Semisupervised Graph Convolution Deep Belief Network for Fault Diagnosis of Electormechanical System With Limited Labeled Data

Xiaoli Zhao, *Student Member, IEEE*, Minping Jia, and Zheng Liu, *Senior Member, IEEE*

***Abstract*—The labeled monitoring data collected from the electromechanical system is limited in the real industries; traditional intelligent fault diagnosis methods cannot achieve satisfactory accurate diagnosis results. To deal with this problem, an intelligent fault diagnosis method for electromechanical system based on a new semisupervised graph convolution deep belief network algorithm is proposed in this article. Specifically, the labeled and unlabeled samples are first employed to design a new adaptive local graph learning method for constructing the graph neighbor relationship. Meanwhile, the labeled samples are applied to describe the discriminative structure information of data via the latest circle loss. Finally, the local and discriminative objective functions are reconstructed under the semisupervised learning framework. The experimental results from the motor-bearing system demonstrate that the method can achieve 98.66% accuracy with only 10% of training labeled data, which indicates that it is a promising semisupervised intelligent fault diagnosis method.**

***Index Terms*—Convolution deep belief network (CDBN), electromechanical system, fault diagnosis, graph neural network (GNN), semisupervised learning.**

I. Introduction

IN THE era of the interconnection of all things, the electromechanical system, as the most common component of industries, is the key bearer of the industrial system [1], [2]. It also undertakes the majority of the manufacturing tasks of industrialization, informationization, and intelligentization [1]–[3]. Nevertheless, the electromechanical system usually has a more complex operating environment with a long operating time and heavy load in the numerous industries [4], [5]. Therefore, the ultralong standby and complex industrial environment may cause certain failures and damage to the electromechanical system, which will cause declining in industrial productivity and some economic losses [3]–[5]. Therefore, it is necessary to carry out efficient and advanced intelligent fault diagnosis approach of electromechanical systems (especially, its core components such as motors, bearings, rotors, gears, etc.) [2], [5], [6].

With the advances in computer science, measurement, and sensing, the most common intelligent fault diagnosis method is currently developed on machine learning [7], [8]. Recently, intelligent fault diagnosis has been transformed traditional fault diagnosis methods based on models or signals into data-driven fault diagnosis methods [4], [5]. According to the evolution of feature extraction and classification, data-driven fault diagnosis can be further divided into the shallow and multiple-layer [9], [10]. The structure of shallow model rarely uses hidden layers or even no hidden layers. It's nonlinear learning performance to solve complex problems is restricted, which may seriously affect detection and diagnosis accuracy. In contrast, the typical representative of multiple-layer fault diagnosis methods is deep learning (DL) algorithm [11]–[13]. DL can integrate feature extraction and fault classification, which can greatly solve shallow learning shortcomings and significantly improve fault diagnostic performance. Generally, DL can be divided into three categories: Discriminant (such as convolutional neural networks, CNN) [9], generative (such as deep Belief networks, DBN) [11], and hybrid [3], respectively. To perform intelligent fault diagnosis on the severity of bearing faults, Yu *et al.* [11] proposed a new intelligent fault diagnosis framework, which combined deep belief network (DBN) and Dempster–Shafer (DS). To reduce the irrelevant information of the data collected by the soft sensor during the industrial process, Yuan *et al.* [13] first proposed a new deep learning network based on a deep quality-driven autoencoder (SQAE) for quality-related feature representation. The authors in [9] designed a new framework based on an improved convolutional neural network (CNN) with transfer learning for fault diagnosis of the rotor-bearing system

Manuscript received July 22, 2020; revised September 16, 2020 and October 8, 2020; accepted October 14, 2020. Date of publication October 29, 2020; date of current version May 3, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 52075095, in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province, China under Grant SJKY19_0064, and Grant KYCX19_0063. Paper no. TII-20-3529. *(Corresponding author: Minping Jia.)*

Xiaoli Zhao is with the School of Mechanical Engineering Southeast University, Nanjing 211189, China, and also with the School of Engineering the University of British Columbia, Kelowna, BC V1V 1V7, Canada (e-mail: zhaoxiaoli5258@163.com).

Minping Jia is with the School of Mechanical Engineering Southeast University, Nanjing 211189, China (e-mail: mpjia@seu.edu.cn).

Zheng Liu is with the School of Engineering the University of British Columbia, Kelowna, BC V1V 1V7, Canada (e-mail: zheng.liu@ubc.ca).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2020.3034189

under different working conditions. To improve the diagnostic accuracy of unbalanced data, Zhao *et al.* [14] first developed an unbalanced fault diagnosis framework based on deep Laplacian autoencoder (DLapAE). To accurately simulate the dynamic and nonlinear performance of the industrial process, the authors in [15] proposed a supervised long short-term memory (SLSTM) network for learning quality-related predictions. To conclude, different DL methods have been widely and successfully applied to the intelligent fault diagnosis [14], [16], industrial process control [15], and other fields.

Inspired by the DBN algorithm and the local weight sharing of CNN, the convolutional deep belief network (CDBN) was first proposed by Lee *et al.* [17], which not only has the unsupervised characteristics of DBN algorithm but also inherits the local feature extraction capabilities of CNN algorithm. Accordingly, a new fault diagnosis framework for the electric locomotive bearings based on CDBN was developed [3]. Zhao *et al.* [18] proposed a fault diagnosis method based on the local-global deep neural network (LGDNN) algorithm. The core is CDBN based on the Fisher optimization criterion (called Fisher-CDBN). However, the aforementioned traditional DL-based intelligent fault diagnosis methods still have the following shortcomings: 1) Most of the abovementioned DL-based fault diagnosis methods are supervised fine-tuning and classification; a large number of unlabeled sample information cannot be fully utilized; 2) the traditional DL algorithm cannot effectively extract the discriminant local structure information of data.

However, most of the existing fault diagnosis methods are established on the presupposition that the collected samples are provided with sufficient labeling information. The fact that the acquisition of the labeled information is costly [18], [19]. According to the labeled and unlabeled data, data-driven fault diagnosis can be further classified into supervised fault diagnosis, semisupervised fault diagnosis, and unsupervised fault diagnosis. Within these methods, semisupervised learning (SSL) [19] can make full use of labeled samples and unlabeled samples. Unlabeled sample information can mine the geometric structure information of data through manifold structure information to obtain better generalization performance [19], [20].

In summary, the labeled industrial data is relatively precious in the real industries, which is much less than unlabeled industrial data. To make full use of the labeled and unlabeled information of data, a new semisupervised deep convolutional belief network (SSDCDBN) algorithm is proposed in this article. On this basis, a new intelligent fault diagnosis framework of electromechanical system based on SSGCDBN algorithm is proposed. To be specific, circle loss and adaptive local graph are introduced to the original CDBN algorithm. Finally, the new objective functions are reconstructed under the semisupervised learning framework. Fig. 1 shows the main research inspiration of this intelligent fault diagnosis method. In the SSL based on manifolds or graphs, the role of the labeled sample is used to be passed down like a torch relay. Naturally, the process of labeled information transfer requires a large number of unlabeled samples to act as an intermediate torchbearer. The main contributions of this article can be summarized and highlighted as follows.

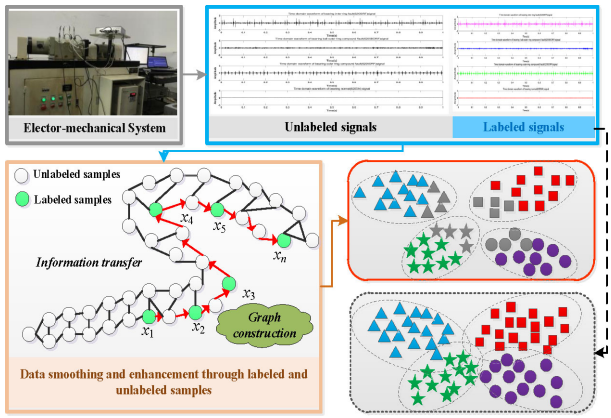


Fig. 1. Illustrations of the proposed SSL method under limited labeled samples.

1) A new graph embedded semisupervised DNN algorithm is proposed, namely, the semisupervised graph convolution deep belief network (SSGCDBN) algorithm.
2) Based on the SSGCDBN algorithm, an intelligent fault diagnosis method for electromechanical system based on SSGCDBN algorithm is proposed in this article.
3) The experimental data of the motor-bearing system verifies the effectiveness of the proposed method and algorithm.

The rest of this article are organized as follows. The related works are presented in Section II. Then, a detailed elaboration on the developed SSGCDBN algorithm and its fault diagnosis method is described in Section III. In Section IV, the effectiveness of the proposed method and algorithm is proved with the motor-bearing system. Finally, Section V concludes the article.

II. Related Works

In this section, the related intelligent fault diagnosis based on SSL and graph learning are briefly reviewed.

A. Intelligent Fault Diagnosis Based on SSL

SSL can be divided into graph-based SSL, pair-constrained SSL, and category information-based SSL, respectively. In [19], a hybrid data-driven semisupervised diagnosis framework based on semisupervised smooth alpha layering (S3AL) algorithm was presented for bearing defect diagnosis. The authors in [21] proposed an intelligent gear fault diagnosis framework in induction machine systems based on the information fusion and semisupervised deep ladder network (SSDLN) algorithm. In addition to supervising information such as class labels and pairwise constraints, samples' graph structure is also an essential manifestation of prior knowledge [21], [22]. Accordingly, the graph-based SSL method has become one of the research hotspots in machine learning. It can simultaneously employ a small number of labeled samples and many unlabeled samples to increase the learning performance of the model. Namely, the data can be represented as a graph, where each sample is regarded as a node, the similarity is measured via the connectivity

between the pair of nodes [22], [23]. The purpose of the graph neural networks (GNNs) is to mine hidden information from the data through the geometric structure relationship between the data. Therefore, learning or constructing an effective graph structure to capture the internal connection between data is a crucial step for GNNs. The authors in [23] developed an intelligent fault diagnosis method based on semisupervised extreme learning machine (SSELM). Chai *et al.* [24] designed a new semisupervised learning method by combining label and sparse constraints to update the structure of AE, named as the label and sparse regularized autoencoder (LSRAE). Lei *et al.* [25] proposed a semisupervised fault diagnosis method based on CNN-LapsELM for the detection of deep soft sensors.

B. Intelligent Fault Diagnosis Based on Graph Learning

The essence of graph-based semisupervised learning methods is to use the structural relationships of graphs to mine the geometric structure relationships of the data. This type of the method closely connects semisupervised learning with graph learning [1], [20]. Under the guidance of a small amount of category labeling information, SSL assists the tasks such as feature extraction and classification or clustering by mining a large number of unlabeled sample structure information [20], [22]. Graph-based semisupervised learning can guide the construction or learning of graphs, and supervised information can also guide the process of graph optimization. Therefore, the relationship between graph and semisupervised learning is the mutual assistance between structural information and category information [26]. Its ultimate goal is to closely link labeled information, unlabeled information, and graph-based structural information together. According to the [1], [22], the key step of graph-based semi-supervised learning method is the process of graph construction and learning.

With the emergence of deep learning, deep graph-based learning methods have been continuously developed, and fault diagnosis techniques based on GNNs have gradually become popular. Zhang *et al.* [27] aimed at traditional deep learning methods that can only learn features from the vertices of the input data, a deep graph convolutional network (DGCN) based on graph theory was proposed for acoustic fault diagnosis of rolling bearings, but this method does not extend it to the field of semisupervised learning framework.

Furthermore, most of the traditional graph construction methods rely on data distribution observation to construct a fixed similarity matrix. The construction or learning of such graphs is independent of downstream tasks [1], [22]. Generally, there are two core issues in composition, one is how to construct edges, and the other is how to weight edges. Common graph structures include $k-$nearest neighbors and ball-radius [1], which employed simple rules to connect the vertices of the graph, but these rules are very sensitive to the noise of the data, and the selection of neighbor parameters is just built on experience. Recently, Nie *et al.* [28] proposed to adaptively update the similarity matrix with rank constraints on the Laplacian operator to merge clustering or classification information in the learning graph. Compared with traditional graph construction methods, the graph structure generated by adaptive graph regularization is more explicit. In [29], the adaptive graph is also constrained by the low-rank representation (LRR). In [30], adaptive regularization is combined with spectral embedding representation to provide a more discriminative data representation. The essence of graph learning is used to utilize graphs to simulate low-dimensional manifolds and improve the data fitting ability.

In summary, as illustrated from the abovementioned analysis and industrial application, the graph-based SSL method mainly involves the following two aspects: 1) Graph construction, the key step of the graph-based SSL method is K-NN and other commonly used methods, but these methods cannot achieve the adaptive learning of local graph neighbor structure information; 2) the optimization goal of the traditional SSL method is not conducive to discriminant analysis [31].

III. Proposed Method

First, the problem formulation is defined in this section. Afterward, the designed semisupervised graph DL (termed as SSGCDBN) algorithm is developed. Finally, a semisupervised fault diagnosis method based on SSGCDBN algorithm is proposed in this section.

A. Problem Formulation

The traditional fault diagnosis method can utilize the labeled fault samples to extract the observed signals, but the geometric structure information of a large number of unlabeled samples is ignored. First, assumed that the sample set is X, where the labeled sample set is X_L and the unlabeled sample set is X_U, $X = \{X_L, X_U\}$, and the category label is $C = \{1, 2, \ldots, c\}$. The graph-based SSL method can make full use of labeled sample information and unlabeled sample information, it also can characterize the internal structural relationship of data through a large number of unlabeled information, thereby the diagnostic performance and generalization can be enhanced. Thus, the objective function of SSL can be expressed

$$\min \sum_{i}^{L} (f(x_i) - y)^2 + \sum_{i,j}^{L+U} w_{i,j}(f(x_i) - f(x_j)) \quad (1)$$

where $f(x)$ is output mapping, y is output target, and w is weight. Where the L is the number of the labeled samples, and the U is the number of unlabeled samples. The $w_{i,j}$ is the connection weight. And the first term is the reconstruction function of the optimized goal. The second term is the regularization, which can ensure that the label information between adjacent samples is as similar as possible. It is used to mine the geometric information of the data. The key part of graph-based SSL involves graph construction and the constraints of the goal optimization.

B. Semisupervised Graph Convolution Deep Belief Network (SSGCDBN)

To improve the generalization performance of fault diagnosis, it is necessary to enhance the ability to extracting the local and discriminant information of the diagnostic model. First, a new type of adaptive local graph structure is constructed by

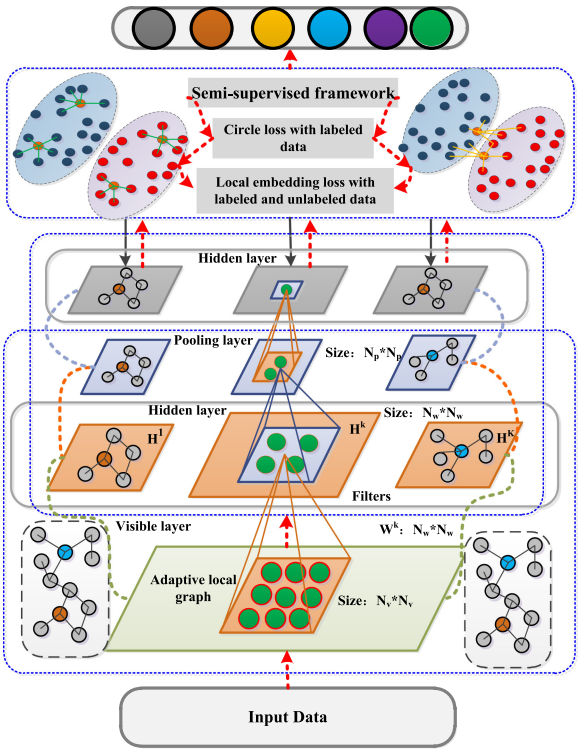


Fig. 2. Structure illustration of the proposed SSGCDBN algorithm.

using the labeled samples and unlabeled samples, then labeled samples are employed to construct discriminative circles loss to obtain the structure information of data. Finally, the semisupervised framework is introduced into CDBN with circle loss and adaptive local graph, the SSGDBNN algorithm is formed. The schematic diagram of the designed SSGCDBN algorithm is shown in Fig. 2. Specifically, the construction process of the SSGCDBN algorithm is displayed as follows.

1) Adaptive Local Graph Learning With Labeled and Unlabeled Samples: The specific expression of the constructed adaptive local graph learning algorithm is described as follows: Defined a nonnegative graph $G(V, E, S)$ containing a similar matrix $S(S_{ij} = W_{ij})$, and $L_S = D - S$ as the Laplacian operator of S can be defined, where D is a dual of diagonal elements as the sum of S rows matrix, the multiplicity c of Laplacian matrix L_S with eigenvalue 0 equals the number of connected components of S. Therefore, the rank constraint $Rank\ (L_S) = N - c$ can be implemented to learn the connected graph structure. Specifically, local adaptive graph learning performs tasks such as low-dimensional embedding or classification to solve the following problems and constraints:

$$\min \sum_{i,j=1}^{N} \|x_i - x_j\|^2 s_{ij} + \varphi_2 s_{ij}^2$$
$$\text{s.t. } S_{ij} \geq 0, I^T s_i = 1, \text{rank}(L_s) = N - c \tag{2}$$

where the φ_2 is adjustment factor, I is the identity matrix, N is the number of samples. The objective of the designed graph is employed to make the close samples to each other assigned a smaller weight, and the category information (i.e., category number c) can be merged into the rank constraint. Thus, a new measure is needed to quantify, which is calculated as the normalized pairwise distance. The normalized paired distances in the original feature space and the normalized paired distances d_{ij} and p_{ij} in the embedded space are defined as

$$d_{ij} = \|x_i - x_j\|_2^2; p_{ij} = \|f_i - f_j\|_2^2. \tag{3}$$

Where f_i and f_j is the Is the mapping of x_i and x_j, respectively. At the same time, the normalized paired distances in the original feature space and the embedded space are defined as

$$\bar{d} = \sqrt{\frac{d_{ij}}{\|d_i\|_2} \cdot \frac{d_{ij}}{\|d_j\|_2}}; \bar{p} = \sqrt{\frac{f_{ij}}{\|f_i\|_2} \cdot \frac{f_{ij}}{\|f_j\|_2}}. \tag{4}$$

The $d_i = [d_{i1}, d_{i2}, \ldots, d_{in}]$ represents the distance between the ith data point and the entire population in the original space, $p_i = [p_{i1}, p_{i2}, \ldots, p_{in}]$ is a distance vector in the embedded space. In summary, the optimized objective function of the adaptive local graph can be replaced by the normalized paired distance of the original feature space and the mapping space

$$\min \sum_{i,j=1}^{N} \bar{d}_{ij} \bar{p}_{ij} s_{ij} + \varphi_2 s_{ij}^2$$
$$\text{s.t. } s_{ij} \geq 0, I^T s_i = 1, \text{rank}(L_s) = N - c \tag{5}$$

where L_s is the Laplacian graph. $\bar{d}_{ij}$ and $\bar{p}_{ij}$ are the normalized distance.

2) Discriminant Reconstruction Objective Function With Labeled Samples: Before introducing the discriminant objective function, a brief introduction to the original CDBN principle is required [20], [21]. The standardized CDBN is a DNN composed of multiple restricted boltzmann machines (CRBM) models. Generally, CRBM consists of two parts: Visible layer V and hidden layer H. Assuming that the visible layer is composed of L channels, each channel is composed of $N_V \times N_V$ real number units. The hidden layer is composed of K groups, and each group is composed of $N_H \times N_H$ hidden units $H_1, H_2, \ldots, H_K$, each H_k is also called a feature map. Each of the K hidden groups are connected to an $N_W \times N_W$ matrix called a filter ($N_W = N_V - N_H + 1, W_1, W_2, \ldots, W_K$). Similar to the standardized RBM, the energy function of the CRBM model is defined as follows:

$$\begin{aligned} &-\log P(\mathbf{v}, \mathbf{h}) \propto E(\mathbf{v}, \mathbf{h}) \\ &= -\sum_{k=1}^{K} \sum_{i,j=1}^{N_H} \sum_{r,s=1}^{N_W} h_{ij}^k W_{rs}^k v_{i+r-1,j+s-1} \\ &\quad -\sum_{k=1}^{K} b_k \sum_{i,j=1}^{N_H} h^k{}_{ij} - c \sum_{i.j=1}^{N_V} v_{ij} \end{aligned} \tag{6}$$

where v and h represent visible and hidden vectors, respectively. And $v_{i,j}$ are the elements in the ith row and jth column of the matrix. W_{rs}^k are the elements in the kth filter in row r and column

s, where N_H and N_V represent the number of hidden and visible layer neurons, respectively. Besides, each hidden group has a deviation b_k, and the visible cells have a shared single deviation c. Therefore, the conditional probability of standard CRBM can be calculated by one-step Gibbs sampling [17] as follows:

$$p(h_{i,j}^k = 1 \,|\mathbf{v}) = \sigma((W^k * v) + b_k) \tag{7}$$

$$p(v_{i,j} = 1 \,|\mathbf{h}) = \sigma((W^k * h)_i + c) \tag{8}$$

where σ is the activation function, and * indicates the convolution operation. To make the training speed of CDBN faster, autoencoder pretraining skills are introduced, and the weights after autoencoder training is trained as the convolution kernel of CDBN, which can improve the generalization ability of DL models. Assuming that there are N training samples, v is used to represent the sample input, and $\vec{v}$ represents the expected output of the output layer, which can be expressed as:$\{(v^{(1)}, \vec{v}^{(1)}), \ldots, (v^{(d)}, \vec{v}^{(d)}), \ldots, (v^{(N)}, \vec{v}^{(N)})\}$. For n samples, the loss function of CDBN that introduces the autoencoder principle can be improved and defined as

$$J(W, b; v, \vec{v}) = \sum \frac{1}{2} \|\vec{v} - h_{W,b}(v)\|^2 \tag{9}$$

where $h_{W,b}$ is the hidden layer value.

To extract the discriminative geometric structure information of the learned features, circles loss is introduced to learn the intraclass and interclass structural information of the fault data. The intraclass (Sn) and interclass (Sp) similarity divergence matrix is used to characterize the data. On the basis of circle loss [31], the divergence of embedded features within a class is minimized, and the divergence between classes is also maximized, thus the purpose of being closer to the same class and farther away from the class after feature extraction is achieved. Specifically, by further generalizing the optimization goal of *Sn-Sp*, circle loss obtains a more flexible optimization approach and a clearer convergence goal, thereby the ability to discriminate the learned features is improved. The traditional deep feature learning methods are to minimize the $(Sn - Sp)$, and the ideal state is ($Sn = 0$, $Sp = 1$). For this reason, Megvii Inc. [31] only made a very simple change, generalizing $(Sn - Sp)$ to $(\alpha_n Sn - \alpha_p Sp)$, α_n and α_p are independent weighting factors, linear with *Sn* and *Sp* relatedly, this not only allows *Sn* and *Sp* to learn at different pace but also adjusts the amplitude to more similar scores. Such an optimization strategy makes $(Sn - Sp = m)$ appear as a circle, so it is called circle loss function. The unified discriminant loss function can be summarized as

$$\begin{aligned} L_{\text{unit}D} &= \log\left[1 + \sum_{i=1}^{K}\sum_{j=1}^{L} \exp(\gamma(s_n^j - s_p^i + m))\right] \\ &= \log\left[1 + \sum_{j=1}^{L} \exp(\gamma(s_n^j + m)) \sum_{i=1}^{K} \exp(\gamma(-s_p^i))\right] \end{aligned} \tag{10}$$

where the similarity between and within the category is $\{s_n^j\}(j = 1, 2, \ldots, L)$ and $\{s_p^i\}(i = 1, 2, \ldots, K)$. Hierarchically, the similarity and weight vector w_i of various types is calculated to obtain the $N - 1$ interclass similarity and the single intraclass similarity

$$s_n^j = \frac{W_j^T x}{(\|w_j\| \, \|x\|)}; s_p = \frac{W_y^T x}{(\|w_y\| \, \|x\|)}. \tag{11}$$

If the residual term m in the above equation is reduced, *Sn* and *Sp* are weighted, the circle loss can be obtained as follows:

$$\begin{aligned} L_{\text{circle}} = \log\Bigg(1 + \sum_{j=1}^{L} \exp(\gamma a_n^j (s_n^j - \Delta_n)) \\ \times \sum_{i=1}^{K} \exp(-\gamma a_p^i (s_p^i - \Delta_p))\Bigg). \end{aligned} \tag{12}$$

According to the decision boundary and [31], the optimization goal of circle Loss is $s_n \to 0, s_p \to 1$. The parameter *m* controls the radius of the decision boundary, which can be regarded as a relaxation factor. In other words, circle loss expects $s_n < m, S_p > 1 - m$. Therefore, there are only two hyperparameters, namely, expansion factor γ and relaxation factor m.

3) Local and Discriminative Objective Functions With Semisupervised Framework: The construction of the adaptive local graph is benefited to extract the local feature information of data. The circle loss can maintain the discriminative performance of data by flexibly optimizing the similarity and intraclass similarity. However, the information of labeled samples in real industries is limited. Massive unlabeled samples are essential to mining useful geometric structure information to improve the generalization performance of the diagnostic model.

For the graph embedding process, the geometric structure of the original feature space of data can be maintained. Specifically, implementing level constraint level $\text{Rank}(L_S) = N - c$ is equivalent to requiring that adjacent samples be constrained in the mapping space. The constraints are described as follows:

$$\sigma_1(L_s) = \ldots = \sigma_c(L_s) = 0, \sigma_{c+1}(L_s) > 0 \tag{13}$$

where σ_i represents the ith smallest eigenvalue of L_s, so $\sum_{i=1}^{c} \sigma_i(Ls)$ is needed to optimize the minimization as follows:

$$\sum_{i=1}^{c} \sigma_i(L_s) = \min_{\mathrm{F} \in R, FF^T = I} tr\left[FL_sF^T\right] = \min \sum_f (f_i - f_j) w_{ij} \tag{14}$$

where $tr[FL_sF^T]$ represents the trace of the matrix. Under the framework of SSL, combined with adaptive local graph learning and discriminant reconstruction objective function, the overall objective function of SSGCDBN can be obtained as follows:

$$\begin{aligned} J_{\text{total}} = \min_{w,f} \frac{1}{l} L_{\text{circleloss}} + \frac{1}{l+u} \Bigg[\varphi_1 \sum_{i,j}^{l+u} \left\|\bar{d}_{ij} \bar{p}_{ij} w_{ij}\right\| \\ + \varphi_2 {w^2}_{ij} + 2\varphi_3 tr(FLsF^T)\Bigg] \\ \text{s.t. } \forall i, w_i^T \mathbf{1} = 1, 0 \le w_i \le 1, \text{rank}(L_s) = n - c \end{aligned} \tag{15}$$

where $[\varphi_1, \varphi_2, \varphi_3]$ are the parameter adjustment coefficients. The graph learned through local and discriminant learning is

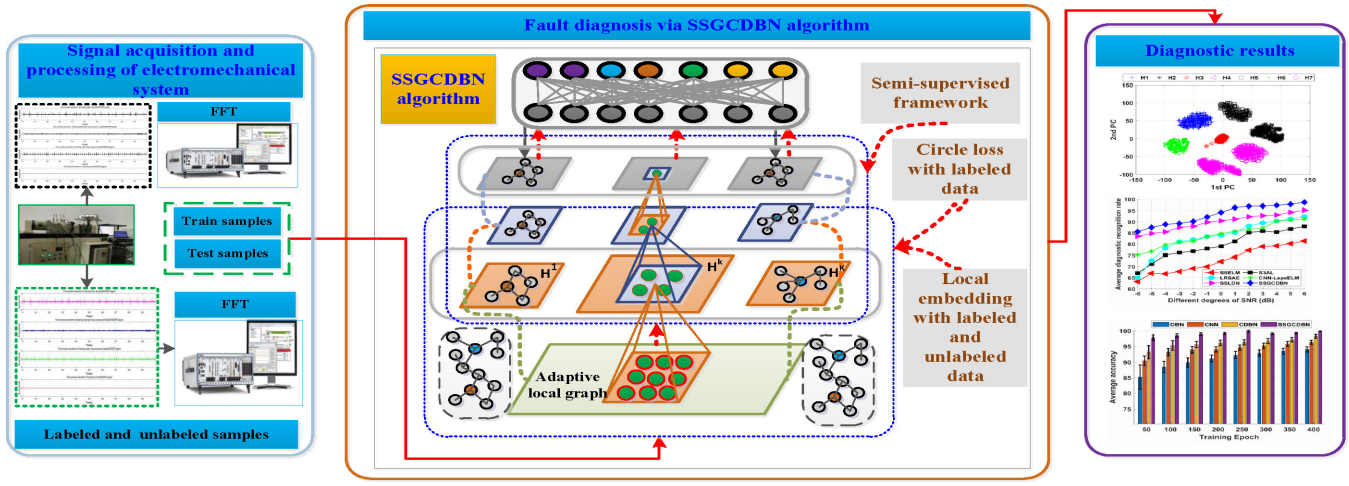


Fig. 3. Framework of the proposed method.

Algorithm 1: SSGCDBN

Input: $G = (V; E; S)$, parameters and Epoch; Batch ;
Output: Y and $(W, b) = Net\_Initialize$;

```
for e = 1 to Number of Epochs do
    for b_n = 1 to Batch Size // Forward propagation; do
        Calculated adaptive local graph by Eq. 5
        Reconstruct discriminant objective by Eq. 12
        Calculate total objective function by Eq. 15
        for k = 1 to K do
            Y = sigm(Y^(K-1) β^(K) + b^(K)) //k−layer
            L^k(θ*_f) = min_θ(l^k_L + l^k_{U+L}) // target
        end
        Optimized Eq.15 via BP
        (W, b) = Net_Update(loss, W, b) //Update
    end
end
return the trained network Model(·) = θ*(w, b)
```

more structured, and there are fewer false connections. The algorithm step designed SSGCDBN is described as Algorithm 1.

4) Algorithm Optimization: The model optimization is achieved by simultaneously minimizing two loss functions under the SSL framework: 1) Supervised circle loss function, that is, the error between the true label of the labeled sample and the classification probability (predicted label); 2) the labeled and unlabeled objective function. Therefore, the designed model can be optimized as

$$J_{\text{total}}(\theta_f^*) = \min_{\theta}(J_L + J_{(U+\text{L})}) \tag{16}$$

where the mapping output is $f(.)$. To optimize the objective function of the abovementioned hybrid embedded network, the partial derivative should be equal to zero. Specifically, the mathematical form of the partial derivative is rewritten as

$$\frac{\partial J_{\text{total}}}{\partial \theta^{(k)}} = \frac{\partial J_L}{\partial \theta^{(k)}} + \frac{\partial J_{(L+U)}}{\partial \theta^{(k)}} = 0, k = 1, 2, \ldots, K \tag{17}$$

where k is the layer index. The gradient descent is employed to update the weight and bias parameters

$$(W^{(k)}, b^{(k)}) \leftarrow (W^{(k)}, b^{(k)}) - \lambda \cdot \frac{\partial J_{\text{total}}}{\partial (W^{(k)}, b^{(k)})} \tag{18}$$

where λ is learning rate.

C. Intelligent Fault Diagnosis Framework Based on SSGDCBN Algorithm

To improve the generalization performance of intelligent fault diagnosis, this article has a comprehensive utilization of labeled information and unlabeled information, a new semisupervised GNN (SSGCDBN algorithm) is proposed. On this basis, a fault diagnosis method of electromechanical system based on SSGDBDN algorithm is proposed in this article. Specifically, the process framework of fault diagnosis method based on SSGCDBN algorithm is shown in Fig. 3, its specific diagnostic steps can be described as follows: 1) Collecting unlabeled and labeled vibration signals from the core components of the electromechanical system; 2) converting the collected vibration signals into spectrum signals through fast fourier transform (FFT) transformation, which is divided into training samples and test samples; 3) the relevant parameters of the SSGCDBN model are initialized; 4) both labeled training samples and unlabeled training samples are used for the SSGCDBN model training; 5) the trained SSGCDBN model is applied to identify testing samples; 6) output the fault diagnosis results and evaluates the performance of the proposed method.

IV. Experimental Studies and Analysis

To evaluate the feasibility of the proposed method, the bearing fault data simulated on the motor-bearing system (accelerated bearing life tester, ABLT-1 A) is used to verify and compare with other semisupervised fault diagnosis methods. All tests in this article were averaged 10 times to eliminate random errors during training.

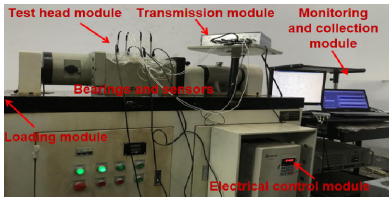


Fig. 4. Overview of the motor-bearing system (ABLT-1 A).

TABLE I
TECHNICAL INDICATORS OF THE MOTOR-BEARING SYSTEM

Parameter/device	Value /type
Speed	1050 r/min
Sampling frequency	10240 Hz
Sensor	PCB vibration acceleration sensor
Data acquisition system	NI 9234
Bearing	HRB6205 deep groove ball bearings
Number of test bearings	8 sets
Loading	0kN

A. Data Acquisition and Processing

The ABLT-1 A used in this section is suitable for bearing fatigue life testing and fault simulation experiments. The overview of this tester is displayed in Fig. 4. Specifically, under the condition of zero loads, there are seven types of health conditions including the compound weak fault of the inner and outer rings (H1), the compound weak fault of the outer ring and ball (H2), the compound fault of the inner and outer rings (H3), the inner ring fault (H4), compound fault of ball and out ring (H5), outer ring fault (H6), and normal (H7) were simulated, respectively. At the same time, the collected signals are intercepted into a set of samples built on 1024 signal points, and the 1024 signal points are intercepted into one sample group. 1000 groups of samples are intercepted for various health conditions, a total of 1000*7=7000 samples are collected. Thus, the constructed sample groups were randomly divided into the training sample subset and test sample subset. Some specific experimental environment parameters of this experiment are displayed in Table I.

B. Implementation Details and Result

It is necessary to set and explain the experimental environment and parameters in advance. First, 50% of the dataset is randomly divided as the training sample subset, and the remainder 50% is set as a testing sample subset. Among them, the random 20% of the training samples are labeled samples, and the rest are unlabeled samples. The designed SSGCDBN network model is mainly composed of the following parameters: The parameters of the original CDBN and the graphic module is embedded with some key parameters, such as the two main parameters of circle loss $[m, \gamma]$ [31] and the adjustment parameters $[\varphi_1, \varphi_2, \varphi_3]$. According to [17], [18], the convolution kernel of the hidden layer and the visible layer can be set to [1-9-16]. Correspondingly, two layers of SSGCRBM is set up. According to the aforementioned fault diagnosis flowchart, the adjustment coefficient can be obtained by the grid search method. Fig. 5 indicates the influence of the regularization adjustment coefficient on the fault diagnosis

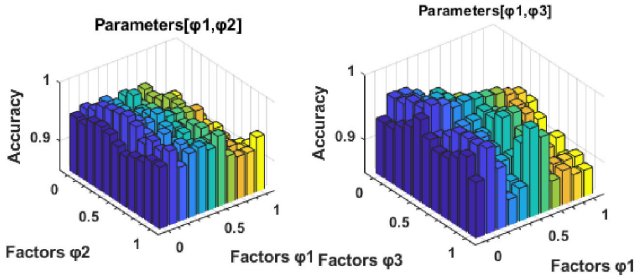


Fig. 5. Different regularization adjustment coefficient for fault diagnosis.

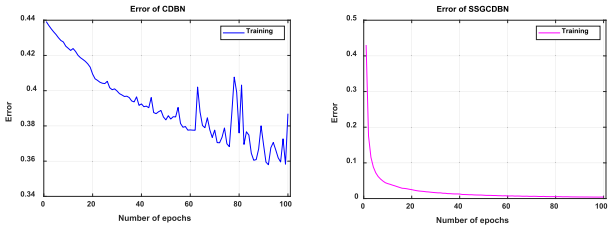
TABLE II
MAIN PARAMETERS SETTING UP OF SSGCDBN ALGORITHM

Parameter name	value	Parameter name	value
Sparsity	0.01	Layer2.n_map.v	9
Number of hidden layer	2	Layer2.n_map.h	16
Learning rate	0.01	Layer2.fliter	[5 5]
Number of iteration	200	Layer2.pool	[2 2]
Batch	20	Regularization	0.01
Layer1.n_map.v	1	Layer1.pool	[2 2]
Layer1.n_map.h	9	Stride	[1 1]
Layer1.fliter	[7 7]	$[\varphi_1, \varphi_2, \varphi_3]$	[0.1, 0.1, 0.3]
Input type	Binary	$[m, \gamma]$	[256, 0.25]

of SSGCDBN. The regularization adjustment coefficient is more appropriate to take the nearby combination [0.1, 0.1, 0.3].

In this experiment, the parameters and diagnosis results are evaluated by using the fivefold nested cross-validation (NCV) technology [2], [19] to select hyperparameters and estimate the performance of the selected model. External cross-validation in NCV is used to evaluate model performance, and internal cross-validation is used to select hyperparameters. First, the samples are evenly divided into five subsets. Then, the outer loop is repeated five times, and onefold is retained for testing each time, and the rest is regarded as the training set. The training set will estimate the parameters through the internal loop. To this end, the training set needs to undergo another cross-validation process, in which onefold samples are reserved for verification, and the other threefold samples constitute the internal training set. In this process, each iteration of the outer loop will be further repeated five times. Through the use of NCV technology, the average diagnostic accuracy of different samples and the deviation of testing samples can be obtained, and the influence of factors such as random errors can be reduced. In summary, the main parameter settings of the designed SSGCDBN algorithm are shown in Table II. This experiment is performed on a PC with Intel Core i7-9750H CPU 2.6 Hz, RAM: 8 GB, and Inter(R) UHD Graphics 630 GPU. Its operating system is the Windows2010, Matlab20018b, and TensorFlow platform with python 3.6 software were used to calculate and accelerate.

Meanwhile, Fig. 6 demonstrates the training error curve of the benchmark CDBN algorithm and SSGCDBN algorithm. As described in Fig. 6 that the convergence speed of the training error reconstruction curve of the SSGCDBN algorithm is much faster than the CDBN algorithm. After 100 time training, the training error of SSGCDBN drops early to below 0.1, and its convergence speed is more rapid. This shows that the SSGCDBN algorithm has better input and output fitting capabilities. From



(a) (b)

Fig. 6. Training error curve of: (a) CDBN. (b) SSGCDBN.

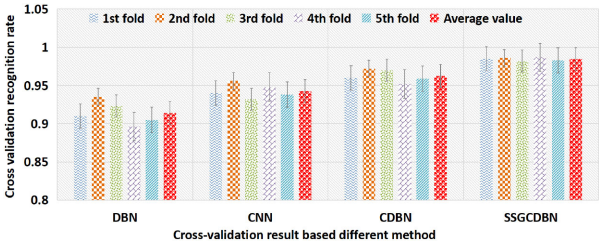


Fig. 7. Testing sample results of cross-validation based on different diagnostic methods.

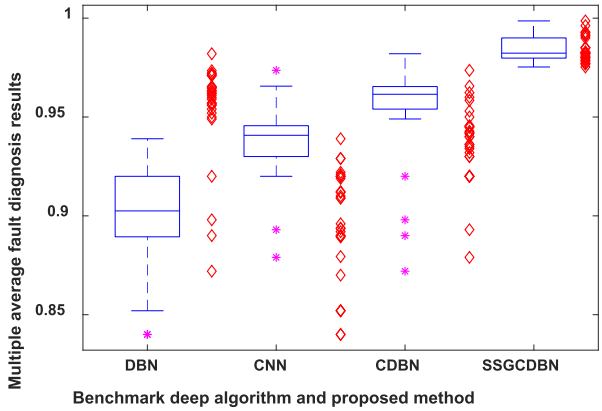


Fig. 8. Box plot of the average results of different fault diagnosis methods.

TABLE III
STATISTIC INDICATORS IN COMPARISON WITH OTHER ALGORITHMS

Methods	Sb	Sw	Sb/SW(J)
DBN	14.296	12.517	1.1421
CNN	168.59	58..35	2.889
CDBN	218.56	28.921	7.557
SSGCDBN	333.09	13.092	26.4423

the perspective of information theory, As explained that the smaller the training error, the less the loss of information, and the stronger the learning ability of the model.

To validate the stability of the designed fault diagnosis algorithms and methods, compared with standard benchmark DL algorithms such as DBN, CNN, CDBN, and other algorithms can be selected, and its data processing and parameter settings are roughly the same with the proposed SSGCDBN algorithm. The number of hidden layers of the network is two, the last layer inputs are fully connected to the BPNN layer. The parameters such as learning rate and batch are also the same with the SSGCDBN algorithm. Simultaneously, the abovementioned four methods are validated by using the fivefold cross-validation, the average result of testing samples is displayed in Fig. 7 . From Fig. 7, the stable performance and diagnostic performance of SSGCDBN are the best. To demonstrate more diagnosis results, 30-time average diagnosis and experimental results were performed on the 6205 dataset. Fig. 8 indicates the box-plot and scatter distribution diagram of the 30 average diagnosis results of the testing samples diagnosed by different fault diagnosis methods. As illustrated in Fig. 8 that the diagnostic performance of the proposed SSGCDBN algorithm is more stable, its diagnostic accuracy is also the highest.

To visualize the output features of the training network, t-distributed stochastic neighbor embedding (t-SNE) technique [12], [32] is used to simplify the high-dimensional output features of the last hidden layer extracted by the above four methods into two-dimensional vector distribution, the visualization results of the testing samples are shown in Fig. 9. As indicated that only the low-dimensional features extracted by the SSGCDBN algorithm can completely separate all faults.

Fig. 10 also displays the average diagnosis results of testing samples via the aforementioned four algorithms in different training periods. As illustrated from Fig. 10 that it can achieve higher diagnostic results in early training periods. At the same time, a large number of unlabeled samples to supplement the deficiencies of labeled samples can be employed via the proposed SSGCDBN, so that the generalization and stability of the models diagnosis are best reflected.

To quantify the extracted fault features, the features extracted from the last layer of the above four methods can be quantitatively evaluated and analyzed. In this case, the indicators based on the dispersion matrix (i.e., intraclass dispersion matrix Sw and interclass dispersion matrix Sb) [33] are introduced to evaluate the fault feature quantitatively and diagnosis performance of SSGCDBN. The intraclass dispersion matrix Sw reflects the clustering degree of each category of the recorded sample, and the interclass dispersion matrix Sb represents the separation ability of different categories. For the output characteristics of the trained classifier, the intraclass dispersion matrix Sw should be as small as possible, and the interclass dispersion matrix Sb should be skewed to a larger value. Table III gives the three average indicators in these 10 trials of statistical data. The quantitative analysis results indicate that the proposed method has a favorable classification ability.

C. Comparison With the State of the Arts

In this subexperiment, the ratio of training samples to test samples to 40%: 60% can be set. Among them, for the training samples, the number of labeled training samples for each type is 40 (that is, 10%, which means that in the case of extremely limited category labels), and the remaining samples are unlabeled training samples. Specifically, the selected related semisupervised fault diagnosis method has similar experimental settings as the method proposed in this article. The details of the relevant comparison methods are selected as SSELM [23],

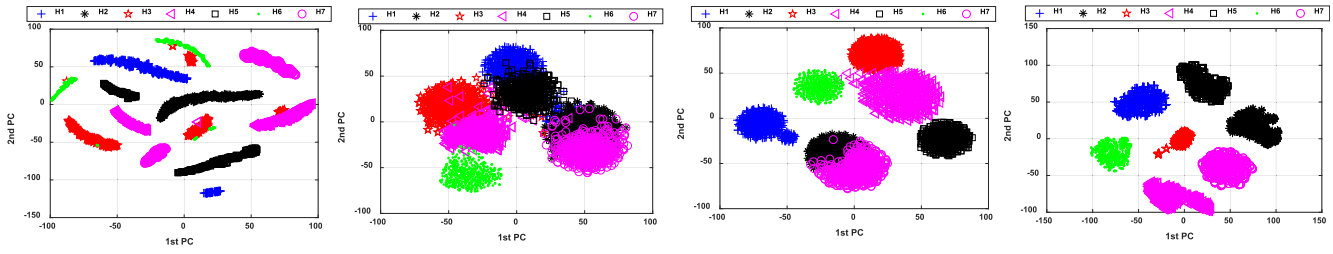


Fig. 9. 2-D visualization result of learned features via: (a) DBN. (b) CNN. (c) CDBN. (d) SSGCDBN.

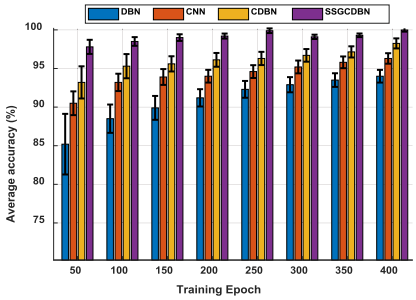


Fig. 10. Comparison of the diagnostic results obtained via different DL algorithm.

TABLE IV
MAIN PARAMETER SETTING OF RELATED SSL ALGORITHMS

SSL algorithms	Main description and setting up
SSELM	"2 trade-off coefficients, [C = 105; labada = 107]"
S3AL	Number of neighbors
LRSAE	"[1024-200-100-1024-100-5]; Sparsity = 0.1 iterations = 200, regularization =0.1; BPNN."
CNN-LapsELM	"input=32*32. C1=6, C2=12, learning 0.1, 2 trade-off coefficients, [C = 105; labada = 107], penalty = 0.1"
SSLDN	"Learning rate 0.1, [1024-200-100],LSTM iterations = 200."

TABLE V
COMPARISON RESULT OF THE DIFFERENT DIAGNOSTIC INDICATOR

Algorithm	Pre(%)	Rec(%)	F-score(%)	Test Time(s)	Train Time(s)
SSELM	82.8±6.74	81.2± 7.723	83.72± 5.96	45.14	64.52
S3AL	89.37±7.32	87.6± 7.85	91.54± 4.367	68.32	96.37
LRSAE	94.2±4.32	93.20± 5.351	95.36± 3.75	234.71	312.63
CNN-LapsELM	90.11±1.475	93.2± 0.965	92.8± 1.12	326.39	421.71
SSLDN	96.4±0.728	94.9± 10.8234	97.8± 0.782	312.57	392.82
SSGCDBN	97.9±0.632	98.12± 0.612	98.92± 0.5932	297.68	361.45

S3AL [19], LRSAE [24], CNN-LapsELM [25], SLNDNN [21], and the proposed method (SSGCDBN). The main parameters of the proposed algorithm and other related SSL algorithms in this subsection are shown in Table IV.

Considering the influence of different labeled samples on the diagnostic performance, the authors in [19], [21] set the labeled sample ratio of the training sample set, which is denoted as $\psi = |L|\backslash(|L+U|)$. Different training labeled sample ratios (i.e., $\psi = 0.1, 0.2, 0.4, 0.6$) is set up. The average diagnostic accuracy based on different SSL algorithm is obtained in Tables V and VI. Combining Tables V and VI, the proposed SSL fault diagnosis method has certain competitiveness compared with other methods. SSGCDBN has a strong generalization ability and stable performance.

To increase the difficulty of the diagnosis tasks, Gaussian white noise with different degrees of SNR can be added to the original data. Therefore, the additional Gaussian noise was

TABLE VI
AVERAGE ACCURACY AND STANDARD DEVIATION AT DIFFERENT LABELED TRAINING SAMPLE RATIOS(%)

Algorithms	0.1	0.2	0.4	0.6	Average
SSELM	81.2±7.673	83.3±6.127	85.4±6.231	88.2±4.456	84.525±6.121
S3AL	88.7±8.320	91.52±7.35	94.45±5.330	96.50±4.932	92.79±6.483
LRSAE	92.36±6.512	93.1±5.012	95.63±4.851	97.23±4.367	94.58±5.1855
CNN-LapsELM	91.89±1.352	93.7±1.105	95.39±0.9287	97.12±0.7635	94.52±1.037
SSLDN	95.6±0.857	96.3±0.8421	97.62±0.532	98.49±0.457	97.0±0.6720
SSGCDBN	98.66±0.568	99.16±0.320	99.23±0.2140	99.82±0.2030	99.21±0.3262

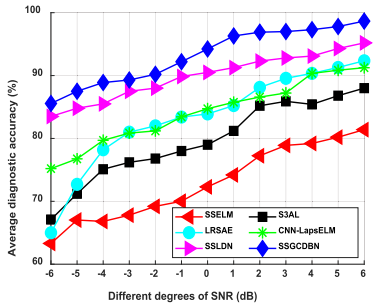


Fig. 11. Fault diagnosis results under different signal-to-noise ratios (SNRs).

added to the original test signal to validate the proposed method's robustness. Specifically, the signal-to-noise ratio (SNR is defined as SNR(dB) = 10log10 (Psignal/Pnoise), where Psignal and Pnoise represent the power of the original signal and additional Gaussian noise. In this subsection, the range of −6 to 6 dB (SNR) noise is employed to evaluate the proposed method. Fig. 11 shows the diagnosis results of testing samples at different SNRs (L/(L+U)=0.1). Under different SNR cases, the proposed method obtains better diagnostic results than other comparison methods.

D. Discussion

It is difficult to obtain the labeled sample information in the real industries. Therefore, this article proposes a new SSL-fault diagnosis method for electromechanical systems based on SSGDBNN algorithm. Its effectiveness is validated with the vibration signal of motor-bearing system. Compared with other SSL and DL methods, this method has been verified in feature extraction ability, classification performance, and generalization performance. From Tables V and VI, the proposed SSGCDBN algorithm and the latest SSDDLN have achieved good diagnostic results to a certain extent. From Figs. 9–11, the designed

fault diagnosis method can effectively capture the relationship between noise and actual data distribution, the proposed method is further verified.

V. Conclusion

Based on the principle of graph embedding and SSL, SSGCDBNN was proposed for accurate intelligent fault diagnosis. Specifically, the designed SSGCDBN algorithm can learn the local and discriminant information of fault data through strategies such as local adaptive graph learning and circle loss. Under the SSL framework, the proposed method can make full use of labeled and unlabeled sample information. The analysis of the experimental data shows that the proposed method also has a strong recognition ability when the labeled training samples are extremely limited.

In addition, the future research work and limitations of this article can be summarized in the following three aspects. 1) To optimize the key parameters of the designed DL model, we only used the cross-validation method to make simple and rough estimates, and further exploration and intelligent optimization are still needed, methods such as particle swarm optimization (PSO) and other algorithms were employed to optimize the model. 2) The actual industrial field data were noisier and more complex, which was only simulated by adding noise, and the further validation of industrial data is still needed. 3) The category distribution of real industrial data was often unbalanced, namely, the number of normal data was far more than the amount of faulty data, but it is assumed that the number of categories samples are equal in this article.

References

[1] L. Qiao, L. Zhang, S. Chen, and D. Shen, "Data-driven graph construction and graph learning: A review," *Neurocomputing*, vol. 312, pp. 336–351, 2018.

[2] X. Zhao, M. Jia, P. Ding, C. Yang, D. She, and Z. Liu, "Intelligent fault diagnosis of multi-channel motor-rotor system based on multi-manifold deep extreme learning machine," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 5, pp. 2177–2187, Oct. 2020.

[3] 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.

[4] 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.*, vol. 69, no. 4, pp. 1505–1518, Apr. 2020.

[5] S. Haidong, C. Junsheng, J. Hongkai, Y. Yu, and W. Zhantao, "Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for early fault prognosis of bearing," *Knowl.-Based Syst.*, vol. 188, 2020, Art. no. 105022.

[6] Z. He, H. Shao, P. Wang, J. J. Lin, J. Cheng, and Y. Yang, "Deep transfer multi-wavelet auto-encoder for intelligent fault diagnosis of gearbox with few target training samples," *Knowl.-Based Syst.*, vol. 191, p. 105313, 2020.

[7] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep residual shrinkage networks for fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4681–4690, Jul. 2020.

[8] 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, 2018.

[9] M. Xia, M. Xia, G. Han, Y. Zhang, and J. Wan, "Intelligent fault diagnosis of rotor-bearing system under varying working conditions with modified transfer CNN and thermal images," *IEEE Trans. Ind. Informat.*, to be published, doi: 10.1109/TII.2020.3005965.

[10] X. Yuan, C. Ou, and Y. Wang, "Development of NVW-SAES with nonlinear correlation metrics for quality-relevant feature learning in process data modeling," *Meas. Sci. Technol.*, vol. 32, p. 015006, 2020.

[11] K. Yu, T. R. Lin, and J. Tan, "A bearing fault and severity diagnostic technique using adaptive deep belief networks and Dempster–Shafer theory," *Struct. Health Monit.*, vol. 19, no. 1, pp. 240–261, 2020.

[12] H. Zhiyi, S. Haidong, J. Lin, C. Junsheng, and Y. Yu, "Transfer fault diagnosis of bearing installed in different machines using enhanced deep auto-encoder," *Measurement*, vol. 152, 2020, Art. no. 107393.

[13] X. Yuan, J. Zhou, B. Huang, Y. Wang, C. Yang, and W. Gui, "Hierarchical quality-relevant feature representation for soft sensor modeling: a novel deep learning strategy," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 3721–3730, Jun. 2020.

[14] X. Zhao, M. Jia, and M. Lin, "Deep Laplacian auto-encoder and its application into imbalanced fault diagnosis of rotating machinery," *Measurement*, vol. 152, 2020, Art. no. 107320.

[15] X. Yuan, L. Li, and Y. Wang, "Nonlinear dynamic soft sensor modeling with supervised long short-term memory network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3168–3176, May 2020.

[16] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, Jul. 2017.

[17] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 609–616.

[18] 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, 2019.

[19] R. Razavi-Far, E. Hallaji, M. Farajzadeh-Zanjani, and M. Saif, "A semi-supervised diagnostic framework based on the surface estimation of faulty distributions," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1277–1286, Mar. 2019.

[20] Y. Gao, L. Gao, X. Li, and X. Yan, "A semi-supervised convolutional neural network-based method for steel surface defect recognition," *Robot. Comput.-Integr. Manuf.*, vol. 61, 2020, Art. no. 101825.

[21] R. Razavi-Far *et al.*, "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.

[22] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Jan. 2009.

[23] R. Razavi-Far, E. Hallaji, M. Saif, and L. Rueda, "A hybrid scheme for fault diagnosis with partially labeled sets of observations," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl.*, 2017, pp. 61–67.

[24] Z. Chai, W. Song, H. Wang, and F. Liu, "A semi-supervised auto-encoder using label and sparse regularizations for classification," *Appl. Soft Comput.*, vol. 77, pp. 205–217, 2019.

[25] Y. Lei, X. Chen, M. Min, and Y. Xie, "A semi-supervised Laplacian extreme learning machine and feature fusion with CNN for industrial superheat identification," *Neurocomputing*, vol. 381, pp. 186–195, 2020.

[26] Y. Zhao, R. Ball, J. Mosesian, J.-F. de Palma, and B. Lehman, "Graph-based semi-supervised learning for fault detection and classification in solar photovoltaic arrays," *IEEE Trans. Power Electron.*, vol. 30, no. 5, pp. 2848–2858, May 2015.

[27] D. Zhang, E. Stewart, M. Entezami, C. Roberts, and D. Yu, "Intelligent acoustic-based fault diagnosis of roller bearings using a deep graph convolutional network," *Measurement*, vol. 156, 2020, Art. no. 107585.

[28] F. Nie, X. Wang, and H. Huang, "Clustering and projected clustering with adaptive neighbors," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2014, pp. 977–986.

[29] J. Wen, X. Fang, Y. Xu, C. Tian, and L. Fei, "Low-rank representation with adaptive graph regularization," *Neural Networks*, vol. 108, pp. 83–96, 2018.

[30] Q. Wang, Z. Qin, F. Nie, and X. Li, "Spectral embedded adaptive neighbors clustering," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 30, no. 4, pp. 1265–1271, Apr. 2019.

[31] Y. Sun, C. Cheng, Y. Zhang, C. Zhang, L. Zheng, Z. Wang, and Y. Wei, "Circle loss: A unified perspective of pair similarity optimization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 6398–6407.

[32] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[33] 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, 2015.

Xiaoli Zhao (Student Member, IEEE) received the M.S. degree from the LanZhou University of Technology, Lanzhou, China, in 2017. He is currently working toward the Ph.D. degree in mechanical engineering with Southeast University, Nanjing, China, and with the School of Engineering, the University of British Columbia, Okanagan, Canada.

His research interests include electromechanical equipment intelligent monitoring and fault diagnosis, rehabilitation robot, artificial intelligence, and signal processing.

Minping Jia 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, Nanjing, China, in 1991.

He is currently a Full Professor in mechanical engineering with Southeast University, Nanjing, China. 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 earth resources engineering from Kyoto University, Kyoto, Japan, in 2000, and the Ph.D. degree in electrical engineering from the University of Ottawa, Ottawa, ON, Canada, in 2007.

From 2012 to 2015, he was a Full Professor with the Toyota Technological Institute, Nagoya, Japan. He is currently Full Professor with the School of Engineering, the University of British Columbia, Vancouver, Okanagan. His research interests include image/data fusion, computer vision, pattern recognition, senor/sensor networks, condition-based maintenance, and nondestructive inspection and evaluation.

Dr. Liu is a Member of SPIE. He is chairing 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, IEEE TRANSACTIONS ON MECHATRONICS, IEEE INSTRUMENTATION AND MEASUREMENT MAGAZINE, IEEE JOURNAL OF RADIO FREQUENCY IDENTIFICATION (RFID), *Information Fusion*, *Machine Vision and Applications*, and *Intelligent Industrial Systems*.