agriculture

MDPI

*Article*

# Rolling Bearing Fault Diagnosis in Agricultural Machinery Based on Multi-Source Locally Adaptive Graph Convolution

**Fengyun Xie [1,2,*], Enguang Sun [1], Linglan Wang [1], Gan Wang [1] and Qian Xiao [1,2]**

[1] School of Mechanical and Vehicle Engineering, East China Jiaotong University, Nanchang 330013, China; sngzm999@163.com (E.S.); wlinglan2022@163.com (L.W.); wanggan813@163.com (G.W.); xiaoqian@ecjtu.jx.cn (Q.X.)

[2] State Key Laboratory of Performance Monitoring Protecting of Rail Transit Infrastructure, East China Jiaotong University, Nanchang 330013, China

* Correspondence: xiefyun@163.com

**Abstract:** Maintaining agricultural machinery is crucial for efficient mechanized farming. Specifically, diagnosing faults in rolling bearings, which are essential rotating components, is of significant importance. Domain-adaptive technology often addresses the challenge of limited labeled data from a single source domain. However, information transfer can sometimes fall short in providing adequate relevant details for supporting target diagnosis tasks, leading to poor recognition performance. This paper introduces a novel fault diagnosis model based on a multi-source locally adaptive graph convolution network to diagnose rolling bearing faults in agricultural machinery. The model initially employs an overlapping sampling method to enhance sample data. Recognizing that two-dimensional time–frequency signals possess richer spatial characteristics in neural networks, wavelet transform is used to convert time series samples into time–frequency graph samples before feeding them into the feature network. This approach constructs a sample data pair from both source and target domains. Furthermore, a feature extraction network is developed by integrating the strengths of deep residual networks and graph convolutional networks, enabling the model to better learn invariant features across domains. The locally adaptive method aids the model in more effectively aligning features from the source and target domains. The model incorporates a Softmax layer as the bearing state classifier, which is set up after the graph convolutional network layer, and outputs bearing state recognition results upon reaching a set number of iterations. The proposed method's effectiveness was validated using a bearing dataset from Jiangnan University. For three different groups of bearing fault diagnosis tasks under varying working conditions, the proposed method achieved recognition accuracies above 99%, with an improvement of 0.30%-4.33% compared to single-source domain diagnosis models. Comparative results indicate that the proposed method can effectively identify bearing states even without target domain labels, showcasing its practical engineering application value.

**Keywords:** agricultural machinery; rolling bearings; subdomain adaptive; transfer learning; graph convolutional network; fault diagnosis



**Citation:** Xie, F.; Sun, E.; Wang, L.; Wang, G.; Xiao, Q. Rolling Bearing Fault Diagnosis in Agricultural Machinery Based on Multi-Source Locally Adaptive Graph Convolution. *Agriculture* **2024**, *14*, 1333. https://doi.org/10.3390/agriculture14081333

Academic Editors: Massimo Brambilla and Maurizio Cutini

Received: 3 July 2024
Revised: 6 August 2024
Accepted: 8 August 2024
Published: 9 August 2024



## 1. Introduction

China, as a large agricultural country, feeds nearly 20% of the global population with 9% of the world's arable land; thus, with the country's high-speed development of industrial growth, agricultural mechanization is increasing on a global level [1]. However, in the face of variable load conditions and a complex working environment, the occurrence of agricultural machinery failure is inevitable. Rolling bearings are widely used in agricultural production as the core components of rotating agricultural machinery, serving the purposes of supporting the rotating body of the machinery and reducing the friction between various components [2]. According to research, bearings can lose effectiveness due to their real-world operation in the environment, their rotational speed, and other factors that cause

wear, fracture, fatigue, and other degrees of failure [3], affecting the operation of agricultural machinery and equipment safety; therefore, rolling bearings require extensive equipment testing and fault diagnosis to ensure their reliability and safety.

Bearings, which are essential components of rotating machinery, are utilized in both agricultural and transportation machinery. The primary differences between bearings for agricultural machinery and transportation vehicles revolve around the operating environment, load requirements, maintenance demands, and design considerations. Agricultural machinery bearings need to be rugged and capable of handling contamination, shock loads, and frequent maintenance [4]. In contrast, transportation vehicle bearings are designed for smoother, more controlled environments with an emphasis on precision, efficiency, and longer maintenance intervals.

Bearings in agricultural machinery often work in harsh environments, exposed to dirt, dust, mud, and moisture, and they need to have good sealing to prevent pollution; due to outdoor use, complex terrain, and seasonal changes, bearings in agricultural machinery often withstand high impact loads and vibration and temperature changes. Bearings in transport vehicles typically operate in more controlled environments such as roads or tracks, where they withstand a more stable temperature range and less impact load than agricultural machinery [5].

Due to harsh operating conditions, bearings for agricultural machinery often require frequent lubrication and maintenance to ensure longevity and performance, and they are designed to be robust and to minimize downtime during critical planting or harvesting seasons [6]. Bearings in transport vehicles typically have longer maintenance intervals compared to bearings in agricultural machinery, and these bearings are often precision-designed to provide high reliability and long-distance performance with minimal maintenance [7].

Agricultural machinery usually needs to work under heavy loads and high impact loads, so bearings must have a high load carrying capacity to cope with these extreme conditions [8]. In addition, agricultural machinery often works in dusty, muddy, and humid environments, so bearings must have good wear resistance to extend the service life; in order to prevent soil, dust, and moisture from entering into the bearings, bearings need good sealing performance. Agricultural machinery is often exposed to water, fertilizers, and pesticides during use, so bearings need to have good corrosion resistance to prevent rust and corrosion. The design of bearings should take into account the ease of maintenance to reduce maintenance time and cost. Agricultural machinery will produce a high temperature when running for a long time, so the bearing needs to have a certain high-temperature resistance to ensure that it can still work normally in a high-temperature environment. In order to improve the comfort of operation and reduce noise pollution, bearings need to have low-noise and low-vibration characteristics [9].

Currently, deep learning is extensively applied in the realm of fault diagnosis; deep learning is widely used in the field of fault diagnosis, with the advantage that there is no need to carry out manual feature extraction steps, as the automatic mining of the hidden nonlinear features in the input signal allows researchers to achieve end-to-end fault classification [10]. In real-world engineering, due to the influences of environmental noise, equipment wear and tear, changing operating conditions, and other factors, the distribution of data collected with sensors is inconsistent, while the collection of a large amount of labeled data requires high labor costs or the presence of field staff who are able to label the state type of the collected signals; therefore, the intelligent fault diagnostic method often fails to achieve satisfactory diagnostic results [11].

The impact of variable operating conditions is as follows [12]: (1) Load variation. Rolling bearings may exhibit different vibration characteristics under different load conditions. As the load increases or decreases, the bearing's vibration pattern will also change, impacting the accuracy of fault detection. (2) Speed change. The change in speed will cause the vibration frequency of the bearing to change, which may mask the fault signal, making fault diagnosis more difficult. (3) Temperature changes. Changes in temperature affect

the physical properties of the bearing material, such as expansion and contraction, which affects the vibration characteristics of the bearing.

The impact of environmental noise is as follows [13]: (1) Background vibration from other components in the mechanical system will also mix with the bearing's fault signal, making signal processing more challenging. (2) Electromagnetic interference. There is strong electromagnetic interference in some industrial environments, which can affect the performance of vibration sensors and lead to inaccurate data acquisition. (3) Environmental sound. In some noisy environments, acoustic sensors may capture a large amount of environmental noise, interfering with the extraction of fault signals.

The impact of the lack of labeled data is as follows [14]: (1) Data acquisition difficulty. Obtaining real fault data is often very difficult because the chance of equipment failure during operation is relatively small. Artificially creating faults is not only costly but can also cause irreversible damage to the equipment. (2) Data imbalance. In most cases, there are far more working data than faulty data, which can lead to data imbalance when training the model, which in turn affects the performance of the model. (3) Labeling errors. Data labeling requires expertise and is prone to errors. Incorrect labeling can lead the model to learn inaccurate information, thereby decreasing the diagnostic accuracy.

In order to obtain accurate results with the diagnostic model in the above situations, scholars introduced the migration learning method.

As a crucial area of research in transfer learning, experts and scholars address the challenges of unlabeled and sparsely labeled data by focusing on features. By measuring and increasing the similarity between source domain data and target domain data, they enhance the target domain's ability to learn from the source domain. The existing methods are based on the mechanism of antagonism and the two aspects of the statistical metrics, and combining these with deep learning algorithms achieves the purpose of bringing the two domains closer together [15]. For example, Zhang et al. [16] improved the diagnostic accuracy for rolling bearings under different operating conditions by proposing a transfer sparse auto-encoder (SAE) based on the local maximum mean difference (LMMD) and K-means. Qian et al. [17] realized fault diagnosis by improving the joint distribution adaptation to be more comprehensive to the marginal and conditional distributions of their source domain and target domain data. Cheng et al. [18] proposed an intelligent diagnostic network inspired by the Wasserstein distance; their model minimized the distribution distance between the source and target domains through adversarial training and achieved the highest diagnostic accuracy in multiple sets of migration tasks. Li et al. [19] utilized the correlation alignment (CORAL) loss to compute the similarity between the source and the target domains in order to solve the difficulties brought to bearing fault diagnosis under variable operating conditions more effectively. Wang et al. [20] proposed a subdomain-based adaptive fault diagnosis model, which takes the local maximum difference as the metric criterion and considers the fine-grained relationship of each category. The aforementioned methods can effectively address the issue of inconsistent distributions between the source and target domains. However, these models primarily focus on the geometrical structure of the data for feature extraction, overlooking the connectivity between data points.

Given that multiple sets of source domain data with varying distributions can be obtained in real-world engineering, a multi-source domain transfer learning approach is theoretically viable and can significantly enhance the model's diagnostic performance. In recent years, MUDA has been widely used in the field of fault diagnosis. Liu et al. [21] proposed a bearing fault diagnosis algorithm based on a multi-feature spatially adaptive network; the algorithm is based on the feature extraction of the input samples through the feature extractor, while multi-core MMD is used to learn the domain-invariant representations of the source and target domain features, and, finally, the classifiers output the results of the target domain data state recognition. Tian et al. [22] established a new multi-source domain migration learning fault diagnosis framework, in which the source–target domain data pairs are constructed using anchor adapters, the shared features are extracted, and then the classifiers are weighted and integrated to ultimately achieve the purpose of fault diagnosis. Nguyen et al. [23]

proposed a multi-source teacher–student learning network for optimal transmission, and the feasibility of the model was verified on multiple datasets. Despite the promising results of the above research, it only performs a global alignment of source and target domain data, overlooking category-specific differences and losing detailed information between categories. Ghorvei et al. [24] introduced local maximum mean difference to realize the transfer of feature knowledge between the source domain and target domain and verified the feasibility of the proposed method under different noise ratios. The above method effectively addresses the issue of distribution inconsistency between the target and source domains. However, most current research primarily focuses on knowledge transfer within a single domain. In complex scenarios, relying on a single source domain may not suffice for diagnostic tasks, leading to reduced diagnostic effectiveness.

While the methods discussed offer innovative solutions and address the issue of limited diagnostic information in a single source domain, there is still a need for researchers to continuously develop new models in the field of multi-source transfer learning for bearing fault diagnosis. Graph neural networks, as an innovative approach through which to capture the information of nodes between samples in a "graph" representation, were first proposed by Scarselli et al. [25], who indicated that graph neural networks can pay more attention to the connectivity of the data. Zhang et al. [26] proposed a deep graph convolutional network for rolling bearing acoustic fault diagnosis and achieved better diagnostic results. Li et al. [27] converted the data into weighted graphs and proposed a fault diagnosis method based on a multisensory wild graph convolution network. Yin et al. [28] proposed a depth graph convolution network, combining multiscale data fusion and multiscale graph convolution fusion, for rotor-bearing fault diagnosis. Li et al. [29] combined a CNN and a graph convolutional network to realize the fault diagnosis of machinery under variable operating conditions, using domain confrontation and maximum mean difference methods.

When domain-adaptive technology addresses the issue of limited labels in data from a single source domain, knowledge transfer often fails to provide sufficient relevant information for the target diagnosis task, leading to poor recognition performance. This study introduces a graph convolutional network approach for diagnosing rolling bearing faults under varying operating conditions, leveraging multi-source domain transfer learning and incorporating a subdomain-adaptive strategy to enhance model performance despite significant data distribution differences. The main objectives are as follows: (1) To propose a multi-source subdomain-adaptive network model for rolling bearing fault diagnosis that integrates feature-related information from various source domains, thereby improving diagnostic accuracy in the absence of target domain labels. (2) To combine the strengths of deep convolutional networks and graph convolutional networks to learn the geometric structural information of sample data and enhance the network's feature learning capability.

## 2. Basic Principles

### *2.1. Spectral Convolution*

The graph convolutional neural network (GCN) approach involves applying convolutional operations to graphs, serving a similar purpose to CNNs in feature extraction. Graph convolutional networks are typically classified into two main groups based on current standards: (1) based on spatial or vertex domains and (2) based on frequency domains or spectrograms [30]. The core of spatial domain convolution lies in aggregating the information of neighboring nodes—for example, summing up the states of all directly connected neighboring nodes as hidden states of the current node—while frequency domain convolution uses graph spectral theory to perform convolution operations on a topological graph. In the field of graph studies, this method can be represented as a graph $G = (V, E)$ with

$N$ nodes, where $V$ and $E$ are the set of nodes and the set of edges, respectively, and the graph can be represented with a Laplace matrix in its general form [31].

$$L = D - A \tag{1}$$

where $D \in R^{n \times n}$ denotes the degree matrix and A is the adjacency matrix.

$$A_{i,j} = \begin{cases} 1 & node < v_i, v_j > is\ connect \\ 0 & node < v_i, v_j > is\ \text{not}\ connect \end{cases} \tag{2}$$

$$D_{i,j} = \begin{cases} d(i) & i = j \\ 0 & i \neq j \end{cases} \tag{3}$$

From the above equation, $D$ is the diagonal matrix and $D_{i,i} = \sum_{j=1}^{n} A_{i,j}$ is the degree of node $v_i$. The conversion relationship of the above formulas is illustrated in Figure 1.

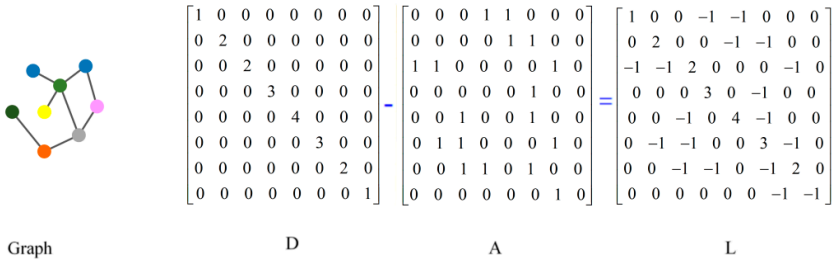


**Figure 1.** Conversion relationship diagram.

The Laplace matrix is represented as follows, and a spectral decomposition of the Laplace matrix is performed:

$$L^{sym} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2} \tag{4}$$

$$L^{sym} = U \Lambda U^{-1} = U \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \cdots & \\ & & & \lambda_n \end{bmatrix} U^{-1} = U \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \cdots & \\ & & & \lambda_n \end{bmatrix} U^T \tag{5}$$

where $I$ represents the unit diagonal matrix as a positive definite matrix, $U$ represents the orthogonal matrix, formed by combining the eigenvectors of the Laplace matrix, $\Lambda$ is the eigenvalue matrix of the Laplace matrix, and $\lambda_{i,i}$ is the matrix eigenvalue.

According to the convolution theorem, which states that "the Fourier transform of the convolution of two signals in the spatial domain equals the product of their Fourier transforms", the following transforms can be applied to graph convolution networks:

$$x *_G g = F^{-1}(F(x) \odot F(g)) = U(U^T x \odot U^T g) = U(U^T g \odot U^T x) \tag{6}$$

where $x$ is the input signal, $*_G$ is the graph convolution operation, $F(x)$ and $F^{-1}$ are the Fourier forward transform and Fourier inverse transform, i.e., the signal is transformed from the spatial domain to the spectral domain and then to the spatial domain signal, and $\odot$ is the Hadamard product.

### *2.2. Local Maximum Mean Difference*

Domain-adaptive methods are generally categorized into feature-based, instance-based, and model parameter-based approaches. Feature-based methods are particularly popular in fault diagnosis because they are not constrained by whether diagnostic samples are labeled. These methods can be further divided into difference-based, adversarial-based, and reconstruction-based approaches. The difference-based method measures the

distribution difference between the source and target domains and maps their data features to a shared feature space, reducing the distribution gap. The adversarial-based approach, inspired by generative adversarial networks, involves generators that produce effective features to mislead domain discriminators, which distinguish between the source and target domains; this interaction enhances the model's performance on target domain samples. The reconstruction-based method incorporates auto-encoders, using an encoder and a decoder to produce features similar to those from both domains. In practice, it is crucial to not only qualitatively assess the similarity between the source and target domains but also to quantitatively measure this similarity. Designing appropriate measurement criteria is essential for effectively narrowing the gap between the domains.

Common adaptive measurement criteria include MK-MMD, JMMD, CORAL, etc. [32], but these methods achieve domain adaptation only by aligning the global distributions of the source and target domains. The local maximum mean difference (LMMD), on the other hand, takes interregional categories into account, with the goal of aligning the differences between each category, rather than narrowing the overall distribution. Suppose samples $x_i^s$ and $x_j^t$ belong to each class, with weights $\omega_i^{sc}$ and $\omega_j^{tc}$, where $y_{ic}$ is the CTH element of the label vector $y_i$, and $c$ is the number of classes. As an extension of the MMD method, MK-MMD adopts a linear combination of multiple nuclei to find the optimal nucleus $K$, as follows:

$$\omega_i^c = \frac{y_{ic}}{\sum\limits_{(x_j, y_j) \in D} y_{jc}} \tag{7}$$

$$\hat{D}(X_s, X_t) = \frac{1}{C} \sum_{c=1}^{C} \left\| \sum_{i=1}^{m} \omega_i^{sc} \phi(x_i^s) - \sum_{j=1}^{n} \omega_j^{tc} \phi(x_j^t) \right\|_H^2 \tag{8}$$

$\phi()$ cannot be calculated directly, so the above equation must be equivalent-transformed, as shown in Formula (9), where $k(z_i^s, z_j^s)$ represents the inner product of the vector.

$$loss_{lmmd} = \hat{D}(X_s, X_t) = \frac{1}{C} \sum_{c=1}^{C} \begin{array}{l} [\sum\limits_{i=1}^{m} \sum\limits_{j=1}^{m} \omega_i^{sc} \omega_j^{sc} k(z_i^s, z_j^s) \\ + \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} \omega_i^{tc} \omega_j^{tc} k(z_i^t, z_j^t) \\ - 2 \sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n} \omega_i^{sc} \omega_j^{tc} k(z_i^s, z_j^t)] \end{array} \tag{9}$$

### *2.3. Ranger Optimization Algorithm*

The Ranger optimization algorithm combines the highly advanced natures of the Radam and Lookahead optimization algorithms [33] and presents new breakthroughs in deep learning optimization. The Ranger optimization algorithm utilizes Radam as the internal optimizer and Lookahead as the external optimizer, where the internal optimizer $\varsigma$ updates the fast weight. The updated rule of the fast weight is as follows:

$$\theta_{t,i+1} = \theta_{t,i} + \varsigma(L, \theta_{t,i-1}, d) \tag{10}$$

where $L$ represents the objective function, $\varsigma$ represents the optimization algorithm, $d$ represents the small sample training batch, i = 1, 2... k represents the exploration of the i batch, and $t$ is the number of iterations. The update of the slow weights is affected by the fast weights, and when the internal optimizer $\varsigma$ completes k batch explorations, the slow weights are updated to the following:

$$\begin{array}{l} \phi_{t+1} = \phi_t + \alpha(\theta_{t,k} - \phi_t) \\ = \alpha \left[ \theta_{t,k} + (1-\alpha)\theta_{t-1,k} + \ldots + (1-\alpha)^{t-1} \theta_{0,k} \right] \\ + (1-\alpha)^t \phi_0 \end{array} \tag{11}$$

## 3. Fault Diagnosis Process

In this article, we present a multi-source locally adaptive (MSLA) fault diagnosis model utilizing the local maximum mean difference measurement method. The model's structure varies based on the number of source domains $N$, and it encompasses three primary stages: signal preprocessing, feature extraction, and classification. During signal preprocessing, bearing vibration signals from Jiangnan University are overlapped, and continuous wavelet transformation is performed using the cmor wavelet on each type of signal sample after expanding the sample size. This process constructs group $N$, a source domain–target domain data pair, facilitating network training. In the feature extraction stage, the model consists of a shared feature extraction network and a domain-specific feature extraction network. Resnet50 [34] is used as the shared feature extraction network to capture common features from the data pairs, which are then fed into the domain-specific feature extraction network. This network employs a graph convolutional network to extract the domain-specific features for each data pair. The Softmax layer, positioned after the feature extraction network, serves as the bearing state classifier. The training loss function of the network includes three components: the local maximum mean difference measurement loss function to minimize distribution differences and aid in learning domain-invariant representations; the difference loss between classifiers to reduce classification errors near the target domain class boundaries; and cross-entropy loss to measure discrepancies between the actual and predicted labels in the source domain. The model uses Radam and Lookahead algorithms for parameter optimization. After reaching the specified number of iterations, the trained model processes the target domain data and generates diagnostic results. The diagnostic flow chart is illustrated in Figure 2.

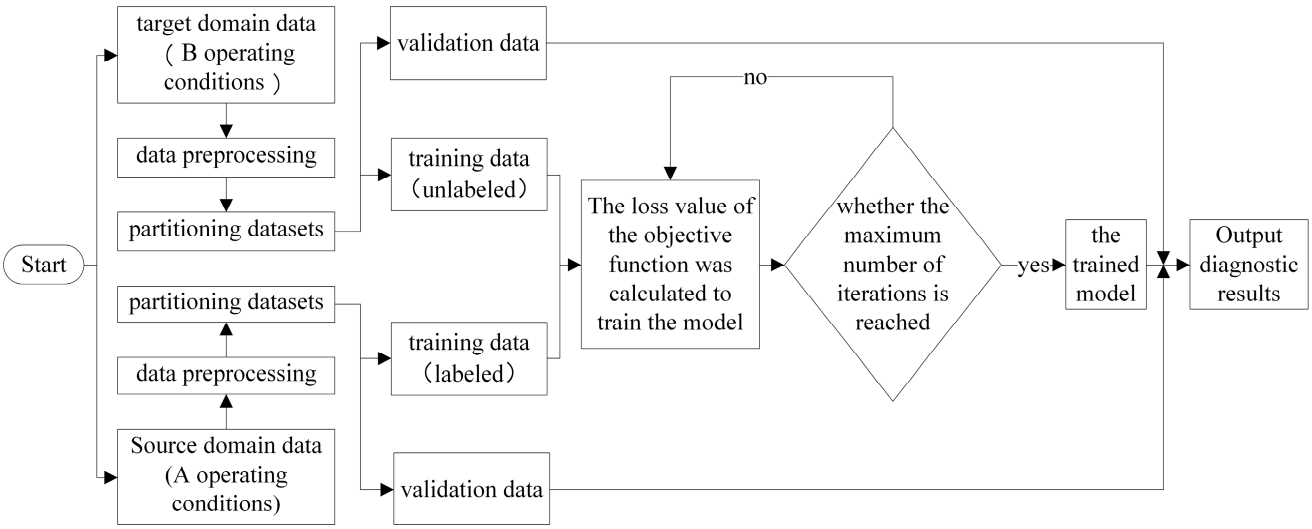


**Figure 2.** The diagnostic flow chart.

In the continuous wavelet transform, it is assumed that the original signal $x(t) \in L^2(R)$ exists, and the continuous wavelet transform formula is obtained by the convolution between $x(t)$ and the wavelet cluster $\psi_{a,b}(\cdot)$:

$$\psi_{e,f}(t) = \frac{1}{\sqrt{e}}\psi\left(\frac{t-f}{e}\right) \tag{12}$$

$$cwt(e,f) = \int x(t)\psi_{e,f}(t)dt \tag{13}$$

In the formula, $e$ and $f$ represent the stretching factor and displacement factor, respectively, $t$ represents time, and $\psi_{e,f}(\cdot)$ represents the wavelet basis function after stretching and displacement.

Since the depth of neural networks impacts their feature extraction capability, shallow networks often have weaker feature extraction abilities. To address this, ResNet50 is used as the shared feature extraction network. ResNet50 consists of 50 layers, organized into five stages. Detailed network parameters are provided in Table 1.

**Table 1.** The values of network parameters.

| Network Architecture | Category | Coverage Area |
|---|---|---|
| input | 224 × 224 × 3 | |
| Con1 | Convolution layer | 7 × 7; 64 |
| Max pool | Maximum pooling layer | 3 × 3; 64 |
| Conv2_x | Residual block1 × 3 | 1 × 1; 64<br>3 × 3; 64<br>…1 × 1; 256 } × 3 |
| Conv3_x | Residual block1 × 4 | 1 × 1; 256<br>3 × 3; 128<br>…1 × 1; 512 } × 4 |
| Conv4_x | Residual block1 × 6 | 1 × 1; 256<br>3 × 3; 256<br>…1 × 1; 1024 } × 6 |
| Conv5_x | Residual block1 × 3 | 1 × 1; 512<br>3 × 3; 512<br>…1 × 1; 2048 } × 3 |
| Average | Average pooling layer | |

The diagram of the fault diagnosis model structure is presented below (Figure 3).

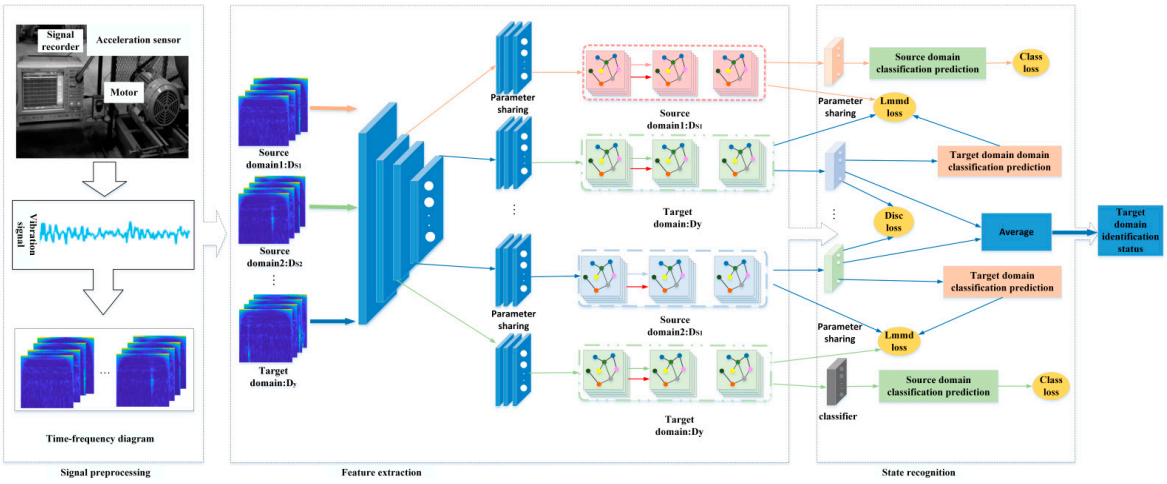


**Figure 3.** Multi-source rolling bearing fault diagnosis under variable working conditions.

The feature extraction network consists of the shared ResNet50 and domain-specific graph convolutional networks, working together to learn high-dimensional features $X$. The structural parameters of this network are detailed in Table 2.

**Table 2.** Feature extraction network structural parameters.

| Network | Network Structure and Parameters |
|---|---|
| Shared feature extraction network *ResNet*50( ) | ResNet50( ) |
| Domain-specific feature extraction network *CNN*( ) | Conv2d(2048,256)-BN2d (256)-ReLu( )-Conv2d(256,256)-BN2d (256)-ReLu( )-Conv2d(256,256)-BN2d (256)-ReLu( )-MRF-GCN( ) |

The high-dimensional feature $X$ serves as the node feature matrix. The adjacency matrix $\hat{A}$ is derived by multiplying the feature vector by its transpose using the graph generation layer (GGL). Subsequently, the sparse adjacency matrix $A$ is obtained through a Top-K sorting mechanism.

$$X_1 = Resnet50(X_{input})X = CNN(X_1) \tag{14}$$

$$\begin{cases} \hat{X} = MLP(X) \\ \hat{A} = nol(X \times \hat{X}) \\ A = Top - k(A) \end{cases} \tag{15}$$

In the formula, $\hat{X}$ represents the feature vector output using node feature matrix $X$ after multilayer perceptron ($MLP$), $nol$ represents the normalized function, and $Top - k(\ )$ represents the calculation of each row of the matrix; the index corresponding to the top $k$ maximum values in each row is returned. The model incorporates a multi-receptive field convolutional layer (MRF-GCN) to extract image features. The MRF-GCN aggregates information from multiple receptive fields to capture more detailed information. In this study, a double-layer MRF-GCN is used to learn the feature representation of sample data, as follows:

$$\begin{gathered} H = \left[\sum_{k_0}^{K_0-1} \theta_{k_0} \Lambda^{k_0} X, \quad \sum_{k_1}^{K_1-1} \theta_{k1} \Lambda^{k_1} X, \quad \ldots \quad \sum_{k_v}^{K_v-1} \theta_k \Lambda^{K_v} X\right] \\ H^0 = MRFConv(AXW^0) \\ H^1 = MRFConv(AHW^1) \end{gathered} \tag{16}$$

where $H^0$ and $H^1$ represent the feature representation extracted from the first and second layers MRF-Conv, respectively.

The Softmax layer was employed for bearing state classification. Once the classifier was integrated into the graph convolutional network, three optimization objectives were established for the network training process: minimizing the classification loss $Loss_{class}$ in the source domain, minimizing the invariant loss $Loss_{lmmd}$ within the domain, and minimizing the dissimilarity loss $Loss_{disc}$ between different classifiers. The loss function is defined as follows:

$$\begin{gathered} loss_{class} = \sum_{j=1}^{N} J(C_j(G_j(H_j(F(X_i^{s_j}))))), Y_i^{s_j}) \\ Loss_{disc} = C_n^2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left| C_j(G_j(H_j(F(x^t)))) - C_i(G_j(H_i(F(x^i)))) \right| \\ Loss = Loss_{class} + \lambda \cdot Loss_{lmmd} + \tau \cdot Loss_{disc} \end{gathered} \tag{17}$$

where $X_i^s$ and $Y_i^s$ represent different distributions of source domain data and labels, respectively; $J$ is the cross-entropy function; $C_j$ represents the prediction results of the label classifier; $G_j$ represents the graph convolutional feature extraction network; $\lambda$ and $\gamma$ represent the hyperparameters of $Loss_{lmmd}$ and $Loss_{disc}$, respectively.

## 4. Experimental Platform and Result Analysis

### *4.1. Experimental Setup*

The datasets from Jiangnan University (JNU) are often used to verify the effectiveness of transfer learning algorithms in the field of fault diagnosis for induction motors, accelerators, signal regulators, etc. [35]. The JNU bearing dataset consists of vibration data recorded at three different speeds (600 RPM, 800 RPM, and 1000 RPM) with a sampling frequency of 50 kHz. This dataset includes four types of bearing conditions: normal, inner ring failure, outer ring failure, and rolling element failure. To represent various working conditions, the data collected at 600 RPM, 800 RPM, and 1000 RPM are labeled as A, B, and C, respectively. The dataset includes a total of six transfer learning tasks. Each vibration data sample from these conditions was processed using overlapping sampling, with a window length of 1024

and a sliding step of 256, resulting in 800 data groups per state. Table 3 details the division of the JNU data used in this study.

**Table 3.** JNU bearing dataset settings.

| Dataset | Speed (r/min) | Bearing Condition | Label | Sample Points per Set | Sample Size |
|---|---|---|---|---|---|
| A/B/C | 600/800/1000 | Normal state | 0 | 1024 | 800 |
| | 600/800/1000 | Inner ring fault | 1 | 1024 | 800 |
| | 600/800/1000 | Outer ring fault | 2 | 1024 | 800 |
| | 600/800/1000 | Rolling element fault | 3 | 1024 | 800 |

The schematic diagram of the experimental platform structure and the data acquisition diagram are shown in Figure 4.

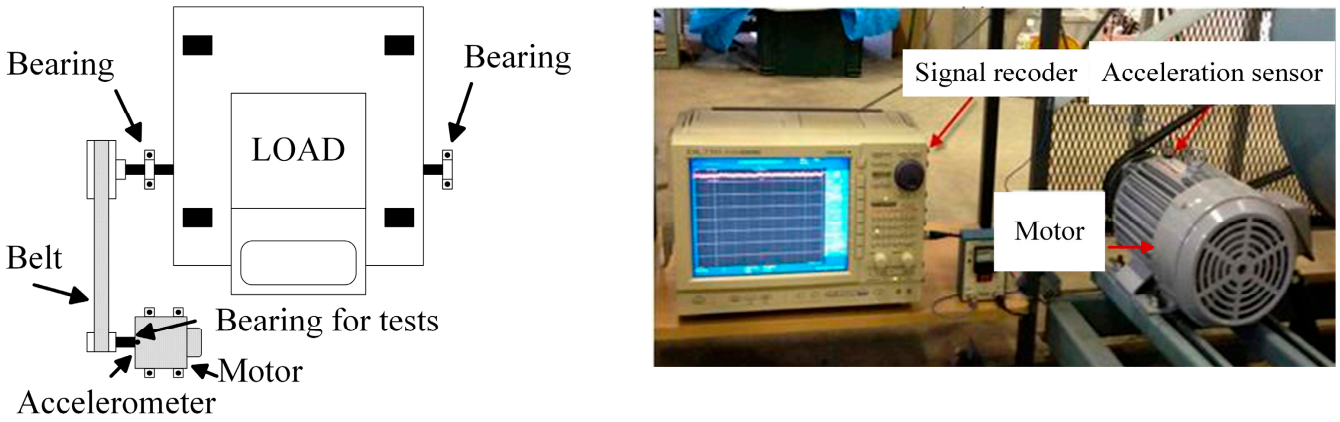


**Figure 4.** Experimental platform: (**a**) experimental platform structure diagram and (**b**) data acquisition diagram.

Comparison experiments were conducted using the Jiangnan University bearing dataset. The source domain samples were used entirely as the training set, while all target domain samples (which were unlabeled) served as both the training and test sets. Network hyperparameters were determined through several preliminary experiments, and the same hyperparameter values were applied across all comparison methods. Details of these hyperparameters can be found in Table 4.

**Table 4.** Model parameter settings.

| Parameter Name | Parameter Value |
|---|---|
| Number of samples per batch | 32 |
| Learning rate $lr_0$ | 0.0005 |
| Learning rate $lr_1$ | 0.000005 |
| Number of iterations | 1000 |
| Activation function | ReLU |

### *4.2. Comparative Analysis of Experimental Results for Adaptive Methods in Different Fields*

Table 5 presents the classification accuracy of various methods for the three migration tasks: B + C→A; A + C→B; and A + B→C. In these tasks, datasets B and C were used as the source domain data with dataset A as the target domain; datasets A and C served as the source domains with dataset B as the target domain; and datasets A and B acted as the source domains with dataset C as the target domain. To demonstrate the effectiveness of the MSLA approach more objectively, joint maximum mean discrepancy (JMMD) loss [36], CORrelation ALignment (CORAL) loss [37], and multiple kernel maximum mean are selected discrepancy (MK-MMD) loss [38] are compared as adaptive algorithms.

**Table 5.** Comparison of experimental mean diagnostic results of different methods.

| Tasks/Adaptive Methods | JMMD | CORAL | MK-MMD | MSLA |
|---|---|---|---|---|
| B + C→A | 96.97% | 97.56% | 97.56% | 99.17% |
| A + C→B | 96.56% | 98.16% | 98.09% | 99.47% |
| A + B→C | 98.83% | 99.54% | 99.29% | 99.88% |

It can be observed from the data in Table 5 that high-precision diagnosis has been obtained by comparing several models, the reason for which may be that, compared to the globally adaptive method, the enhanced subdomain-adaptive approach allows the model to capture detailed information from each category and more effectively align the subdomain distributions within the same category. At the same time, the diagnostic accuracy of the MSLA method was further improved, and the diagnostic accuracy results from the three groups of tasks were 99.17%, 99.47%, and 99.88%. Compared with other models, the diagnostic accuracy results for the MSLA model were increased by 1.61–2.20%, 1.31–2.91%, and 0.34–1.05%, thus showing that the graph convolution layer can help the model learn the structural features between samples and achieve the purpose of extracting richer feature information.

Figure 5 illustrates the confusion matrix of the output categories after the model performs fault diagnosis for task A + C→B. In this matrix, the horizontal axis represents the predicted categories for the target domain samples, while the vertical axis shows their true labels. Comparing the confusion matrices of the four methods, it is observed that the first three methods (a–c) misclassified 110, 59, and 61 target domain samples, respectively. In contrast, the MSLA method misclassified only 17 out of 3200 target domain samples. Specifically, the misclassifications included nine rolling element fault samples labeled as normal, three rolling element fault samples labeled as outer ring fault, one inner ring fault sample labeled as normal, two normal samples labeled as outer ring fault, and two normal samples labeled as rolling element fault. All other target domain samples were correctly classified, resulting in a recognition accuracy of 99.47%.

To further assess the effectiveness of the multi-source domain method, we compared it with experiments involving single-source domain migration models. Specifically, datasets B, C, and B + C were used as the source domains, with dataset A as the target domain. For simplicity, these cases are denoted as B→A, C→A, and B + C→A, respectively. Given that the difference between single-domain and multi-domain sample data might impact the results, Table 6 presents the average diagnostic outcomes for various models. In this context, SSLA refers to a single-source local domain-adaptive transfer learning method enhanced with graph convolutional networks.

The data in Table 6 show that, compared with A→C and B→C, the diagnostic accuracy results of A + B→C increased by 1.29%, 1.20%, and 1.20%; compared with A→B and C→B, the diagnostic accuracy results of A and C→B were increased by 1.16%, 0.59%, and 0.3%, respectively. The diagnostic accuracy results of B + C→A increased by 2.94%, 4.33%, and 2.24%, compared with B→A and C→A. The results indicate that the MSLA model significantly enhances the diagnostic performance compared to any single-transfer learning task. Simply aggregating source domain data does not substantially improve the model's diagnostic

accuracy, and the results for the three single-source migration tasks are lower than those of the MSLA model. Thus, it can be concluded that by effectively integrating multi-source domain models, we can capture more domain-invariant information, boost classification performance, and address the limitations of single-source domain migration methods.

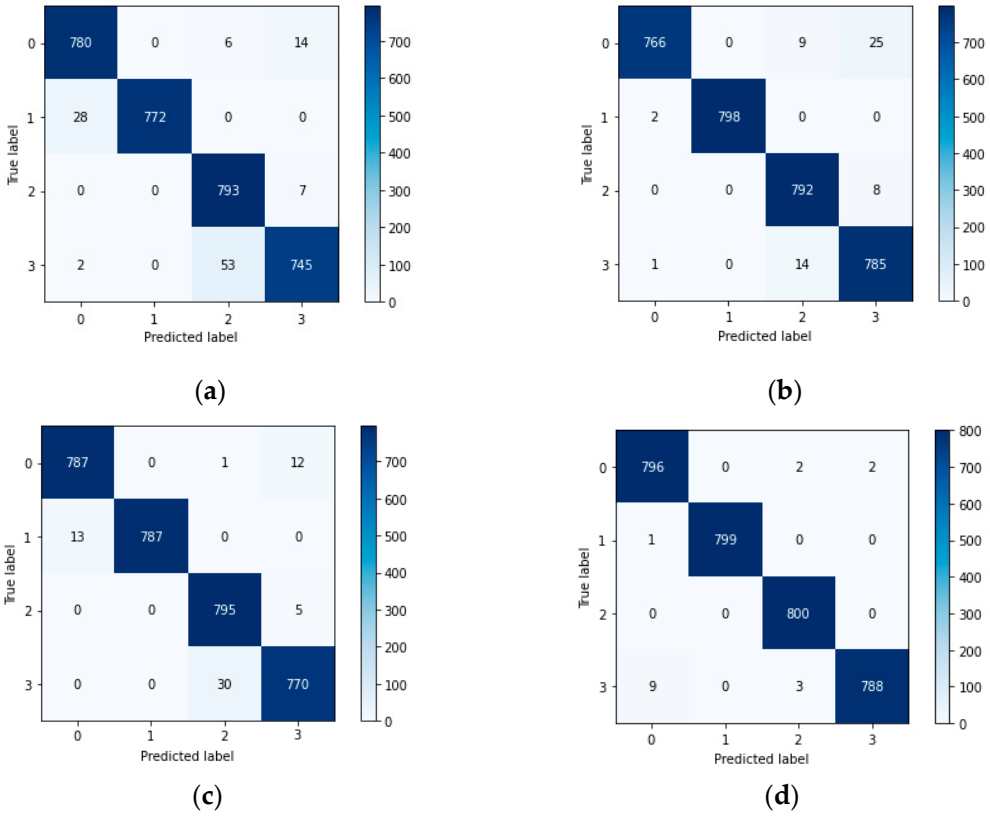


**Figure 5.** Recognition accuracy confusion matrix for fault diagnosis in task A + C→B by different models is as follows: (**a**) accuracy of the JMMD method for the A + C→B task; (**b**) accuracy of the CORAL method for the C→A task; (**c**) accuracy of the MK-MMD method for the B + C→A task; and (**d**) accuracy of the MSLA method for the B, C→A tasks.

**Table 6.** Comparison of experimental average diagnostic results.

| Tasks | MSLA | Tasks | SSLA |
|---|---|---|---|
| | | B→A | 96.23% |
| B + C→A | 99.17% | C→A | 94.84% |
| | | B + C→A | 96.93% |
| | | A→B | 98.73% |
| A + C→B | 99.89% | C→A | 99.30% |
| | | A + C→B | 99.59% |
| | | A→C | 98.59% |
| A + B→C | 99.88% | B→C | 98.68% |
| | | A + B→C | 98.68% |

In this study, t-SNE visualization was used to examine the results of experiments migrating to dataset A, specifically for B + C→A. t-SNE is a nonlinear dimensionality reduction technique that effectively reveals clustering relationships among different categories, making it useful for visual analysis in fault diagnosis. The results for this study are shown in Figure 6. Compared to the SSLA model, the MSLA model exhibits less overlap in visualization and clearer boundaries between different states. This indicates that after training with the MSLA model, the feature information is more effectively differentiated,

allowing the classifier to identify bearing states with greater accuracy. To highlight the advantages of the MSLA method, the paper uses t-SNE visualization for a group of variable condition migration tasks that showed the most distinct effects. The dimensionality of the features from the last layer of the network was reduced to a two-dimensional plane, as shown in Figure 6. Compared to the visualizations from the three SSLA models, the MSLA model exhibits less overlap and more pronounced boundaries between states, demonstrating improved feature differentiation and more accurate bearing state recognition.

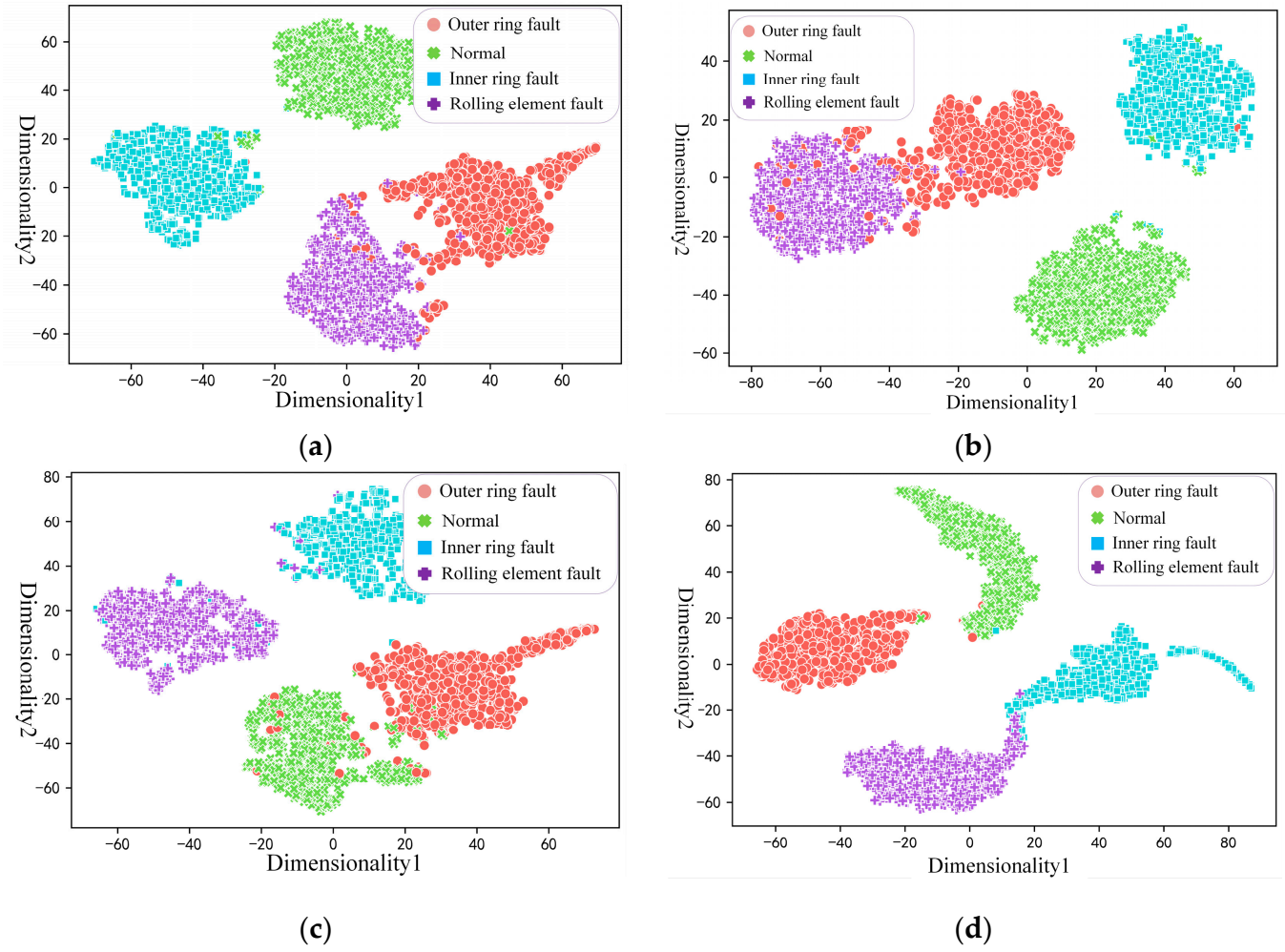


**Figure 6.** The feature distribution maps extracted from the 4 models: (**a**) feature distribution of SSLA methods in the B→A task; (**b**) feature distribution of SSLA methods in the C→A task; (**c**) feature distribution of SSLA methods in the B + C→A task; and (**d**) feature distribution of MSLA methods in the B + C→A tasks.

After the single working condition migration tasks B→A, C→A, and B + C→A, the recognition accuracy results were 96.23%, 94.84%, and 96.93%. As can be seen from Figure 6, different bearing states overlap more, making it difficult for the classifier to accurately determine the actual state. After multi-source transfer learning, by using the proposed MSLA method, the four types of fault samples are essentially separated, resulting in an improved classification accuracy of 99.17%. Model validation was carried out on the JNU bearing dataset, and the results showed that the recognition accuracy of the MSLA model was increased by 0.30–4.33% compared with its corresponding single-source domain diagnosis model, which indicates that the proposed method can still effectively identify the bearing state in the absence of target domain labels. It can learn more domain-invariant

information, improve the classification performance of the model, and make up for the shortcomings of the single-source domain migration method.

## 5. Conclusions

To address the issues of a lack of labels in target domains and poor model training, we propose a multi-source locally adaptive (MSLA) fault diagnosis model that leverages data geometry and data connectivity. This model combines the strengths of deep convolutional networks and graph convolutional networks to learn more comprehensive features. Validation using the JNU bearing dataset demonstrated that the MSLA model achieved a recognition accuracy improvement of 0.30–4.33% over its corresponding single-source domain diagnosis model. This indicates that the proposed method can effectively identify bearing states even in the absence of target domain labels.

The main contributions are as follows: (1) We propose a multi-source subdomain-adaptive network model for rolling bearing fault diagnosis, which learns feature knowledge from various source domains, enhancing diagnosis accuracy when target domain label information is absent. (2) This study integrates the strengths of deep convolutional networks and graph convolutional networks to learn the geometric structure information of sample data, thereby improving the network's feature learning capability.

In addition to time–frequency analysis, in the future, this can be combined with the data from a variety of sensors, such as image sensors, lidar, temperature and humidity sensors, etc., for multi-modal data fusion to provide more comprehensive and accurate condition monitoring and fault diagnosis of agricultural machinery. The existing approach has been successful with specific types of agricultural machinery, but its potential can also be validated and applied in other types of agricultural machinery, such as plant protection drones, planters, and irrigation systems. This not only helps to improve the intelligence of agricultural machinery but also improves the overall efficiency and benefits of agricultural production. The next step will be to develop real-time monitoring and remote diagnosis systems based on the Internet of Things (IoT) to achieve the real-time monitoring, data acquisition, and remote fault diagnosis of agricultural machinery, thereby reducing downtime and improving production efficiency.

**Author Contributions:** Conceptualization, F.X. and E.S.; methodology, L.W.; validation, G.W.; investigation, Q.X.; writing—original draft preparation, F.X.; writing—review and editing, F.X. and E.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (52265068), the Natural Science Foundation of Jiangxi Province (20224BAB204050), Equipment Key Laboratory Project of the Ministry of Education (KLCEZ2022-02), and the Project of Jiangxi Provincial Department of Education (GJJ2200627).

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** Data are available in a publicly accessible repository that does not issue DOIs. Publicly available datasets were analyzed in this study. These data can be found in the following: https://github.com/ClarkGableWang/JNU-Bearing-Dataset (accessed on 3 July 2024).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Zhang, Q.; Chu, Y.; Xue, Y.; Ying, H.; Chen, X.; Zhao, Y.; Ma, W.; Ma, L.; Zhang, J.; Yin, Y.; et al. Outlook of China's agriculture transforming from smallholder operation to sustainable production. *Glob. Food Secur.* **2020**, *26*, 100444. [CrossRef]
2. Wang, Y.; Li, D.; Nie, C.; Gong, P.; Yang, J.; Hu, Z.; Li, B.; Ma, M. Research Progress on the Wear Resistance of Key Components in Agricultural Machinery. *Materials* **2023**, *16*, 7646. [CrossRef] [PubMed]
3. Liu, C.H.; Chen, X.Y.; Gu, J.M.; Jiang, S.N.; Feng, Z.L. High-speed wear lifetime analysis of instrument ball bearings. *Proc. Inst. Mech. Eng. Part J J. Eng. Tribol.* **2009**, *223*, 497–510. [CrossRef]
4. Scolaro, E.; Beligoj, M.; Estevez, M.P.; Alberti, L.; Renzi, M.; Mattetti, M. Electrification of agricultural machinery: A review. *IEEE Access* **2021**, *9*, 164520–164541. [CrossRef]

5. Craessaerts, G.; De Baerdemaeker, J.; Saeys, W. Fault diagnostic systems for agricultural machinery. *Biosyst. Eng.* **2010**, *106*, 26–36. [CrossRef]
6. Mishra, D.; Satapathy, S. Reliability and maintenance of agricultural machinery by MCDM approach. *Int. J. Syst. Assur. Eng. Manag.* **2023**, *14*, 135–146. [CrossRef]
7. Alimova, Z.X.; Kholikova, N.A.; Kholova, S.O.; Karimova, K.G. Influence of the antioxidant properties of lubricants on the wear of agricultural machinery parts. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *868*, 012037. [CrossRef]
8. Han, J.; Zhang, J.; Zeng, B.; Mao, M. Optimizing dynamic facility location-allocation for agricultural machinery maintenance using Benders decomposition. *Omega* **2021**, *105*, 102498. [CrossRef]
9. Celenta, G.; De Simone, M.C. Retrofitting techniques for agricultural machines. In *New Technologies Development and Application III*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 388–396.
10. Niazian, M.; Niedbała, G. Machine Learning for Plant Breeding and Biotechnology. *Agriculture* **2020**, *10*, 436. [CrossRef]
11. Cheng, Z.; Lu, Z. Research on Load Disturbance Based Variable Speed PID Control and a Novel Denoising Method Based Effect Evaluation of HST for Agricultural Machinery. *Agriculture* **2021**, *11*, 960. [CrossRef]
12. Moshrefzadeh, A. Condition monitoring and intelligent diagnosis of rolling element bearings under constant/variable load and speed conditions. *Mech. Syst. Signal Process.* **2021**, *149*, 107153. [CrossRef]
13. Lv, Y.; Zhao, W.; Zhao, Z.; Li, W.; Ng, K.K. Vibration signal-based early fault prognosis: Status quo and applications. *Adv. Eng. Inform.* **2022**, *52*, 101609. [CrossRef]
14. Cheng, C.; Zhou, B.; Ma, G.; Wu, D.; Yuan, Y. Wasserstein distance based deep adversarial transfer learning for intelligent fault diagnosis with unlabeled or insufficient labeled data. *Neurocomputing* **2020**, *409*, 35–45. [CrossRef]
15. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A Survey of Transfer Learning. *J. Big Data* **2016**, *3*, 1–40. [CrossRef]
16. Zhang, X.; He, L.; Wang, X.; Wang, J.; Cheng, P. Transfer Fault Diagnosis based on Local Maximum Mean Difference and K-means. *Comput. Ind. Eng.* **2022**, *172*, 108568. [CrossRef]
17. Qian, W.; Li, S.; Yi, P.; Zhang, K. A Novel Transfer Learning Method for Robust Fault Diagnosis of Rotating Machines under Variable Working Conditions. *Measurements* **2019**, *138*, 514–525. [CrossRef]
18. Cheng, C.; Zhou, B.; Ma, G.; Wu, D.; Yuan, Y. Wasserstein Distance Based Deep Adversarial Transfer Learning for Intelligent Fault Diagnosis. *arXiv* **2019**, arXiv:1903.06753.
19. Li, X.; Zhang, Z.; Gao, L.; Wen, L. A New Semi-supervised Fault Diagnosis Method via Deep CORAL and Transfer Component Analysis. *IEEE Trans. Emerg. Top. Comput. Intell.* **2021**, *6*, 690–699. [CrossRef]
20. Wang, Z.; He, X.; Yang, B.; Li, N. Subdomain Adaptation Transfer Learning Network for Fault Diagnosis of Roller Bearings. *IEEE Trans. Ind. Electron.* **2021**, *69*, 8430–8439. [CrossRef]
21. Liu, J.; Guan, R.; Li, Z.; Zhang, J.; Hu, Y.; Wang, X. Adaptive Multi-feature Fusion Graph Convolutional Network for Hyperspectral Image Classification. *Remote Sens.* **2023**, *15*, 5483. [CrossRef]
22. Tian, J.; Han, D.; Li, M.; Shi, P. A Multi-source Information Transfer Learning Method with Subdomain Adaptation for Cross-domain Fault Diagnosis. *Knowl.-Based Syst.* **2022**, *243*, 108466. [CrossRef]
23. Nguyen, T.; Le, T.; Zhao, H.; Tran, Q.H.; Nguyen, T.; Phung, D. Most: Multi-source domain adaptation via optimal transport for student-teacher learning. In *Uncertainty in Artificial Intelligence*; PMLR: London, UK, 2021; pp. 225–235.
24. Ghorvei, M.; Kavianpour, M.; Beheshti, M.T.; Ramezani, A. An unsupervised bearing fault diagnosis based on deep subdomain adaptation under noise and variable load condition. *Meas. Sci. Technol.* **2021**, *33*, 025901. [CrossRef]
25. Scarselli, F.; Gori, M.; Tsoi, A.C.; Hagenbuchner, M.; Monfardini, G. The graph neural network model. *IEEE Trans. Neural Netw.* **2008**, *20*, 61–80. [CrossRef] [PubMed]
26. Zhang, N.; Wu, L.; Yang, J.; Guan, Y. Naive Bayes Bearing Fault Diagnosis Based on Enhanced Independence of Data. *Sensors* **2018**, *18*, 463. [CrossRef] [PubMed]
27. Li, T.; Zhao, Z.; Sun, C.; Yan, R.; Chen, X. Multireceptive Field Graph Convolutional Networks for Machine Fault Diagnosis. *IEEE Trans. Ind. Electron.* **2021**, *68*, 12739–12749. [CrossRef]
28. Yin, P.; Nie, J.; Liang, X.; Yu, S.; Wang, C.; Nie, W.; Ding, X. A Multi-scale Graph Convolutional Neural Network Framework for Fault Diagnosis of Rolling Bearing. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 2520713. [CrossRef]
29. Li, T.; Zhao, Z.; Sun, C.; Yan, R.; Chen, X. Domain Adversarial Graph Convolutional Network for Fault Diagnosis under Variable Working Conditions. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3515010. [CrossRef]
30. Zhang, S.; Tong, H.; Xu, J.; Maciejewski, R. Graph Convolutional Networks: A Comprehensive Review. *Comput. Soc. Netw.* **2019**, *6*, 11. [CrossRef] [PubMed]
31. Das, K.C. The Laplacian Spectrum of a Graph. *Comput. Math. Appl.* **2004**, *48*, 715–724. [CrossRef]
32. Gatgash, Z.E.; Sadeghi, S.H. Comparative Effect of Conventional and Adaptive Management Approaches on Watershed Health. *Soil Tillage Res.* **2024**, *235*, 105869. [CrossRef]
33. Natarajan, S.; Kurian, S.; Divakarachari, P.B.; Falkowski-Gilski, P. An Automated Learning Model for Twitter Sentiment Analysis using Ranger AdaBelief Optimizer based Bidirectional Long Short Term Memory. *Expert Syst.* **2024**, *41*, e13610. [CrossRef]
34. Chakravarty, N.; Dua, M. Feature extraction using GTCC spectrogram and ResNet50 based classification for audio spoof detection. *Int. J. Speech Technol.* **2024**, *27*, 225–237. [CrossRef]
35. Fan, L. *Fault Diagnosis and Performance Degradation Assessment of Rolling Bearings*; Jiangnan University: Wuxi, China, 2021.

36. Xu, F.; Hong, D.; Tian, Y.; Wei, N.; Wu, J. Unsupervised Deep Transfer Learning Method for Rolling Bearing Fault Diagnosis Based on Improved Convolutional Neural Network. *J. Phys. Conf. Ser.* **2024**, *2694*, 012050. [CrossRef]
37. Liu, X.; Cheng, W.; Zhang, L.; Chen, X.; Wang, S. An Intelligent Hybrid Bearing Fault Diagnosis Method Based on Transformer and Domain Adaptation. In Proceedings of the 2021 IEEE International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC) 2021, Weihai, China, 13–15 August 2021; pp. 304–310.
38. Li, X.; Yu, T.; He, D.; Xie, Z.; Kong, X. Fusion with Joint Distribution and Adversarial Networks: A New Transfer Learning Approach for Intelligent Fault Diagnosis. In Proceedings of the PHM Society Asia-Pacific Conference, Tokyo, Japan, 11–14 September 2023; Volume 4.