Research Article

Research on the Initial Fault Prediction Method of Rolling Bearings Based on DCAE-TCN Transfer Learning

Huaitao Shi [ID],[1] **Yajun Shang** [ID],[1] **Xiaochen Zhang** [ID],[1] **and Yinghan Tang**[2]

[1]*School of Mechanical Engineering, Shenyang Jianzhu University, Shenyang 110168, China*
[2]*Bourns College of Engineering, University of California Riverside, Los Angeles 90001, CA, USA*

Correspondence should be addressed to Xiaochen Zhang; xczhang@sjzu.edu.cn

In actual working conditions, the initial faults of rolling bearings are difficult to effectively predict due to the lack of evolution knowledge, weak fault information, and strong noise interference. In this paper, a rolling bearing initial fault prediction model that is based on transfer learning and the DCAE-TCN is presented. Firstly, a deep autoencoder (DAE as the first two hidden layers and CAE as the last hidden layer) is used to extract fault features from the rolling bearing vibration signal data. Then, the balanced distributed adaptation (BDA) is used to minimise the distribution difference and class spacing between extracted fault features, and a common feature set is constructed. The temporal features of the original vibration signal in the target domain are extracted using the advantages of the TCN. The experiments are conducted on the publicly available XJTU-SY dataset. The experimental results show that the proposed method can effectively learn the transferable features and compensate the differences between the source and target domains and has a promising application with higher accuracy and robustness for the prediction of early failures of rolling bearings.

1. Introduction

Rolling bearings, as one of the core components of rotating machinery [1, 2], are extremely vulnerable to damage due to installation, temperature, lubrication, and other factors during long-term operation [3]. The faults of rolling bearings will cause damage to other parts and even the life of the entire machine. If the initial fault is not effectively identified and removed in time, the accuracy of the rotating machinery will continue to decrease until the system is completely damaged, or it even causes irreversible consequences.

The information on the initial fault characteristics of rolling bearings is very weak. So far, the main methods for the initial fault are mathematical model and data drive [4]. The mathematical modelling approach relies on the operational vibration signal or acoustic signal of rotating machines for the fault diagnosis [5], which are analysed in the time domain, frequency domain, and time-frequency domain, and a large number of characteristic parameters are extracted from the

signal as criteria for traditional fault diagnosis. However, acoustic signals have a lower signal-to-noise ratio than vibration signals and are susceptible to environmental noise interference in the acquisition process, which affects the accuracy of rolling bearing fault diagnosis [6]. Therefore, time-domain features (RMS and kurtosis) are widely used in bearing performance evaluation [7–9]. Chen et al. [10] proposed a time-domain statistical indicator based on the root mean square (RMS) value for use in rolling bearing condition identification; the RMS reflects the signal energy and is not effective relative to the weak early fault amplitude values. Yuan based on the time-domain statistical index of kurtosis [11] for rolling bearing condition identification, which has a high sensitivity to initial abnormal fault signals, but is prone to false alarms and has an impact on the reliability of assessing rolling bearing faults. RMS and kurtosis are the most common time-domain characterisation methods as they effectively reflect the real-time changes in rolling bearing operating vibrations. The complexity of rotating machinery

systems and the uncertainty of their operating conditions make them difficult to describe accurately with mathematical models.

The data-driven deep learning method can automatically learn rich and different features and have a strong feature extraction capability, which makes it possible to apply them to the initial fault study of rolling bearings. The method of traditional artificial neural network (ANN) cannot accurately diagnose bearing faults in the field of initial fault research of rolling bearings [12, 13]. Deep learning automatically learns representative features from rolling bearing vibration signals and gets rid of the dependence on signal processing technology. Lu et al. [14] proposed a new DAE model to diagnose bearing faults, which improved the ability and robustness of the hidden layer to extract features. Du et al. [15] entered the bearing vibration signal in the form of a time-frequency image into a convolutional neural network and achieved a diagnostic accuracy of over 90%. Chen and Li [16] input bearing vibration signals directly into deep sparse autoencoders performing automatic character extraction and analysis. A deep self-coding network for rolling bearing fault diagnosis incorporating advanced features was proposed by Shao et al. [13], which improved the robustness of feature extraction. Although all of the above research studies have adopted deep models to analyze rolling bearing faults more successfully, the models are independent of each other, ignoring the time series correlation between the data. Temporal convolutional network (TCN) [17] is a kind of network structure based on the convolution structure, which is specially used to process sequence information. It can not only grasp the overall information of the sequence from local to global but also use convolution instead of recursion, so it also has a great advantage in training speed. Chen et al. [18] used TCN to mine the time series characteristics of the bearing degradation trend, obtain the health index of the bearing, and predict the residual life of the bearing. Jaya-singhe et al. [19] combined two models of TCN and LSTM to construct a residual life prediction model for industrial equipment with good prediction results for data obtained from complex environments.

All of the above methods require a large amount of rolling bearing fault characteristic data, but in the area of initial fault diagnosis for rolling bearings, initial fault characteristic data are very scarce. Without abundant label data, it is very complicated to construct a rolling bearing initial fault prediction model. Transfer learning (TL) [20] breaks the assumption of independent identical distributions between different domains so that knowledge can be transferred through training data where the domain test data (target domain) are related but have differences in edge distributions or differences in conditional distributions. At present, transfer learning has been proved to be applied in machine vision [21], bioinformatics [22], object recognition [23], and natural language processing [24]. A domain adaptive machine fault diagnosis method based on a depth model was proposed by Li et al. [25], and the performance of the method was tested with rolling bearing datasets collected under different conditions of operation. Wen et al. [26] developed a deep transfer learning model to verify the validity of fault diagnosis with a dataset obtained at a bearing test rig operating under different load conditions. A feature-based transfer learning network (FTNN) approach was presented by Yang et al. [27]. The data from the laboratory rotating machinery are used to learn transferable features to identify the health status existing in the actual rotating machinery. Through the use of transfer learning methods, the above methods have achieved better results compared with other methods.

Based on the above research and the successful application of transfer learning in other fields, a rolling bearing initial fault prediction (TDCTCN) model that uses transfer learning and DCAE-TCN is proposed. In the fault feature extraction stage, the denoising autoencoder (DAE) is used as the first two layers, and the contractive autoencoder (CAE) is used as the last hidden layer to construct a deep autoencoder (DCAE) to gain the initial fault data for rolling bearings. The labelled source sample data and the unlabelled or with only a few labels target sample data are, respectively, used as the input to the feature extraction module to obtain the feature datasets ξ_s and ξ_t in the source samples and the target samples, respectively. In the domain adaptation stage, the feature dataset obtained in the previous phase is used as the input. The equilibrium factor μ of the two feature sets is calculated by A-distance, and the distribution adaptation (edge probability distribution or conditional probability distribution) method is chosen to first optimise that to minimise the distance between the source and target domain feature sets and then to construct the common feature set. In the rolling bearing early failure prediction stage, in order to predict the time of the initial fault occurrence, the obtained common feature set is used as the input to the TCN as the pseudo-label of the target data, the data temporal features are extracted from the target data, and the predicted value of the next moment is output to predict the operation status of the rolling bearing at the next moment. The experimental validation is conducted using the XJTU-SY full life cycle dataset. The experimental results showed that, on the basis of transfer learning and the DCAE-TCN model, the operating condition of the rolling bearing can be well described, and the initial fault of the rolling bearing can be predicted earlier and more stably, which is very important for the degradation monitoring of the rolling bearing running state.

Based on the above model analysis methods, the main contributions of this paper to the initial failure prediction of rolling bearings are as follows:

(1) We propose a rolling bearing initial fault prediction model based on transfer learning and DCAE-TCN. By combining the domain adaptive method of transfer learning with the deep autoencoder network and temporal convolution network, the data features of rolling bearings in the target domain can be extracted more accurately, and the evolution knowledge of rolling bearings is transferred to the

target domain for more accurate prediction of the next moment of operation of rolling bearings in the target domain.

(2) The DAE and the CAE are used as the first two implicit layers and the last implicit layer to construct the DCAE that can extract the data features of rolling bearings. Meanwhile, the temporal correlation between the data extracted by the TCN can extract more complete initial fault features and temporal sequence, and the improved model can make the accuracy and smoothness optimal.

(3) A balanced distribution adaptation approach is adopted. A common feature set is constructed between different domains by reducing the difference in the distribution of transferable features between the source and target domain datasets layer by layer. This common feature set is then used for model training of the TDCTCN network.

This paper is organised as follows. In Section 2, the basic theory of initial failure prediction of rolling bearings is briefly introduced. Section 3 introduces the model for initial fault prediction based on DCAE-TCN for transfer learning. Section 4 experiments the method on the basis of a full life cycle of a rolling bearing and analyses and discusses the experimental results of early failure prediction in rolling bearings. Conclusions of the full paper are given in Section 5.

2. Basic Methods

2.1. Description of Transfer Learning Problems. Transfer learning (TL) breaks the assumption of independence and identity between data from different domains and reduces the distance and class range between source and target samples' data by the domain adaptive method [28]. In this paper, the knowledge of rolling bearing operation evolution in the source samples is transferred to the target samples' data, and early failures of rolling bearings are predicted in the target domain. Assume that D^s and D^t are the source domain data and the target domain data, respectively. We will get a labelled source domain dataset $\{X_{si}, Y_{si}\}_{i=1}^{n}$ and an unlabelled target domain dataset $\{X_{tj}\}_{j=1}^{m}$. Assume a sample feature space $\chi_s \in D^s$, $\chi_t \in D^t$, and label space $\Upsilon_s = \{1, 2, 3 \ldots \ldots, k\}$ containing the evolution of rolling bearings from a normal state to k failure state. From the perspective of data generation, different data distributions have marginal distribution $P_s(X_s) \neq P_t(X_t)$ and conditional distribution $P_s(Y_s|X_s) \neq P_s(Y_t|X_t)$. The aim of TL is to transfer the learned knowledge from the source domain to the unknown target domain. Balanced distribution adaptation methods solve transfer learning problems by adaptively minimizing the variance in the distribution between the source and target domains and deal with the class imbalance problem, which minimizes the differences between $P_s(X_s)$ and $P_t(X_t)$ and between $P(Y_s|X_s)$ and $P(Y_t|X_t)$.

At the stage of transfer learning domain adaptation, knowledge of the health evolution of the rolling bearing in the source domain is transferred to the target domain to predict the next moment of operation of the rolling bearing in the target domain. The following assumptions should be met:

(1) All data include the evolution of degradation from normal to faulty states, and the evolution of the experimental data in the source and target samples from the normal state to the initial fault state is intrinsically consistent

(2) The source domain data have temporal association, which provides the necessary knowledge of evolution from the normal state to the initial fault state for the initial fault prediction of the target object

(3) For different working situations, the sample space of target domain data and derived data is $\chi_t \in D^t$, and the data have the characteristics of temporal association

Unlike fault diagnosis, the nature of initial fault prediction is more inclined to the problem of anomaly detection, that is, only the data at the start of the operational phase are marked as normal, and the time of appearance of an early fault is expected based on the deviation of the anomalous data from the normal data. Through iterative training with labelled data in the source domain, a nonlinear relationship $f: \chi_s \longrightarrow y_s$ from the sample space $\chi_s \in D^s$ to the labelled sample space Υ_s can be established, which is the predictive knowledge of initial faults in rolling bearings. Because the label data of the target domain are usually few or even missing, it is difficult to establish the prediction model $f: \chi_t \longrightarrow y_t$ for the initial fault prediction accuracy of rolling bearings in the target domain.

In this paper, knowledge of the prediction of initial fault of rolling bearings in the source domain is used to predict early failures of rolling bearings in the target domain. The functions of transfer learning in the proposed model are mainly reflected in the following: (1) through the principle of balanced distribution adaptation, the data distribution of the source and the target domains is adapted to establish the evolutionary knowledge representation from the normal state to the initial fault state. (2) With the help of the evolution knowledge and distinguishing information from the normal state to the initial fault state of the source domain, the effectiveness and robustness of initial fault prediction of rolling bearing data in target domain are promoted.

2.2. Balanced Distribution Adaptation (BDA). Balanced distribution adaptation (BDA) is an excellent tool for migration learning, which solves the migration learning problem by minimizing the difference in distribution between the source and target samples and deals with the class imbalance problem, that is, minimizing the difference

between $P(X_s)$ and $P(X_t)$ and $P(Y_s|X_s)$ and $P(Y_t|X_t)$. The balanced distribution adaptation can be adaptively adjusted to the probability distribution according to the relationship between the source and target samples:

$$
D(D_s, D_t) \approx (1 - \mu) D(P(X_s), P(X_t)) \\
+ \mu D(P(Y_s|X_s), P(Y_t|X_t)). \tag{1}
$$

In the above equation, $\mu \in [0, 1]$. When $\mu \longrightarrow 0$, the greater the difference between the datasets, the more obvious the marginal distribution. When $\mu \longrightarrow 1$, the more similar the datasets are, the more suitable the conditional distribution is than the marginal distribution.

It is also important to notice the impossibility to evaluate the conditional distribution because the target domain data have no or few labels. Therefore, using the class-conditional distribution $P(X_t|Y_t)$, $P(X_t|Y_t)$ is very similar to $P(Y_t|X_t)$ based on sufficient statistics for very large data samples. In the purpose of calculating $P(X_t|Y_t)$, the data with labels D^s are trained to obtain the pseudo-label of D^t and to make predictions for D^t.

For calculating the probability distributions in equation (1), the marginal and conditional distributions are estimated empirically using the maximum mean difference (MMD) method. As a nonparametric measure, MMD is widely used in existing migration learning algorithms, so equation (1) is rewritten in the following form:

$$
D(D_s, D_t) \approx (1 - \mu) \left\| \frac{1}{n} \sum_{i=1}^{n} X_{s_i} - \frac{1}{m} \sum_{j=1}^{m} X_{t_j} \right\|_{H}^{2} + \mu \sum_{c=1}^{C} \left\| \frac{1}{n_c} \sum_{X_{s_i} \in D_s^{(c)}} X_{s_i} - \frac{1}{m_c} \sum_{X_{t_{j}\in D_t^{(c)}}} X_{t_j} \right\|_{H}^{2}, \tag{2}
$$

where H is the regenerative kernel Hilbert space, $c \in \{1, 2, 3, 4 \cdots\cdots C\}$ denotes the count of labels, n, m are the number of samples, $D_s^{(c)}$ and $D_t^{(c)}$ belong to C and represent samples, and $n_c = |D_s^{(c)}|$ and $m_c = |D_t^{(c)}|$ belonging to $D_s^{(c)}$ and $D_t^{(c)}$ in the source and the target domains, respectively, denote the number of samples. The first and second parts indicate the distances of the edge distributions and the distances of the conditional distributions between different domains, respectively.

2.3. Deep Autoencoder Network (DCAE).

The autoencoder (AE) is a type of unsupervised feature representation network consisting of an input layer (x), a hidden layer (h), and an output layer (r). The aim for the AE is to minimise the differences between input and output data by encoding and decoding the reconstructed data. The DAE obtains robust reconstructed data from data samples that are corrupted or noisy. The CAE enhances the robustness of the learned features by adding penalty terms (Jacobi) to the loss function. In other words, the CAE enhances the robustness of feature learning by using the internal information of the data, mainly for the output of the data, and the DAE enhances the robustness of feature learning by using the external information of the data, mainly for the data input. Therefore, a CAE is combined with a DAE to build the DCAE to enhance the data feature learning capability.

Based on their own merits, DAEs and CAEs are applied to learn potential features from rolling bearing vibration data and to learn deeper features on top of potential features, respectively. Finally, the deep feature of learning is robust to the small disturbance of the input. Figure 1 is a structure diagram of a deep autoencoder network.

As shown in Figure 1, the DCAE is composed of three hidden layers: the first and second implied layers consist of the DAE and the third implied layer consists of the CAE. Feature learning is performed based on input and output data to increase the robustness of initial fault feature extraction for rolling bearings.

The encoding formula of the AE to the hidden layer is

$$
h = f_e(a_h) = \sigma(W_{xh}x + b_{xh}), \tag{3}
$$

where f_e is the encoding function for the hidden layer, σ is the activation function of sigmoid, W_{xh} is its weight, and b_{xh} is its bias.

The decoding formula of the AE to the output layer is

$$
r = f_d(h) = \sigma(W_{hr}h + b_{hr}), \tag{4}
$$

where $f_d(\ )$ is a decoding function to the output layer, W_{hr} is the weight, and b_{hr} is the bias.

The central idea of the AE is to keep the input data consistent with the output data through encoding and decoding, so the reconstruction error of the AE is

$$
J_{AE}(\theta) = L(x, r) = - \sum_{i=1}^{N} [x_i \log(r_i) + (1 - x_i)\log(1 - r_i)], \tag{5}
$$

where $L(x, r)$ is the calculation difference between coding and decoding loss functions.

The unprocessed vibration data contain a large amount of noise, so the noise-laden $\tilde{x}$ is used instead of the noise-free input data x. The encoding and decoding equations for the denoising autoencoder (DAE) are shown in the following:

$$
\tilde{h} = f_e(a_h) = \sigma(W_{xh}\tilde{x} + b_{xh}), \tag{6}
$$

$$
\tilde{r} = f_d(\tilde{h}) = \sigma(W_{hr}\tilde{h} + b_{hr}). \tag{7}
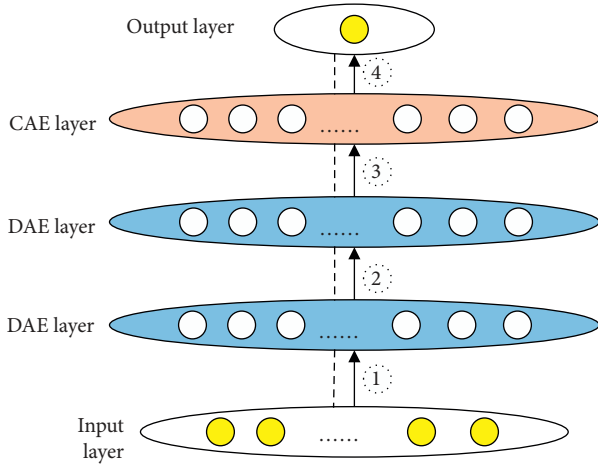$$

Therefore, the reconstruction error of the DAE is

FIGURE 1: Deep autoencoder network structure.

FIGURE 2: TCN structure diagram.

$$J_{\text{DAE}}(\theta) = L(x, \widetilde{r}) = -\sum_{i=1}^{N} \left[x_i \log(\widetilde{r}_i) + (1 - x_i) \log(1 - \widetilde{r}_i) \right]. \tag{8}$$

A contractive item is inserted into the loss function to improve the robustness of the CAE in learning information inside the data, and the reconstruction error of the CAE is

$$J_{\text{CAE}}(\theta) = L(x, r) + \lambda \|J_h(x)\|_F^2, \tag{9}$$

$$\|J_h(x)\|_F^2 = \sum_{ij} \left(\frac{\partial h_j(x)}{\partial x_i} \right)^2, \tag{10}$$

$$J_h(x) = \begin{pmatrix} \dfrac{\partial h_1}{\partial x_1} & \cdots & \dfrac{\partial h_1}{\partial x_N} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial h_M}{\partial x_1} & \cdots & \dfrac{\partial h_M}{\partial x_N} \end{pmatrix}, \tag{11}$$

where λ is the constriction factor, $\|J_h(x)\|_F^2$ is the constriction penalty item, and $J_h(x)$ is a Jacobian matrix.

2.4. Temporal Convolutional Neural Network. The temporal convolutional neural network (TCN), first proposed in 2018, adds residual blocks and inflated causal convolutions to the convolutional neural network (CNN) to obtain a temporal dilated convolution that possesses causality and longevity. Causality means that the current convolution result only depends on the previous and present input so that it can accept long-term historical information, while the long-term effect brings a large perception field for convolution so that it can build a time series network with memory ability. The TCN structure is illustrated in Figure 2.

Suppose the input data are $X = \{X_0, X_1, \cdots, X_{n-1}, X_n\}$ and output data are $Y = \{Y_0, Y_1, \cdots, Y_{n-1}, Y_n\}$, which have the same sequence length, and the output data Y_n only involve time n and sequence elements before n. Assuming
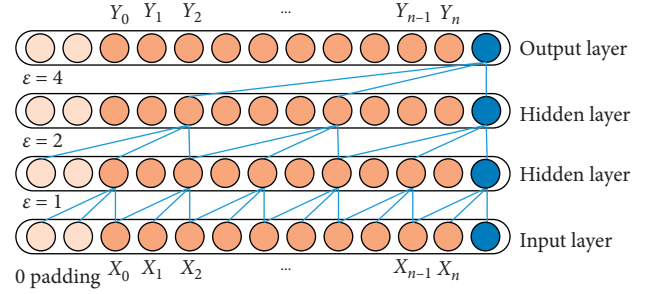
that the current time is n_0, the one-dimensional causal convolution replaces the input X_0 of all $n < n_0$ moments with 0:

$$f(x) = \begin{cases} x_n, & n < n_0 \\ 0, & n > n_0 \end{cases}. \tag{12}$$

In addition, the TCN also uses the expansion convolution method to increase the receptive field, and the expansion convolution transport is the convolution of the unit with a fixed gap. As shown in Figure 2, ε is the expansion factor, which increases with the number of network layers. It is obvious that when the expansive convolution is used, the expansion of the receptive field is much faster than that of the ordinary convolution.

3. The Proposed Method

We construct a rolling bearing early failure prediction model based on transfer learning and the integration of the DCAE and TCN to predict the initial faults for rolling bearings in this paper.

3.1. Overview of the Proposed Method. The proposed method comprises four parts (see Figure 3): domain division, feature extraction, balanced distribution adaptation, time series feature extraction, and initial fault prediction. In terms of domain division, the source domain contains evolutionary knowledge of various stages of the degradation process in the state of normal to failure, while the target sample provides no fault information. Thus, the source evolutionary knowledge is taken to predict the early fault of the target rolling bearing. In terms of feature extraction, a combination of the DAE and CAE is performed to extract the transferable features of source ξ_s and target ξ_t. In addition, the source and target datasets need to have the same or similar fault evolution process. In the balanced distribution adaptation part, the balance factor μ is obtained by calculating the ratio between the overall distance and the class distance between the transferable features of source ξ_s and target ξ_t and then calculating the nonparametric distance between their transferable features. The obtained similarity matrix of the target domains is used as the optimisation target, and backpropagation will be used to train the parameters of the nonlinear characteristic mapping and obtain the common feature set Y. In terms of temporal feature extraction and
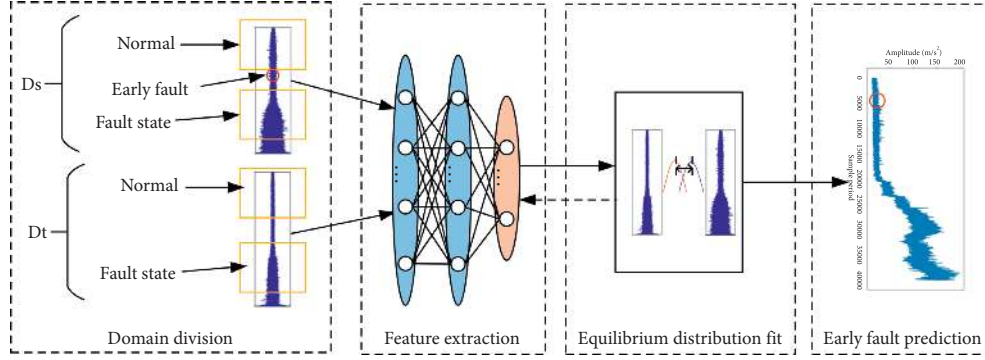
FIGURE 3: DCAE-TCN model structure based on transfer learning.

early fault prediction, the TCN extracts temporal features from the target rolling bearing data to predict the operating state of rolling bearings at the next moment. And based on the knowledge of the evolution of rolling bearings contained in the common feature set, the initial faults of rolling bearings are predicted accurately and efficiently.

3.2. Prediction Model Establishment.

The construction of the DCAE-TCN rolling bearing initial fault prediction model based on migration learning includes the feature extraction stage, domain adaptive stage, and initial fault prediction stage. In the feature extraction stage, the source and target sample data are input into the deep autoencoder stacked by DAE and CAE, respectively, so as to, respectively, affect the input and output of source and target domain data and extract weak early fault characteristic information. In the domain adaptation stage, for the feature information extracted from the source and the target domains in the previous stage, the BDA is used to reduce the distance between the feature information of the source and target domains, and then the common feature set is obtained. In the initial fault prediction stage, the TCN is taken to extract the time sequence features of the target sample data, and the common feature set obtained in the previous stage is taken to predict the next operating condition of rolling bearings in the target domain and the initial fault by using the evolution information from the normal state of the source data to the initial fault state. Figure 4 is a flowchart of initial fault prediction of rolling bearings.

3.2.1. Feature Extraction Stage

(1) Use the source domain dataset $\{X_{si}, Y_{si}\}_{i=1}^{n}$ and the target domain dataset $\{X_{tj}\}_{j=1}^{m}$ input to the DCAE model.

(2) The source domain dataset is iteratively trained according to formulas (6)–(11), and the weight W_s and bias b_s are updated in real time in the training process.

(3) Judge whether the accuracy meets the requirements or whether it reaches the set maximum number of training times. If not satisfied, repeat execution (2); if satisfied, stop the training.

(4) After the model is trained, save the extracted source domain data feature set ξ_s^3, weights, and biases of the model.

(5) Fix the trained DCAE model, and input the target data into the DCAE. The saved weights and biases are used as model parameters and will not be updated.

(6) Based on the feature extraction of the target data, the feature set ξ_t^3 of the target data is obtained.

3.2.2. Domain Adaptive Stage

(1) The balance factor μ is obtained by using the A-distance method to calculate between the transferable features of source ξ_s^3 and target ξ_t^3

(2) If $\mu \longrightarrow 1$, then the conditional probability distribution is preferentially fitted between the transferable feature sets; if $\mu \longrightarrow 0$, the marginal probability distribution is adapted first

(3) According to the distance between the source transferable feature set and the target transferable feature set, the target similarity matrix $R = \{R_0, R_1, \cdots, R_n\}$ is obtained

(4) According to the source label sample space Υ_s, the target pseudo-label is constructed, and the public sensitive feature set $F = \{F_0', F_1', \cdots, R_n'\}$ is established

3.2.3. Initial Fault Prediction Stage

(1) The common feature set Y obtained in the previous step is used as a pseudo-label and input into the TCN to perform temporal feature extraction on the target dataset to obtain the predicted value for the next moment.

(2) Determine whether the predicted value obtained exceeds the normal state threshold; if yes, determine the presence of an abnormal vibration signal. If not, repeat cycle (1).

4. Experimental Analysis and Prediction

4.1. Experimental Dataset.

In order to verify the effectiveness of the proposed method in the initial fault prediction of
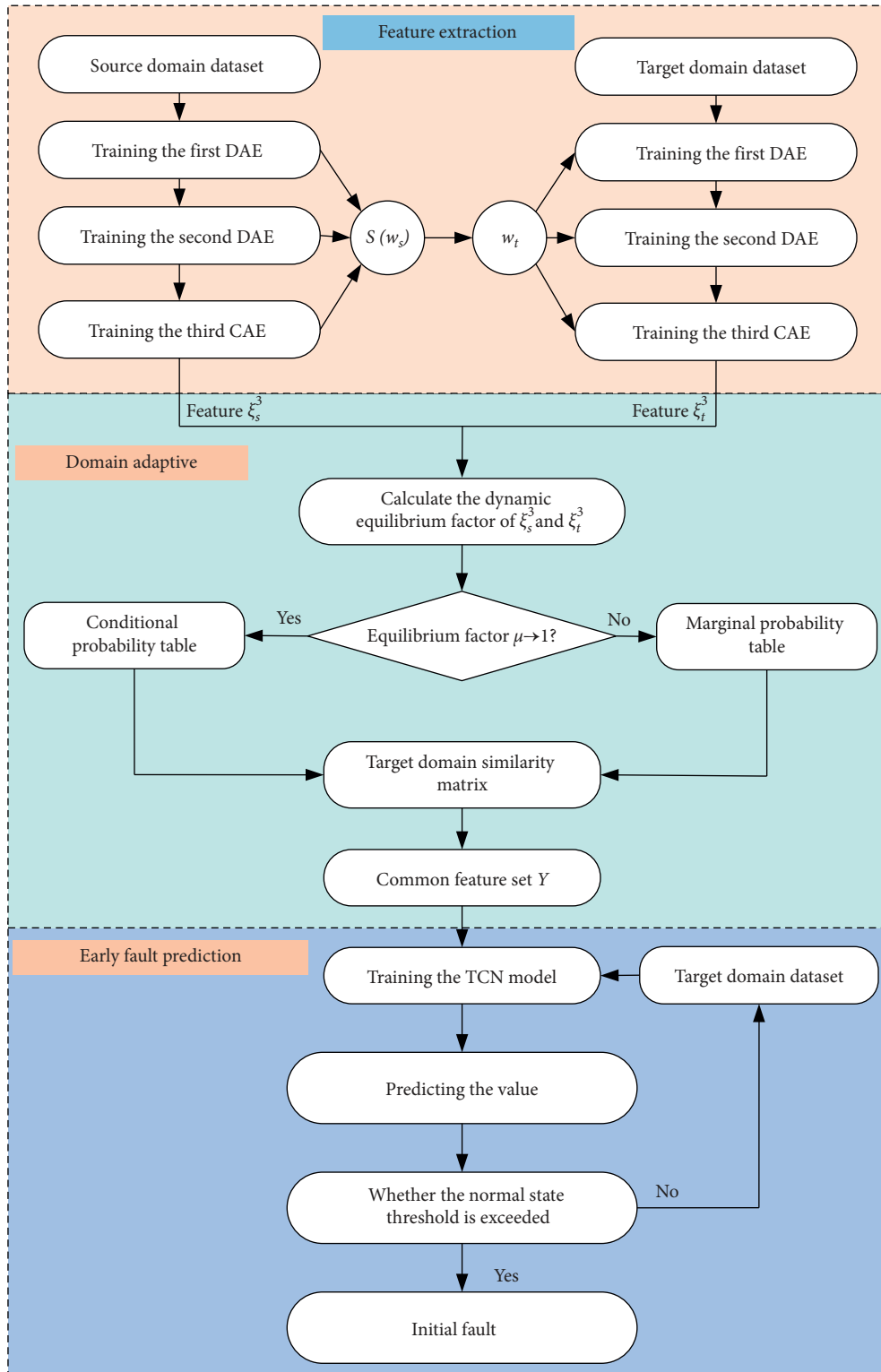
Figure 4: Flowchart of the rolling bearing initial fault prediction model.

rolling bearings, experiment and analysis are carried out by using the life cycle data of rolling bearings obtained from the rolling bearing test platform (XJTU-SY) [29]. Figure 5 shows a rolling bearing testbed, which consists of an AC motor, a motor speed controller, a supporting shaft, two supporting

bearings (heavy roller bearings), and a hydraulic loading system. The rotational speed is provided by the AC motor. A hydraulic loading system loaded onto the bearing seat being tested provides the system with radial forces. The platform can carry out speeding up the degradation tests of bearings

| Digital force display | Motor speed controller | Vertical accelerometer | Tested bearing |

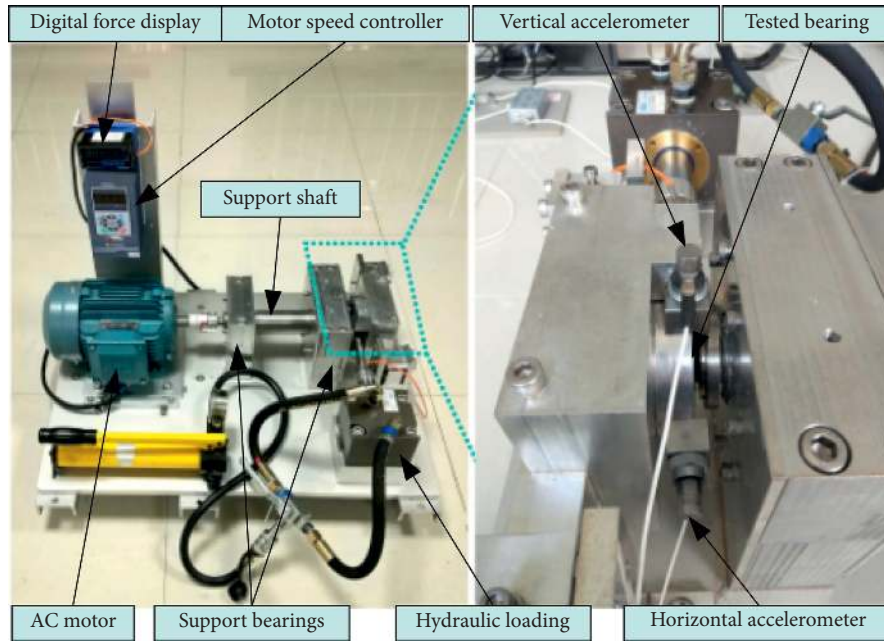| AC motor | Support bearings | Hydraulic loading | Horizontal accelerometer |

Support shaft

FIGURE 5: Rolling bearing test platform.

in different working states so that the full data of rolling bearings from operation to failure can be obtained.

As shown in Table 1, there are several sets of experiments under three working conditions. There are 5 bearings per operating condition, which are sampled at 25.6 kHz, with 32,768 data points per sample.

According to Table 1, two transfer learning tasks are created, that is, A $\longrightarrow$ B and A $\longrightarrow$ C. The A dataset is used as the source sample data to provide the evolution knowledge of rolling bearings, and the B and C datasets are taken as the target sample data. The reason for transfer learning is for predicting, as much as possible, the time period of initial faults of rolling bearings under B and C conditions.

4.2. Initial Fault Prediction of the TDCTCN Rolling Bearing. Bearing1_3 and bearing1_5 in experiment A are selected as the source domain data. Bearing2_3 and bearing2_5 in experiment B and bearing3_3 and bearing3_4 in experiment C are selected as target domain datasets. The original time-domain vibration signal plots for the whole life cycle of the four rolling bearings in the target domain are shown in Figure 6.

As seen in Figure 7(a), the time-domain response of bearing2_3 is relatively stable before about 410 samples (7 hours), and then the time-domain signal shows very obvious abnormal vibration, that is, the 2_3 rolling bearing is obviously damaged, but the damage degree of the rolling bearing cannot be judged. In Figure 7(b), the signal of bearing2_5 before about 230 samples (4 hours) is cross-stationary, then the time-domain signal shows an abnormal vibration phenomenon, and the rolling bearing is obviously damaged. In Figure 7(c), since the radial force of the C group experiment is smaller than that of the other two groups of

experiments, bearing3_3 and bearing3_4 run relatively smoothly, bearing3_3 shows obvious abnormality after about 440 samples (7.5 h), and the rolling bearing has obvious damage. In Figure 7(d), bearing3_4 shows obvious abnormal vibration and obvious damage after 1800 samples (30 hours), but like bearing2_3, the degree of damage to the roller bearing cannot be determined by the three bearings of bearing2_5, bearing3_3, and bearing3_4. As shown in the figure, the signal in the time domain can reflect the degradation trend to a certain extent for the rolling bearings, but it cannot well reflect the change process of initial faults of rolling bearings nor can it judge the degree of damage of rolling bearings. It is also impossible to judge when rolling bearings need to take corresponding measures to reduce losses.

The datasets of bearing2_3, bearing2_5, bearing3_3, and bearing3_4 in experiment B and experiment C were brought into the model for training. As shown in Figure 8, the DCAE-TCN model combined with transfer learning proposed in this paper describes the running states of the rolling bearings, as well as the prediction of the initial fault time of the rolling bearings and the classification of the running states of the rolling bearings.

As shown in Figure 8, the TDCTCN model can better reflect the running states of the rolling bearings and predict the time of the initial faults of the rolling bearings and can also judge the damage degree of the rolling bearings according to the acceleration. In Figure 8(a) bearing2_3 at about the 5500th sampling period, the first abnormal test data increased compared with the normal signal amplitude, the rolling bearing ran for about 3 hours, and then the data tended to normal. Compared with Figure 7(a), abnormalities were detected about 4 hours in advance. In about 20,000–25,000 sampling periods, the vibration of the test data increased significantly compared with the test data of

TABLE 1: Accelerated life condition of bearings.

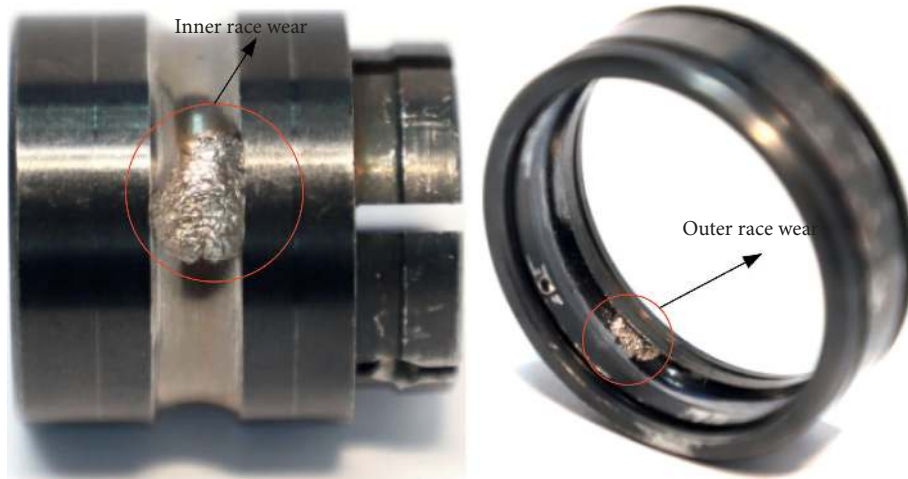| Experimental condition	Radial force (KN)	Rotational speed (r/min)	Bearing dataset
A	12	2100	Bearing1_1 Bearing1_2 Bearing1_3 Bearing1_4 Bearing1_5
B	11	2250	Bearing2_1 Bearing2_2 Bearing2_3 Bearing2_4 Bearing2_5
C	10	2400	Bearing3_1 Bearing3_2 Bearing3_3 Bearing3_4 Bearing3_5

FIGURE 6: Diagram of faulty bearings.

the first abnormal signal, and then there were no normal data, so it can be judged that a moderate fault has occurred. By analogy, a severe fault occurred in about 25,000–34,000 sampling periods, after which the signal vibration increased obviously, and the rolling bearing is in the failure stage. Similarly, bearing2_5 first detected abnormal data in about 7000th sampling period (bearing running time is about 2.7 h), which is about 1.3 h earlier than Figure 7(b). Bearig3_3 first detected abnormal data in about 25,000 sampling periods (bearing running time is about 6.8 h), which is about 1 hour earlier than Figure 7(c). Bearing3_4 first detected abnormal data in about 42,000 sampling periods (bearing running about 23.8 hours), about 6 hours earlier than Figure 7(d). Similar to bearing2_3, bearing2_5, bearing3_3, and bearing3_4 can also determine the damage degree of rolling bearings according to the comparison of signal amplitude with the normal state. The presence of abnormal vibration signals in rolling bearings is better judged by using the data characteristic of the normal condition of the rolling bearing as a control line. In summary, the TDCTCN model can well learn the existing knowledge of rolling bearings' evolution in the source domain and more accurately reflect the initial faults and degradation stages of rolling bearings.

4.3. Comparative Results and Analysis

4.3.1. TDCTCN Model versus Traditional Feature Extraction Metrics. To verify the effectiveness of the proposed method in this paper, the proposed method TDCTCN algorithm was compared with the time-domain statistical indicators

kurtosis, RMS, DCAE + RMS, and DCAE + kurtosis. Both the comparison method and the method proposed in this paper use a BDA to reduce the distribution differences and class spacing between the target and source data for early fault prediction of the target rolling bearing operating data. A graph of the comparison results is shown in Figure 9; the blue line represents the TDCTCN algorithm proposed in this paper, the green line shows the pending variable kurtosis statistics, the red line shows the table RMS statistics, the orange line represents the features extracted by DCAE and RMS, and the purple line represents the features extracted by DCAE and kurtosis.

Figure 9 depicts the progression of the four target bearing faults. As can be seen in Figure 9(a), an abnormal vibration signal appears near about 5500 more samples, and both kurtosis and the method proposed in this paper show higher than normal vibration signals. The other conventional time-domain indicators do not show any abnormal fluctuations in the vicinity of this sample. For kurtosis, it is very sensitive to abnormal changes in the signal. When a rolling bearing runs abnormally, there will be a transient spike in the vibration signal, so kurtosis will change significantly, but it will cause the diagnostic results to be unstable, and the chance of causing a false diagnosis will increase. At around 22,000 sample cycles, rolling bearings showed significant faults, where DCAE + RMS showed abnormal fluctuations, in line with the method proposed in this paper, but their vibration signals did not produce abnormalities at the time of the initial fault of the rolling bearing and only fluctuated at the later stages of the fault, so they cannot be used as an assessment indicator to give a good indication of the operating condition of the rolling bearing.
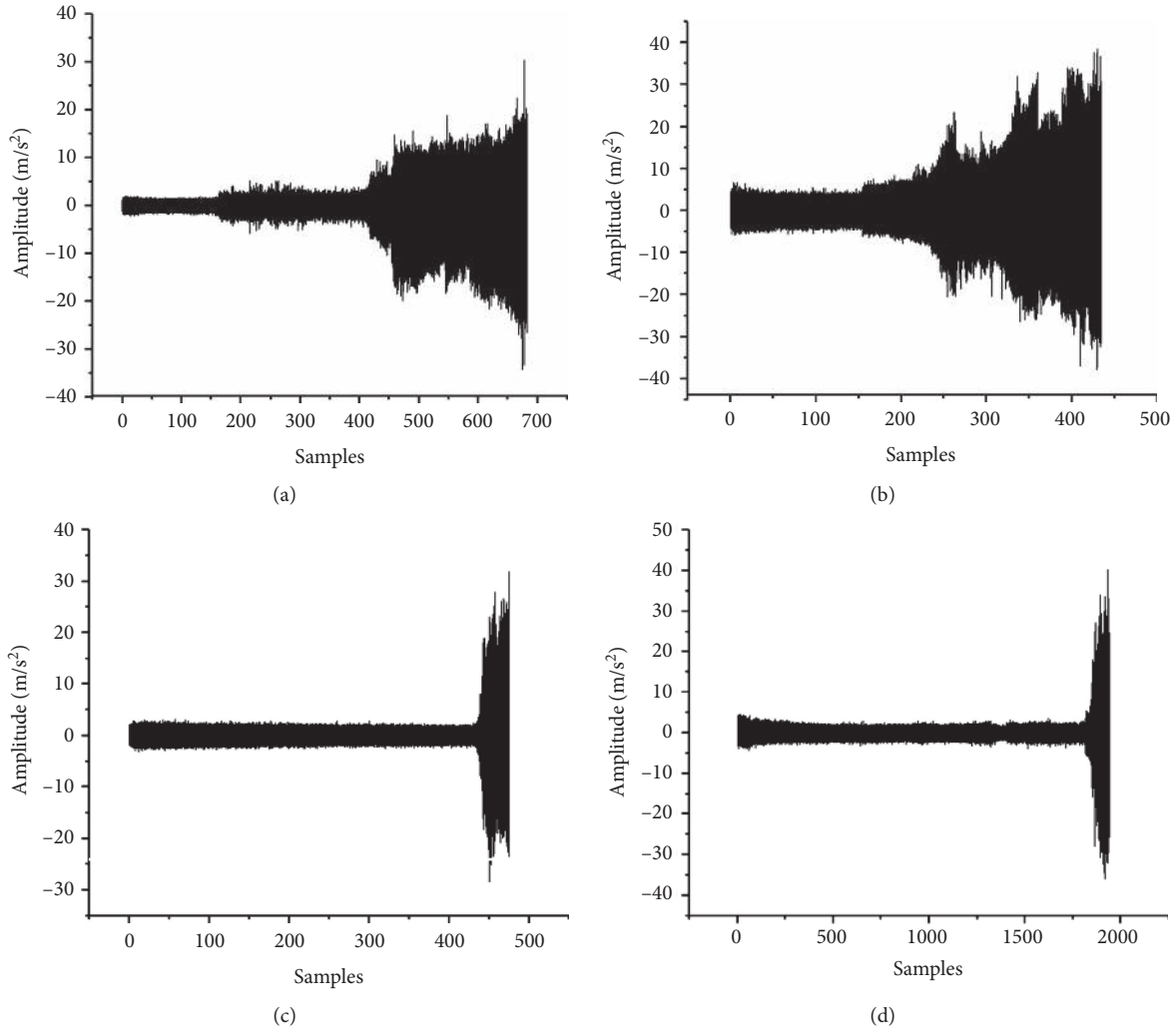
FIGURE 7: Plot of the original time-domain vibration signal for the full life cycle of a rolling bearing. (a) Bearing2_3. (b) Bearing2_5. (c) Bearing3_3. (d) Bearing3_4.

The other two methods, RMS and SDAE + kurtosis, are not sensitive to changes in the condition of the target rolling bearing. By analogy, similar results appear in Figures 9(b)–9(d). And the TDCTCN algorithm proposed in this paper can not only be very sensitive to the initial abnormalities of the target rolling bearing but also can reflect the degradation trend of the target rolling bearing at a later stage, and the signal of the rolling bearing is stable during the diagnosis process, so it can predict the initial fault of the target rolling bearing.

4.3.2. DCTCN Model without Transfer Learning. To verify the validity of the TDCTCN initial fault prediction model of rolling bearings and the necessity of transfer learning, we trained the data of the target domain with the DCTCN model without transfer learning. Relying on the fault character extraction ability of the deep autoencoding network and the time sequence feature extraction capability of the TCN network, the initial faults of rolling bearings were predicted. Figure 10 is a result graph of the training accuracy and loss rate of the DCTCN model without migration learning.

As shown in Figure 10, for the DCTCN model without transfer learning, the accuracy rises to near 1, and the loss rate decreases to around 0 after 30 iterations. The accuracy and loss rate fluctuate at the beginning, which means that the model is constantly trying to converge to the best point, and then the model is gradually stable. In summary, the DCAE-TCN model has high accuracy, which proves that the DCAE-TCN model has good feature extraction capabilities.

Figure 11 shows the description of the operating condition of the rolling bearings by the DCTCN model without transfer learning.

To validate the necessity of migration learning, the target sample data are directly imported into the deep autoencoder network for fault feature extraction, and then the trained data are taken as the input to the temporal convolution neural network for temporal feature extraction. Instead of
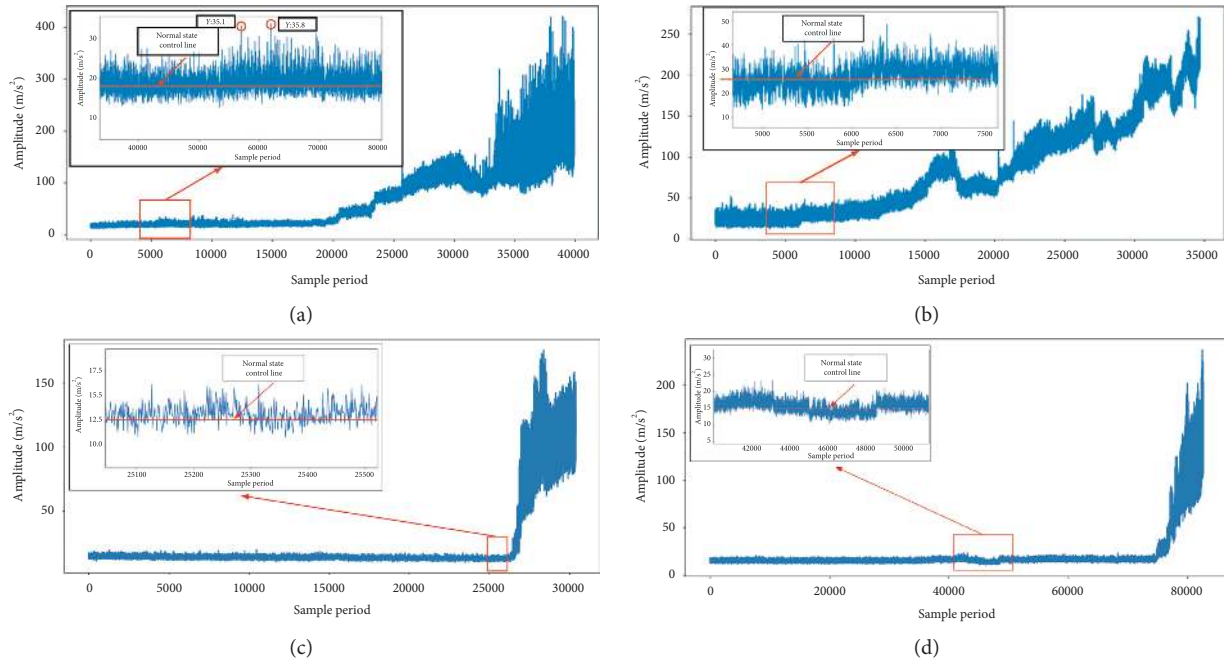
FIGURE 8: Description curve of the running state of rolling bearings. (a) Bearing2_3. (b) Bearing2_5. (c) Bearing3_3. (d) Bearing3_4.
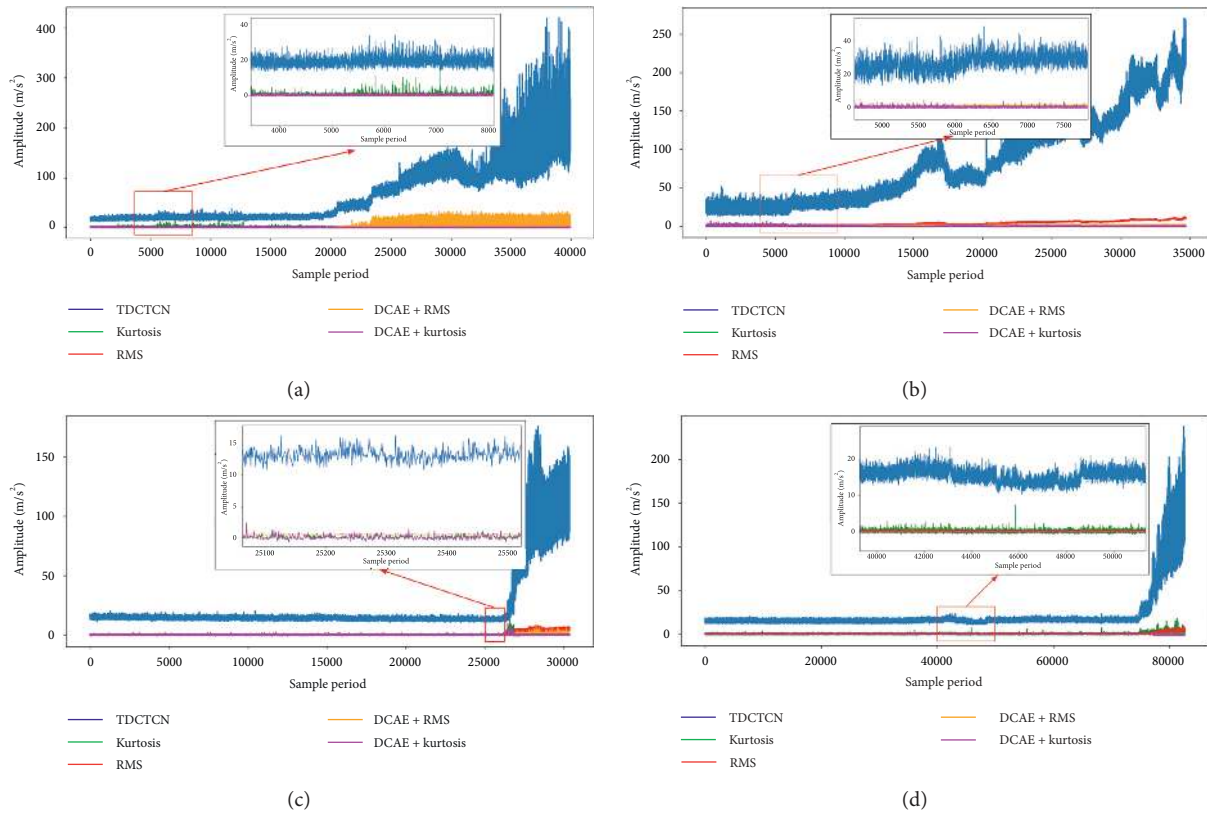

FIGURE 9: Description of the running state of rolling bearings. (a) Bearing2_3. (b) Bearing2_5. (c) Bearing3_3. (d) Bearing3_4.

transferring the existing rolling bearing evolution knowledge to the target domain, the DCAE-TCN model is used to predict the initial faults of the target bearing.

Figure 11 is a description of the operation status of rolling bearings in the target domain without transfer learning. Compared with Figure 8, Figure 11 is zoomed in in
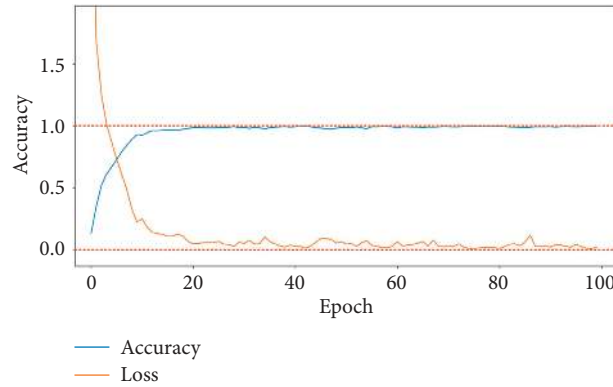
FIGURE 10: DCTCN model training accuracy.
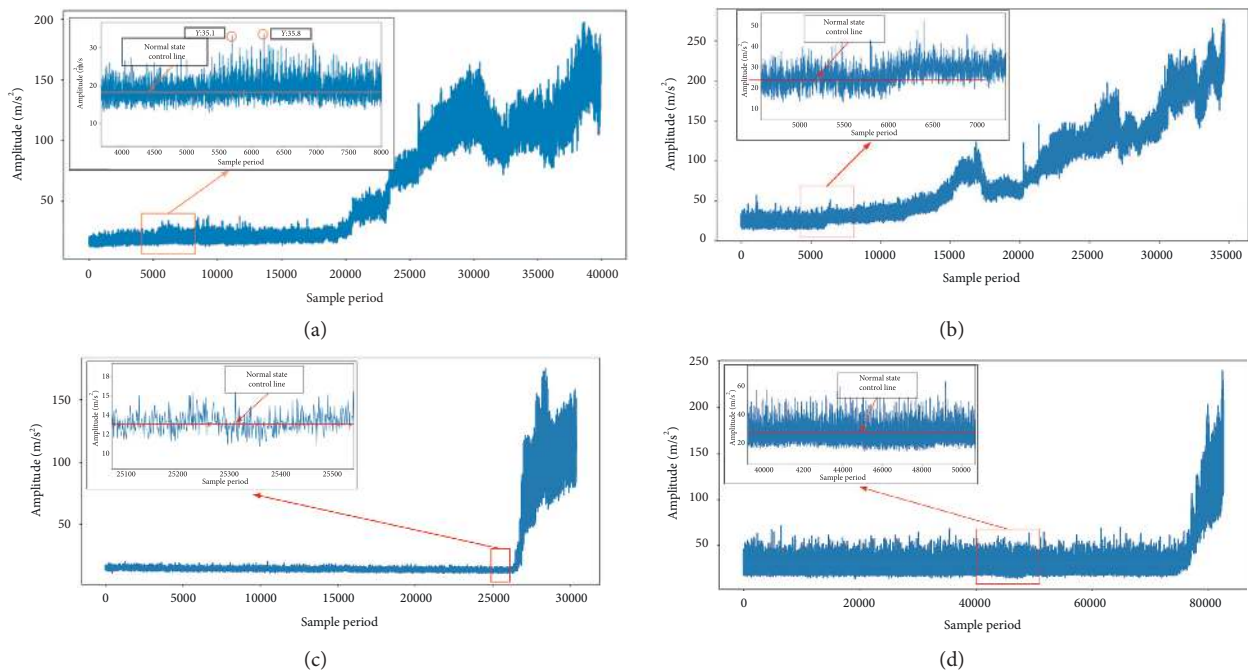

(a)

(b)

(c)

(d)

FIGURE 11: Description of the running state of rolling bearings. (a) Bearing2_3. (b) Bearing2_5. (c) Bearing3_3. (d) Bearing3_4.

the same place to more clearly reflect the presence of abnormal test data. In Figure 11(a), bearing2_3 has no abnormal detection data during about the 5500 sampling period, but abnormal test data appeared near the 18,000 sampling period, and the acceleration increased continuously compared with normal data until the rolling bearing failed completely. In Figure 11(b), the abnormal data of bearing2_5 first appeared near about 7000 revolutions' sampling period, but the amplitude of the abnormal data relative to bearing2_5 in Figure 7 is relatively small, then the acceleration increased continuously, and the abnormal test data appeared again in the sampling period of about 12,000 revolutions until the rolling bearing failed completely. In

Figure 11(c), the amplitude of the test data in about 25,000 sampling periods is somewhat larger than that of the normal data, but the increase is not very large, and then the data returned to a stable level, but the number of abnormal data continued to increase until the rolling bearing failed completely. In Figure 11(d), there are no abnormal test data around about 42,000 sampling periods. It can be seen that although the amplitude of the bearing2_3 and bearing2_5 signal can roughly determine the degree of damage of the rolling bearings, compared with Figure 7, it cannot effectively predict in advance. In addition, the fault characteristics of bearing3_3 and bearing3_4 are not obvious, that is, the damage degree of the rolling bearing cannot be clearly

shown. In summary, although the DCTCN model without transfer learning has higher accuracy and higher loss rate, it cannot well reflect the occurrence of initial faults of rolling bearings. It is further demonstrated that the proposed DCTCN model based on transfer learning is effective.

4.3.3. Quantitative Comparative Analysis. The migration accuracy of the proposed method is compared with transfer component analysis (TCA) [30], joint distribution adaptation (JDA) [31], transfer joint matching (TJM) [32], deep domain confusion (DDC) [33], and domain adversarial training of neural networks (DANN) [34].

TCA is a new regional adaptive feature extraction method as a comparison that learns a group of common migration components under two interrelated domains so that it is possible to project the data distribution differences in different domains into a subspace. The data distribution differences can be significantly reduced by the TCA method. JDA is a transfer learning method that deals with different probability distributions of the source domain and the target domain (no marked data). JDA extends the nonparametric maximum average deviation to measure the separation of marginal and conditional allocations. It is combined with principal component analysis (PCA) to construct an efficient and robust feature representation of distribution differences. TJM is an unsupervised domain adaptive joint distributed migration method. Its goal is to reduce domain differences through the principle process of dimensionality reduction, as well as joint matching features and cross-domain weights. And new feature representations are constructed, which are invariant to distribution differences and unrelated instances. In order to ensure the domain invariance of learning features, DDC adds adaptive layer and distribution matching term MMD to the structure of CNN. DANN adds other components of domain judgment to the deep neural network to learn the data characteristics provided by the source domain. As shown in Table 2, after selecting the optimal parameters for each method, 10 experiments were carried out, and the average values were taken.

In order to see the migration accuracy of each of the above methods more intuitively, as shown in Figure 12, the migration accuracy of each method is intuitively represented in a chart.

From Figure 12, we can see that the TDCTCN's accuracy is 80.56%, which is the highest among the six methods. Compared with JDA and TJM, TCA does not have joint distribution adaptation and lacks domain adaptation capabilities. Therefore, the accuracy of TCA is lower than that of the other five methods, which is 51.78%. The average accuracy of JDA is 53.18%. The average accuracy of TJM is 63.18%. Although these two methods adopt joint distribution adaptation, they cannot extract weak fault features from rolling bearing data samples, so their accuracy for initial fault prediction of rolling bearings is not as good as that of the method presented in this paper. Both DDC and DANN

TABLE 2: Transfer accuracy for transfer tasks (%).

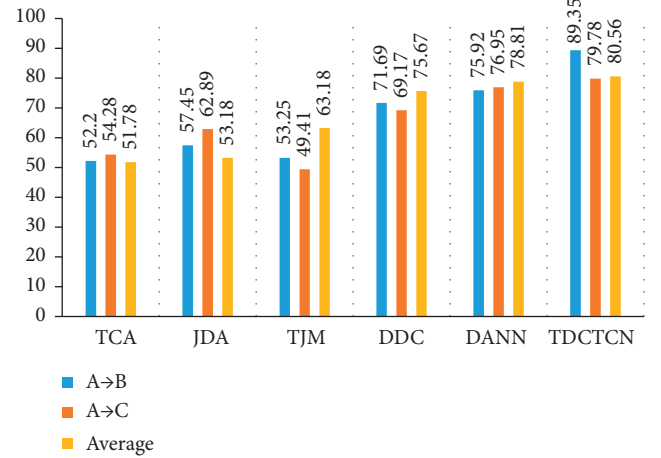| Transfer learning	A⟶B	A⟶C	Average
TCA	52.20	54.28	51.78
JDA	57.45	62.89	53.18
TJM	53.25	49.41	63.18
DDC	71.69	69.17	75.67
DANN	75.92	76.95	78.81
TDCTCN (ours)	89.35	79.78	80.56

FIGURE 12: Transfer accuracy for transfer tasks (%).

networks are deep learning and migration learning methods with the function of reducing the range between the source and target domains. Therefore, the migration accuracy is significantly higher than TCA, JDA, and TJM, but still inferior to the TDCTCN method. In summary, the validity and accuracy of the TDCTCN model in the initial fault prediction for rolling bearings are verified.

5. Conclusions

The initial fault vibration signal information of rolling bearings is weak. The weak fault characteristic information and strong noise interference make it difficult to effectively predict the initial faults of rolling bearings. Aiming at the problems, a TDCTCN rolling bearing initial fault prediction method based on transfer learning is proposed. The following conclusions were obtained:

(1) By combining the domain adaptive method of transfer learning with the deep autoencoder network and the temporal convolutional network, the data features of rolling bearings in the target field are accurately extracted. In addition, the evolution knowledge of rolling bearings in the source samples is transferred to the target sample, and the running state of target rolling bearings is accurately predicted.

(2) The experimental data verified by XJTU-SY show that the proposed TDCTCN method can predict the

initial faults of rolling bearings at least 1 hour in advance. For bearings running smoothly, the initial failure of rolling bearings can be predicted after 23.8 hours of operation.

(3) Compared with other traditional feature extraction methods, the method proposed in this paper can predict the early faults of rolling bearings earlier and has strong robustness. Compared with other transfer learning methods, the method has higher transfer accuracy. Compared with the DCTCN model without transfer learning, the proposed model has better stability and robustness, which is of great significance for predicting early faults and life cycle calculation of rolling bearings.

In the future research, we will combine the advantages of the proposed method with other intelligent diagnosis methods to carry out real-time online detection and diagnosis of early faults of rolling bearings.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Huaitao Shi summarized the research results and existing deficiencies in this field. A rolling bearing initial fault prediction model that was based on transfer learning and the DCAE-TCN was presented by Yajun Shang. The effective verification of the network simulation was carried out by Yajun Shang and Xiaochen Zhang. Yinghan Tang provided and analysed the experimental data of this article. Yajun Shang and Huaitao Shi summarized the findings and wrote this article.

Acknowledgments

References

[1] 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 Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1505–1518, 2020.

[2] J. Sanz, R. Perera, and C. Huerta, "Fault diagnosis of rotating machinery based on auto-associative neural networks and wavelet transforms," *Journal of Sound and Vibration*, vol. 302, no. 4-5, pp. 981–999, 2007.

[3] L. Song, H. Wang, and P. Chen, "Step-by-Step fuzzy diagnosis method for equipment based on symptom extraction and trivalent logic fuzzy diagnosis theory," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 6, pp. 3467–3478, 2018.

[4] M. X. Hou and H. T. Shi, "Stator-winding incipient shorted-turn fault detection for motor system in motorized spindle using modified interval observers," *IEEE Trans-actions on Instrumentation and Measurement*, vol. 70, 2020.

[5] Z. Li, Y. Jiang, C. Hu et al., "Difference equation based empirical mode decomposition with application to separation enhancement of multi-fault vibration signals," *Journal of Difference Equations and Applications*, vol. 23, no. 1-2, pp. 457–467, 2017.

[6] A. Glowacz, "Recognition of acoustic signals of induction motor using fft, smofs-10 and LSVM," *Eksploatacja i Niezawodnosc*, vol. 17, no. 4, pp. 569–574, 2015.

[7] Y. Hao, L. Song, L. Cui, and H. Wang, "A three-dimensional geometric features-based SCA algorithm for compound faults diagnosis," *Measurement*, vol. 134, pp. 480–491, 2019.

[8] L. Cui, B. Li, J. Ma, and Z. Jin, "Quantitative trend fault diagnosis of a rolling bearing based on Sparsogram and Lempel-Ziv," *Measurement*, vol. 128, pp. 410–418, 2018.

[9] J. Singh, A. K. Darpe, and S. P. Singh, "Bearing damage assessment using Jensen-Rényi Divergence based on EEMD," *Mechanical Systems and Signal Processing*, vol. 87, pp. 307–339, 2017.

[10] Z. Q. Chen, C. Li, and R. V. Sanchez, "Gearbox fault identification and classification with convolutional neural networks," *Shock and Vibration*, vol. 2015, Article ID 390134, 10 pages, 2015.

[11] Y. L. Yuan, "Fault Diagnosis of rolling bearing based on kurtosis-wavelet packet analysis," *New Technology & New Process*, vol. 5, pp. 43–46, 2008.

[12] S. Jurgen, "Deep learning in neural networks: an over-view," *Neural Networks*, vol. 61, no. 1, pp. 85–117, 2015.

[13] H. Shao, H. Jiang, and F. Zhao, "An enhancement deep feature fusion method for rotating machinery fault diagnosis," *Knowledge-Based Systems*, vol. 119, pp. 200–220, 2017.

[14] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Processing*, vol. 130, no. 1, pp. 377–388, 2017.

[15] X. L. Du, Z. G. Chen, G. N. Zhan et al., "Faults diagnosis of rolling bearings based on synchro squeezed S transform and deep learning," *Modular Machine Tool and Automatic Manufacturing Technique*, vol. 543, no. 5, pp. 90–93+97, 2020.

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

[17] Y. Chen, Y. Kang, Y. Chen, and Z. Wang, "Probabilistic forecasting with temporal convolutional neural network," *Neurocomputing*, vol. 399, no. 5, pp. 491–501, 2020.

[18] W. T. MaoJ. L. He et al., "Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1594–1608, 2019.

[19] L. Jayasinghe, T. Samarasinghe, C. Yuen et al., "Temporal convolutional memory networks for remaining useful life estimation of industrial machinery," *Machine Learning*, 2018.

[20] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[21] Y. Zhang, S. Miao, and R. Liao, "Structural domain adaptation with laten graph alignment," in *Proceedings of the 25th IEEE International Conference on Image Processing*, pp. 3753–3757, Athens, Greece, October 2018.

[22] V. Jayaram, M. Alamgir, Y. Altun et al., "Transfer learning in brain-computer interfaces," *IEEE Computa-Tional Intelligence Magazine*, vol. 11, no. 1, pp. 20–31, 2015.

[23] X. Chai, Q. Wang, Y. Zhao et al., "A fast, efficient domain adaptation technique for cross-domain electroencephalography (EEG)-Based emotion recognition," *Sensors*, vol. 17, no. 5, p. 1014, 2017.

[24] Z. Dai, Z. Yang, Y. Yang et al., "Transformer-XL: attentive language models beyond a fixed-length context," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, August 2019.

[25] X. Li, W. Zhang, and Q. Ding, "A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning," *Neurocomputing*, vol. 310, pp. 77–95, 2018.

[26] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 1, pp. 136–144, 2019.

[27] B. Yang, Y. G. Lei, F. Jia et al., "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," *Mechanical Systems and Signal Processing*, vol. 122, pp. 692–706, 2019.

[28] H. SHI, M. HOU, Y. WU, J. Guo, and D. Feng, "Leader-following consensus of first-order multi-agent systems with dynamic hybrid quantizer," *International Journal of Control, Automation and Systems*, vol. 18, no. 11, pp. 2765–2773, 2020.

[29] B. Wang, Y. G. Lei, N. P. Li et al., "A hybrid prognostics approach for estimating remaining useful life of rolling element bearings," *IEEE Transactions on Reliability*, vol. 69, no. 1, pp. 401–412, 2020.

[30] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[31] M. Long, J. Wang, G. Ding et al., "Transfer feature learning with joint distribution adaptation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2200–2207, Sydney, Australia, December 2013.

[32] M. Long, J. Wang, G. Ding et al., "Transfer joint matching for unsupervised domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition*, Columbus, OH, USA, June 2014.

[33] E. Tzeng, J. Hoffman, N. Zhang et al., "Deep domain confusion: maximizing for domain invariance," 2014, https://arxiv.org/abs/1412.3474.

[34] Y. Ganin, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.