Fusion with Joint Distribution and Adversarial Networks: A New Transfer Learning Approach for Intelligent Fault Diagnosis

Xueyi Li[1], Tianyu Yu[2], David He[3], Zhijie Xie[4], Xiangwei Kong[5]

[1,2,4] *College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin, 150040, China*
lixueyiphm@163.com
yty15052761682@126.com
xiezhijie111@sina.com

[3] *Department of Mechanical and Industrial Engineering, University of Illinois at Chicago, Chicago, IL 60607, USA*
davidhe@uic.edu

[5] *School of Mechanical Engineering and Automation, Northeastern University, Shenyang, 110819, China*
xwkong@me.neu.edu.cn

ABSTRACT

Bearings and gears are important components in rotating machinery, and the diagnosis of faults in bearings and gears has always been an important topic. Currently, data-driven fault diagnosis is a better method. However, under actual working conditions, domain shift can easily occur due to different operating conditions, leading to difficulties in transfer learning and significantly reducing the diagnostic performance of the model. Re-labeling the fault types of the model is time-consuming and costly. To overcome these difficulties, a new unsupervised transfer learning framework based on the fusion of joint distribution and adversarial networks has been introduced for the fault diagnosis of bearings and gears in rotating machinery. The joint adaptation network learns the transfer network by aligning the joint distribution of multiple specific domain layers across domains, based on Joint Maximum Mean Discrepancy (JMMD) to achieve domain alignment. At the same time, the domain classifier in the adversarial network is used to minimize the domain classification loss as domain distribution difference to minimize domain shift. The fusion of these two methods achieves domain alignment, reduces model training time, and improves the accuracy and stability of the model. The experimental results demonstrate that the proposed model framework exhibits excellent performance in detecting and classifying different types of faults. The new model framework also demonstrates outstanding performance across various fault detection and classification tasks.

1. INTRODUCTION

With the rapid development of manufacturing industry, rotary machinery has gradually become a key equipment in the mechanical manufacturing field, and higher requirements have been put forward for its intelligent maintenance [1]. The reliability and performance of bearings and gears are one of the key factors in mechanical system. For high-demand applications such as aerospace and medical equipment, the quality and reliability requirements for bearings and gears are higher. With the development of intelligent fault diagnosis field, methods based on deep learning have become popular. Janssens *et al*. [2] first applied convolutional neural network (CNN) to the fault diagnosis of bearings. Although traditional deep learning methods have achieved certain achievements in the field of fault diagnosis, past fault diagnosis methods are mainly based on the same working environment and conditions, which have great limitations. When facing the need for re-labeling due to different workloads, the model needs to be retrained, wasting a lot of time. Therefore, this paper proposes a diagnostic method based on transfer learning, which speeds up the learning process. Transfer learning can be mainly divided into four categories: (1) Network-based, (2) Instance-based, (3) Mapping-based, and (4) Adversarial-based [3]. Network-based fault diagnosis refers to directly transferring the parameters trained in the source domain as part of the testing process, or fine-tuning the network parameters using a small amount of labeled data in the target domain. Zhou *et al*. [4] proposed a deep learning framework combining transfer learning and transposed convolution, Li *et al* [5] processed data into images as inputs to the transfer learning model, and trained and fine-tuned VGG19. The basic concept of mapping-based method refers to mapping instances from the source domain and the target domain to the feature space through a feature extractor. Mapping-based methods mainly include: (1) Minkowski distance, (2) KL divergence (Kullback Leibler, KL), (3) CORAL alignment, (4) Maximum mean discrepancy (MMD), and (5) Multikernel MMD (MK-MMD). Qian *et al*. [6] proposed a new method for assessing distribution differences called Auto-balanced High-order Kullback-Leibler (AHKL) divergence, which can evaluate first-order and high-order moment differences and automatically adjust the weights between them. They also developed Smooth Conditional Distribution Alignment (SCDA), which aligns the conditional distribution by introducing soft labels instead of using widely used pseudo-labels. Zhu *et al*. [7] calculated domain loss through a linear combination of multiple Gaussian kernels, which results in greater adaptability compared to a single kernel. These two methods reduce distribution differences and learn transferable features. The effectiveness of the proposed methods was validated through experiments on transfer fault diagnosis.

Adversarial-based fault diagnosis refers to reducing the feature distribution differences between the source and target domains by utilizing a domain discriminator. Ganin *et al*. [8] proposed the Domain Adaptation by Backpropagation (DANN) architecture, which equates the loss of the domain discriminator in GAN to the distribution distance between domains.

Compared with traditional fault diagnosis methods, whether based on single-kernel MMD, multi-kernel MK-MMD, or DANN architecture, they were all proposed to solve $P(X_s) \neq Q(X_t)$, based on the common assumption of $P(X_s \mid Y_s) = P(X_t \mid Y_t)$. This assumption is relatively strong and does not conform to actual working conditions, posing higher requirements for training process. Therefore, this paper proposes a new assumption, $D(P_s(\mathbf{X}_s, Y), P_t(\mathbf{X}_t, Y))$ $\approx D(P_s(Y \mid \mathbf{X}), P_t(Y \mid \mathbf{X})) + D(P_s(\mathbf{X}), P_t(\mathbf{X}))$, based on the conditions that the conditional and marginal distributions are different. In order to address this issue, a multilinear adjustment is proposed to capture the inter-covariance between feature representation and classifier prediction for improved discriminability, and an entropy adjustment is proposed to control the uncertainty of classifier prediction for ensuring transferability. Finally, a framework combining JMMD and adversarial networks is proposed. We proposed an effective method to enhance transfer learning models for fault diagnosis. At the same time, we conducted comparative experiments to compare the proposed method with other commonly used transfer learning methods. The proposed method is verified on the gear dataset from Northeast Forestry University and the bearing dataset from Jiangnan University, and the results show that the proposed method effectively improves the stability and accuracy of the transfer learning model for fault diagnosis.

2 Preparatorwork

2.1 Problem definition

Unsupervised deep transfer learning with overlapping categories is defined as follows: it assumes that the labeled data from the source domain is available to predict the unlabeled data from the target domain. In this study, we focus on the case where the fault categories of the source and target domains are the same. We introduce some basic notation and assume that the labels in the source domain are available. The source domain is defined as follows:

$$\mathsf{D}_s = \left\{\left(x_i^s, y_i^s\right)\right\}_{i=1}^{n_s} \quad x_i^s \in \mathbf{X_s}, \quad y_i^s \in \mathbf{Y_s}$$

We define some basic symbols, where D_s represents the source domain, $x_i^s \in \mathsf{R}^d$ denotes the *i*th sample, $\mathbf{X_s}$ is the union set of all samples, y_i^s represents the i th label of the *i*th sample, $\mathbf{Y_s}$ is the union set of all different labels, and n_s represents the total number of samples in the source domain. Additionally, assuming that the labels of the target domain are not available, the following definitions are made for the source domain.

$$\mathsf{D}_t = \left\{\left(x_i^t\right)\right\}_{i=1}^{n_t} \quad x_i^t \in \mathbf{X_t}$$

In the above equation, D_t represents the target domain, $x_i^t \in \mathsf{R}^d$ is the i th sample, $\mathbf{X_t}$ is the union of all samples, and n_t is the total number of target domain samples. Fig.1 shows the comparison before and after domain adaptation.

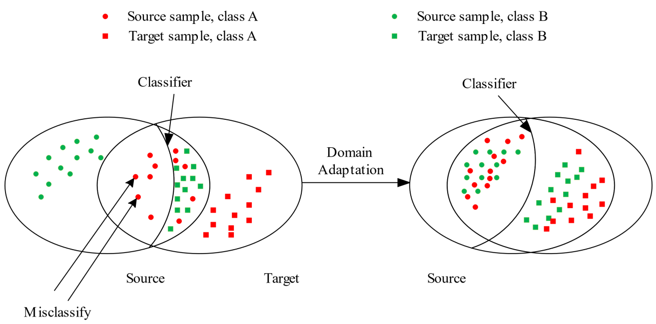

Fig 1. Domain adaptive before and after comparison chart.

2.2 JMMD(Joint Maximum Mean Discrepancy

JMMD is a method used to measure the distance between joint distributions $P(X_s,Y_s)$ and $Q(X_t,Y_t)$, as shown in Fig.2, which is a system diagram of a joint distribution network structure. Many methods aim to solve the transfer learning problem by adding the difference between the marginal distributions $P(X_s)$ and $Q(X_t)$ of the source and target domain errors. MMD, as a kernel two-sample test statistic, has been widely used to measure the distribution differences between $P(X_s)$ and $Q(X_t)$.The JMMD formula is as follows, as proposed in [9]:

$$\mathcal{L}_{\text{JMMD}}(P,Q) = \Big\| \mathbb{E}_P\left(\otimes_{l=1}^{|L|}\phi^l\left(z_l^s\right)\right) - \mathbb{E}_Q\left(\otimes_{l=1}^{|L|}\phi^l\left(z_l^t\right)\right)\Big\|^2_{\otimes_{l=1}^{|L|}\mathcal{H}^l}$$

which $\otimes_{l=1}^{|L|}\phi^l(z_l) = \phi^1(z_1)\otimes\cdots\otimes\phi^{|L|}(z_{|L|})$ is the feature map in the tensor product Hilbert space, L is the higher-level network architecture, $|L|$ denotes the layer number, z_l^s represents the l th layer activation generated by the source domain, and z_l^t represents the l th layer activation generated by the target domain.

Therefore, the total loss function of applying the JMMD method to the field of bearing fault diagnosis is shown in Eq.4, where $\mathcal{L}_c$ represents the classification loss and λ_{JMMD} represents the coefficient for measuring distance using the JMMD method.

$$\mathcal{L} = \mathcal{L}_c + \lambda_{\text{JMMD}}\mathcal{L}_{\text{JMMD}}\left(\mathcal{D}_s,\mathcal{D}_t\right)$$

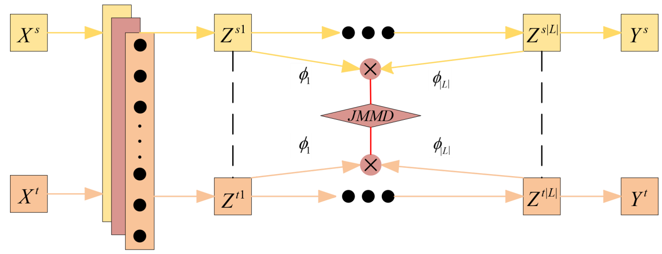

Fig.2. The architectures of Joint Adaptation Network.

2.3 Conditional Adversarial Domain Adaptation

The key to the Conditional Domain Adversarial Network (CDAN) [10] model is a novel conditional domain discriminator, which takes the cross-covariance between the domain-specific feature representation and the classifier prediction as a condition. Further, the condition of the domain discriminator is placed on the uncertainty of the classifier prediction, and the discriminator is prioritized for examples that are easy to transfer. The entire system can be solved in linear time through backpropagation [11]. To understand the CDAN structure, we first define some basic symbols. Firstly, we need to define a multilinear map $\otimes$, which represents the outer product of multiple random vectors. If given two random vectors x and y, the average mapping $x\otimes y$ can capture the complex multimodal structure within the data. In addition, cross-covariance can successfully model the joint distribution $P(x,y)$, therefore, the loss for conditional adversarial is defined as follows:

$$w(H(p)) = 1 + e^{-H(p)}$$

$$\mathcal{L}_{\text{CDAN}}(\theta_f,\theta_d) = -\mathbb{E}_{x_i^s\in\mathcal{D}_s}\log[G_d(G_f(x_i^s)\otimes G_c(G_f(x_i^s)))] \\ -\mathbb{E}_{x_i^t\in\mathcal{D}_t}\log\left[1-G_d\left(G_f(x_i^t)\right)\otimes G_c\left(G_f(x_i^t)\right)\right)$$

In order to reduce prediction uncertainty, entropy criterion $H(p) = -\sum_{c=0}^{c-1} p_c \log p_c$ is used to define the uncertainty of the classifier prediction. p_c corresponds to the predicted probability of the true label c. According to the entropy-aware weight function in Eq. 5, difficult-to-transfer samples are reweighted with lower weights in the modified conditional adversarial loss in Eq.6.

Through the aforementioned adversarial training, the feature generator can generate feature distributions that are similar to the input data, while the domain discriminator has stronger domain discrimination ability.

$$\mathrm{L}(\theta_f,\theta_c,\theta_d)=\mathrm{L}_c(\theta_f,\theta_c)-\lambda_{\mathrm{CDAN}}\mathrm{L}_{\mathrm{CDAN}}(\theta_f,\theta_d)$$

2.4 Proposed Methods

The proposed approach in this paper combines the two methods presented in Sections 2.2 and 2.3. One is based on mapping, and the other is based on adversarial training, both of which have achieved some success in transfer learning. However, they are still far from sufficient. The main goal of this paper is to improve the model's performance by combining these two methods to achieve breakthroughs in the field of fault diagnosis. Fig.3 illustrates the flowchart of the proposed method, where the total loss is obtained by combining Eq.4 and Eq.6.

$$\mathrm{L}=\mathrm{L}_c+\lambda_{\mathrm{JMMD}}\mathrm{L}_{\mathrm{JMMD}}\left(\mathrm{D}_s,\mathrm{D}_t\right)-\lambda_{\mathrm{CDAN}}\mathrm{L}_{\mathrm{CDAN}}(\theta_f,\theta_d)$$

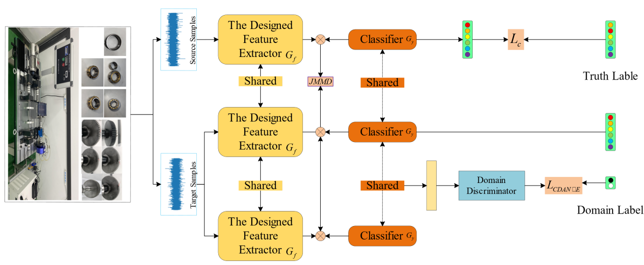

Fig.3 . This paper presents the model architecture diagram.

3. Experimental Validation

The experiment in this study uses both bearing and gear datasets to fully verify the advantages of the proposed model that combines JMMD and domain adversarial CDAN algorithms. By comparing with traditional transfer learning algorithms, In comparison to conventional methods, the model framework exhibits higher accuracy and stability in detecting and classifying different types of faults.

The experiment is designed to compare the proposed method with two existing domain adaptation methods: Deep Adaptive Network (DAN) based on MK-MMD and Conditional Adversarial Domain Adaptation with Entropy minimization (DANN+E) based on CDAN. Six groups of comparison experiments are conducted: Method 1, a composite method based on JMMD and CDAN+E; Method 2, a composite method based on MK-MMD and CDAN+E; Method 3, a method based on JMMD; Method 4, a composite method based on MK-MMD; Method 5, a method based on CDAN+E; and Method 6, a baseline method without using any transfer learning method.

3.1 Dataset Description

The two datasets mainly used in this paper are the Jiangnan University Bearing Dataset (JNU) and the Northeast Forestry University Gear Dataset (NEFU), which are described in detail below.

(1) Jiangnan University Bearing Dataset

The JNU bearing dataset has a sampling rate of 50 kHz. This study conducted fault diagnosis tests using two types of rolling bearings, N205 and NU205, to obtain signals under normal and faulty conditions. N205 bearings were used for normal, outer ring defect, and rolling ball defect states, while NU205 bearings were used for inner ring defect states. These faulty bearings were manufactured using a wire cutting robot, and Table 1 lists the specifications, fault sizes, and other necessary information of the test bearings. Each type of fault data (including normal bearings) was sampled at three different speeds: 600, 800, and 1000 rpm. Three different transfer learning conditions were established based on these speeds. 600 rpm corresponds to G1, 800 rpm corresponds to G2, and 1000 rpm corresponds to G3. This paper constructed a total of six transferable states for experimental validation in the JNU bearing dataset: G1→G2、G1→G3、G2→G3、G2→G1、G3→G1, and G3→G2.

Table 1. Bearing information for verification.

Contents	N205	Nu205
Bearing outer diameter	52mm	52mm
Bearing outer diameter	25mm	25mm
Bearing width	15mm	15mm
Bearing roller diameter	7mm	7mm
The number of the rollers	10	11
Contact angle	0 rad	0 rad
Out-race defect(width×depth)	0.3×0.25 mm Early stage	
Rolling element defect(width×depth)	0.5×0.15 mm Early stage	
Inner-race defect(width ×depth)		0.3×0.25 mm Early stage

3.2 Experimental Validation Using the Bearing Dataset from the JNU

Case 1:six groups of experiments were set up in this study, compared with the method proposed in this paper. The accuracy is shown in Table 2, the confusion matrix is shown in Fig.4, and the F1-score is shown in Fig.6. This score takes into account both precision and recall, providing a balanced measure of the model's classification performance across different fault types. In addition, the t-SNE plot of G1→G2 in Method 1 was used to visualize the classification process of model for different faults is shown in Fig.5.

Furthermore, the results presented in Table 2 demonstrate the superior performance of our proposed method across various transfer scenarios. With accuracies of 97.79%, 96.92%, 98.70%, 99.18%, 88.72%, and 99.72%, our approach consistently outperformed other existing methods. This improvement in accuracy highlights the effectiveness and reliability of our method in handling different types of faults.

Moreover, Fig. 6 provides a visual representation of the F1-Score for each fault type in different transfer states. Notably,

the red bars representing our method (Method 1) consistently exhibit higher F1-Scores compared to the other colors. This observation further supports the robustness and efficacy of our proposed method in accurately classifying and identifying fault types.

In essence, the effectiveness of our method is mainly due to the combination of JMMD and CDAN+E, which measures the distance between domains at multiple scales. JMMD evaluates the visible distance between domains, while CDAN+E evaluates the implicit distance between domains. Compared with traditional MK-MMD, JMMD considers both the marginal and conditional distributions. Although MK-MMD is an improvement of the MMD method, the MK-MMD used in this paper is still based on the assumption of the same conditional distribution. This assumption will have different negative effects under different working conditions, especially in cases of significant differences in working conditions. CDAN+E quantifies the uncertainty of the classifier's predictions through entropy criterion $H(g) = -\sum_{c=1}^{C} g_c \log g_c$, where C is a quantity and g_c is the probability of predicting class C. Each training factor of the conditional domain discriminator is reweighted.

Table 2. The accuracy rate in each migration state in Case1.

Task	G1→G2	G1→G3	G2→G1	G2→G3	G3→G1	G3→G2
Method1	**97.79**	**96.92**	**98.70**	**99.18**	**88.72**	**99.72**
Method2	97.54	95.38	93.44	98.55	88.64	99.43
Method3	97.52	96.26	96.46	98.83	87.79	98.92
Method4	96.95	95.04	97.22	97.29	87.07	98.98
Method5	94.72	96.06	80.64	96.75	84.75	98.76
Method6	95.76	89.93	57.28	96.03	78.03	96.53

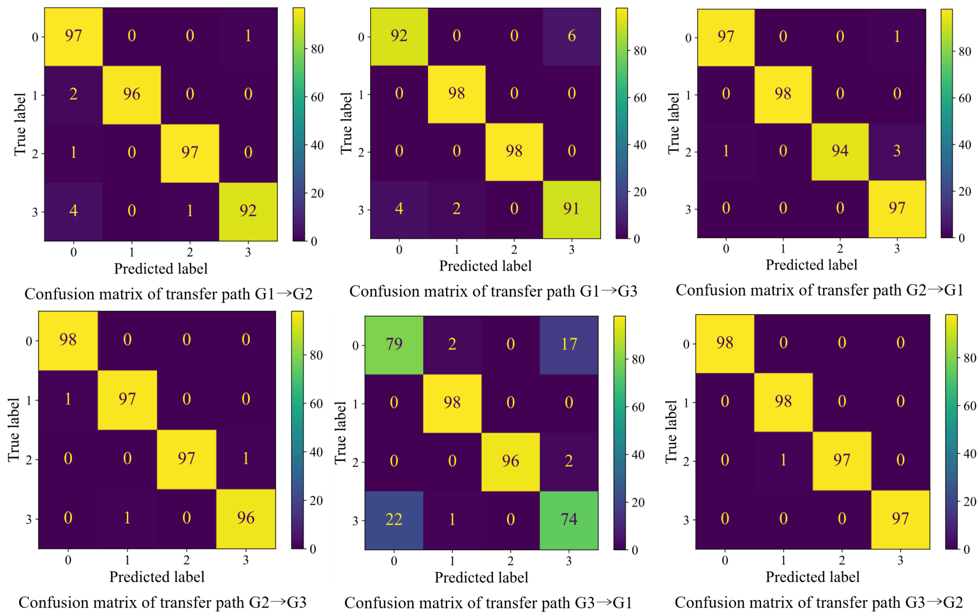

Fig. 4. Confusion matrices for the proposed method in Case1.

Fig. 5. Feature visualization based on t-SNE for the proposed method in Case1.

4 Conclusion

Unsupervised cross-domain fault diagnosis has always been a research challenge. In order to further improve the accuracy of unsupervised cross-domain diagnosis and enhance the model's cross-domain performance and generalization ability, this paper proposes a diagnostic model based on the combination of JMMD explicit distance and CDAN+E implicit distance. Firstly, the joint distribution properties of JMMD are used to solve the weak conditional assumptions of traditional methods based on approximating conditional distributions or marginal distributions to be equal, which can be more closely related to real industrial applications.

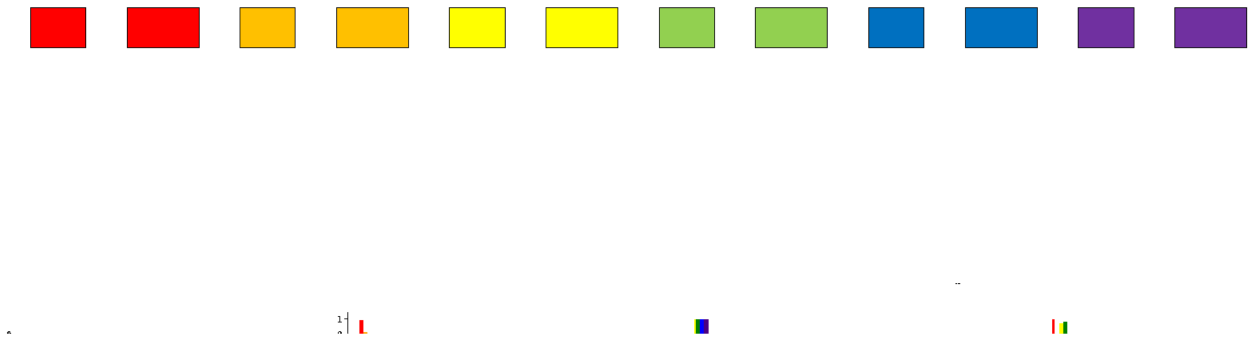

Fig. 6. F1-score evaluation results for each comparative method in Case 1.

In order to further improve the performance of the model, CDAN+E is introduced to solve the problem of adjusting classification in the original multi-modal distribution of different domains in traditional domain adversarial models. Finally, through the fusion of architectures and verification on two datasets, the effectiveness of the proposed method is demonstrated. Unsupervised learning is the mainstream direction for the future development of fault diagnosis, and it is also a challenging direction. Solving problems in the target domain without labels is difficult. Moreover, the limitation of this study is that the fault types in the source and target domains are completely the same. Therefore, we plan to conduct research in the open domain direction, which can classify even when faced with fault features that have not appeared in the source domain.

Acknowledgement

This work is supported in part by the Fundamental Research Funds for the Central Universities(No.2572022BF07) and in part by the Key Laboratory of Vibration and Control of Aero-Propulsion System, Ministry of Education, Northeastern University (VCAME202209).

References

[1] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," Advances in adaptive data analysis, vol. 1, no. 01, pp. 1-41, 2009.

[2] O. Janssens, V. Slavkovikj, B. Vervisch et al., "Convolutional neural network based fault detection for rotating machinery," Journal of Sound and Vibration, vol. 377, pp. 331-345, 2016.

[3] Z. Zhao, Q. Zhang, X. Yu et al., "Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-28, 2021.

[4] J. Zhou, X. Yang, L. Zhang et al., "Multisignal VGG19 network with transposed convolution for rotating machinery fault diagnosis based on deep transfer learning," Shock and Vibration, vol. 2020, pp. 1-12, 2020.

[5]Y. Li, N. Wang, J. Shi et al., "Adaptive batch normalization for practical domain adaptation," Pattern Recognition, vol. 80, pp. 109-117, 2018.

[6] W. Qian, S. Li, and X. Jiang, "Deep transfer network for rotating machine fault analysis," Pattern recognition, vol. 96, p. 106993, 2019.

[7] J. Zhu, N. Chen, and C. Shen, "A new deep transfer learning method for bearing fault diagnosis under different working conditions," IEEE Sensors Journal, vol. 20, no. 15, pp. 8394-8402, 2019.

[8] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in International conference on machine learning, 2015: PMLR, pp. 1180-1189.

[9] M. Long, H. Zhu, J. Wang et al., "Deep transfer learning with joint adaptation networks," in International conference on machine learning, 2017: PMLR, pp. 2208-2217.

[10] M. Long, Z. Cao, J. Wang et al., "Conditional adversarial domain adaptation," Advances in neural information processing systems, vol. 31, 2018.

[11] X. Li, K. Su, Q. He, et al., "Research on Fault Diagnosis of Highway Bi-LSTM Based on Attention Mechanism," Eksplo atacja i Niezawodność – Maintenance and Reliability, vol. 25, no. 2, 2023.