang, N.-X. Xu, and Q. Ding, "Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places," *IEEE Trans. Ind. Electron.*, Early Access, doi: 10.1109/tie.2019.2935987.
- [5] Z. Chen, K. Gryllias, and W. Li, "Mechanical fault diagnosis using convolutional neural networks and extreme learning machine," *Mech. Syst. Signal Process.*, vol. 133, Nov. 2019, Art. no. 106272.
- [6] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, Aug. 2018.
- [7] F. Zhang, J. Huang, F. Chu, and L. Cui, "Mechanism and method for outer raceway defect localization of ball bearings," *IEEE Access*, to be published, doi: 10.1109/ACCESS.2019.2950284.
- [8] X. Zhao and M. Jia, "A novel deep fuzzy clustering neural network model and its application in rolling bearing fault recognition," *Meas. Sci. Technol.*, vol. 29, no. 12, Dec. 2018, Art. no. 125005.
- [9] Y. Chen, X. Liang, and M. J. Zuo, "Sparse time series modeling of the baseline vibration from a gearbox under time-varying speed condition," *Mech. Syst. Signal Process.*, vol. 134, Dec. 2019, Art. no. 106342.
- [10] Q. Ni, K. Wang, and J. Zheng, "Rolling element bearings fault diagnosis based on a novel optimal frequency band selection scheme," *IEEE Access*, vol. 7, pp. 80748–80766, 2019.
- [11] X. Yan and M. Jia, "Application of CSA-VMD and optimal scale morphological slice bispectrum in enhancing outer race fault detection of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 122, pp. 56–86, May 2019.
- [12] X. Ding, Q. He, and N. Luo, "A fusion feature and its improvement based on locality preserving projections for rolling element bearing fault classification," *J. Sound Vib.*, vol. 335, pp. 367–383, Jan. 2015.
- [13] X. Zhu, J. Xiong, and Q. Liang, "Fault diagnosis of rotation machinery based on support vector machine optimized by quantum genetic algorithm," *IEEE Access*, vol. 6, pp. 33583–33588, 2018.
- [14] K. Yu, T. R. Lin, H. Ma, H. Li, and J. Zeng, "A combined polynomial chirplet transform and synchroextracting technique for analyzing nonstationary signals of rotating machinery," *IEEE Trans. Instrum. Meas.*, to be published, doi: 10.1109/tim.2019.2913058.
- [15] M. Farajzadeh-Zanjani, R. Razavi-Far, and M. Saif, "Efficient sampling techniques for ensemble learning and diagnosing bearing defects under class imbalanced condition," in *Proc. IEEE Symp. Comput. Intell. (SSCI)*, Dec. 2016, pp. 1–7.
- [16] W. Teng, X. Ding, X. Zhang, Y. Liu, and Z. Ma, "Multi-fault detection and failure analysis of wind turbine gearbox using complex wavelet transform," *Renew. Energy*, vol. 93, pp. 591–598, Aug. 2016.
- [17] X. Liu, L. Bo, and H. Luo, "Bearing faults diagnostics based on hybrid LS-SVM and EMD method," *Measurement*, vol. 59, pp. 145–166, Jan. 2015.
- [18] Z. Li, J. Chen, Y. Zi, and J. Pan, "Independence-oriented VMD to identify fault feature for wheel set bearing fault diagnosis of high speed locomotive," *Mech. Syst. Signal Process.*, vol. 85, pp. 512–529, Feb. 2017.
- [19] X. Zhao and M. Jia, "Fault diagnosis of rolling bearing based on feature reduction with global-local margin Fisher analysis," *Neurocomputing*, vol. 315, pp. 447–464, Nov. 2018.
- [20] X. Dai and Z. Gao, "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis," *IEEE Trans. Ind. Informat.*, vol. 9, no. 4, pp. 2226–2238, Nov. 2013.
- [21] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.
- [22] Z. Su, B. Tang, J. Ma, and L. Deng, "Fault diagnosis method based on incremental enhanced supervised locally linear embedding and adaptive nearest neighbor classifier," *Measurement*, vol. 48, no. 1, pp. 136–148, Feb. 2014.
- [23] Y. Li, X. Liang, Y. Wei, and X. Wang, "A method based on refined composite multi-scale symbolic dynamic entropy and ISVM-BT for rotating machinery fault diagnosis," *Neurocomputing*, vol. 315, pp. 246–260, Nov. 2018.
- [24] R. Razavi-Far, E. Hallaji, M. Farajzadeh-Zanjani, M. Saif, S. H. Kia, H. Henao, and G.-A. Capolino, "Information fusion and semi-supervised deep learning scheme for diagnosing gear faults in induction machine systems," *IEEE Trans. Ind. Electron.*, vol. 66, no. 8, pp. 6331–6342, Aug. 2019.
- [25] P. Tamilselvan and P. Wang, "Failure diagnosis using deep belief learning based health state classification," *Rel. Eng. Syst. Saf.*, vol. 115, pp. 124–135, Jul. 2013.
- [26] G. Jiang, H. He, J. Yan, and P. Xie, "Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3196–3207, Apr. 2019.
- [27] H. Shao, H. Jiang, H. Zhang, and T. Liang, "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2727–2736, Mar. 2018.
- [28] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, Jul. 2016.
- [29] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Dec. 2010.
- [30] 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.
- [31] W. Jiang, J. Zhou, H. Liu, and Y. Shan, "A multi-step progressive fault diagnosis method for rolling element bearing based on energy entropy theory and hybrid ensemble auto-encoder," *ISA Trans.*, vol. 87, pp. 235–250, Apr. 2019.
- [32] J. Zhang, K. Li, Y. Liang, and N. Li, "Learning 3D faces from 2D images via stacked contractive autoencoder," *Neurocomputing*, vol. 257, pp. 67–78, Sep. 2017.
- [33] H. Shao, H. Jiang, Y. Lin, and X. Li, "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders," *Mech. Syst. Signal Process.*, vol. 102, pp. 278–297, Mar. 2018.
- [34] Z. Zhao, S. Wu, B. Qiao, S. Wang, and X. Chen, "Enhanced sparse period-group lasso for bearing fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 66, no. 3, pp. 2143–2153, Mar. 2019.
- [35] Z. Zhang, S. Li, J. Wang, Y. Xin, and Z. An, "General normalized sparse filtering: A novel unsupervised learning method for rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 124, pp. 596–612, Jun. 2019.
- [36] T. Y. Li and N. H. Rhee, "Homotopy algorithm for symmetric eigenvalue problems," *Numer. Math.*, vol. 55, no. 3, pp. 265–280, May 1989.
- [37] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, vol. 2, Apr. 2005, pp. 985–990.
- [38] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.
- [39] G. Huang, S. Song, J. N. D. Gupta, and C. Wu, "Semi-supervised and unsupervised extreme learning machines," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2405–2417, Dec. 2014.
- [40] X.-S. Yang and S. Deb, "Cuckoo search: Recent advances and applications," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 169–174, Jan. 2014.
- [41] X. Jia, M. Zhao, Y. Di, P. Li, and J. Lee, "Sparse filtering with the generalized l_p/l_q norm and its applications to the condition monitoring of rotating machinery," *Mech. Syst. Signal Process.*, vol. 102, no. 1, pp. 198–213, Mar. 2018.
- [42] C. Lu, S. Wang, and C. Zhang, "Fault diagnosis of hydraulic piston pumps based on a two-step EMD method and fuzzy C-means clustering," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, vol. 230, no. 16, pp. 2913–2928, Oct. 2016.
- [43] F. Chen, B. Tang, and R. Chen, "A novel fault diagnosis model for gearbox based on wavelet support vector machine with immune genetic algorithm," *Measurement*, vol. 46, no. 1, pp. 220–232, Jan. 2013.



XIAOLI ZHAO (Student Member, IEEE) received the M.S. degree from the Lanzhou University of Technology, Lanzhou, China, in 2017. He is currently pursuing the Ph.D. degree with Southeast University, Nanjing, China. He is a Visiting Ph.D. Student with the School of Engineering, The University of British Columbia, Okanagan. His main research interests include mechanical signal processing, electromechanical equipment intelligent monitoring and fault diagnosis, and brain-computer interaction.



MINPING JIA (Member, IEEE) received the B.S. and M.S. degrees in mechanical engineering from the Nanjing Institute of Technology (now Southeast University), Nanjing, China, in 1982 and 1985, respectively, and the Ph.D. degree in mechanical engineering from Southeast University, in 1991. He is currently a Full Professor with Southeast University. His research interests include dynamic signal processing, machine fault diagnosis, and vibration engineering applications.



ZHENG LIU (Senior Member, IEEE) received the Ph.D. degree in engineering from Kyoto University, Kyoto, Japan, in 2000, and the Ph.D. degree from the University of Ottawa, Canada, in 2007. From 2000 to 2001, he was a Research Fellow with Nanyang Technological University, Singapore. He then joined the Institute for Aerospace Research (IAR), National Research Council Canada, Ottawa, ON, Canada, as a Governmental Laboratory Visiting Fellow nominated by NSERC. After being with IAR for five years, he transferred to the NRC Institute for Research in Construction, where he was a Research Officer. From 2012 to 2015, he was a Full Professor with the Toyota Technological Institute, Nagoya, Japan. He is currently with the School of Engineering, The University of British Columbia, Okanagan. His research interests include image/data fusion, computer vision, pattern recognition, sensor/sensor networks, condition-based maintenance, and non-destructive inspection and evaluation. He is a member of SPIE. He is the Chair of the IEEE IMS Technical Committee on industrial inspection (TC-36). He holds a Professional Engineer License in British Columbia and Ontario. He serves on the Editorial Board for journals: the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, the IEEE TRANSACTIONS ON MECHATRONICS, the *IEEE Instrumentation and Measurement Magazine*, the IEEE JOURNAL OF RADIO FREQUENCY IDENTIFICATION, *Information Fusion*, *Machine Vision and Applications*, and *Intelligent Industrial Systems*.

• • •