Proceedings of the Sixth International Conference on
Railway Technology: Research, Development and Maintenance
Edited by: J. Pombo
Civil-Comp Conferences, Volume 7, Paper 28.1
Civil-Comp Press, Edinburgh, United Kingdom, 2024
ISSN: 2753-3239, doi: 10.4203/ccc.7.28.1

Varying Speed Diagnosis of High Speed Train Bogie Rolling Bearing: Real Train Experiment, Comparison and Prediction of Transfer Learning Performance

B. Yang[1] , T Wang[1], J. Xie[2]

[1]College of Mechanical and Vehicle Engineering,
Hunan University, China
[2]School of Traffic & Transportation Engineering,
Central South University, China

Abstract

Due to the increasing speed of high-speed trains, intelligent fault diagnosis of train bogies is facing new problems: the diagnostic model for lower speeds is no longer applicable, and the lack of fault labels for high speed data makes the training of deep learning diagnostic models even more difficult. In order to be able to reliably diagnose unlabeled signals of different speeds using a limited number of labeled fault samples, this paper conducts an experiment and comparison of the effects of transfer learning on bearing faults of a real train. experiments and migration learning effects are compared. The axle bearing failure simulation experiment was conducted on a real high-speed train car using a rolling test bed, and the monitoring data with fault labels at different speeds were obtained; then the cross-migration was conducted using multiple speed monitoring data and multiple migration learning methods to obtain the cross-speed migration learning fault diagnosis effect; finally, the comparison of the distributional differences and the migration learning effect dataset was used to ensure the migration learning model could accomplish the monitoring at higher speeds. model can accomplish higher speed monitoring data migration diagnosis. At the same time, this paper uses a variety of signal pre-processing methods, network models and migration learning methods in the proposed framework for comparison, further verifying the feasibility and stability of the prediction method, and gives the optimal application of reference suggestions.

Keywords: deep transfer learning, domain adaptation, variable working conditions diagnosis, high-speed train bogie, axle bearing diagnosis, deep learning

1 Introduction

The rapid development of machine learning methods enables data-driven fault diagnosis methods to overcome the limitations of professional knowledge and manual analysis and to be widely used for monitoring industrial machinery that generates huge amounts of data[1, 2]. However, for high-speed trains, there are still difficulties in the application of intelligent fault diagnosis methods. Monitoring data for high-speed trains have two distinct characteristics[3, 4]: Firstly, high-speed train running usually contain multiple stable speed operating intervals with large speed differences that caused by vary railway line conditions and running dispatching strategies. Secondly, early failures and minor level damages are more difficult to be detected that caused by complex structure of high-speed train bogies and more conservative maintenance strategies. These characteristics of monitoring data make high-speed train bogie bearing diagnosis a label less (only have labels in some specific working conditions) and multiple working condition fault classification problem[5].

Earlier intelligent diagnosis research has achieved distinguished model performance under sufficient labeled data condition, but when there are few labels and many working conditions such as the diagnosis of axle box bearings in high-speed trains, high diagnostic accuracy cannot be guaranteed, and even effective fault diagnosis cannot be completed[6]. For this problem, the domain transfer learning approach has the unique advantage of mapping the source domain (labeled samples) and the target domain (unlabeled samples) to the same feature space and aligning the distributions, so that the pre-trained model of the source domain can be used to diagnose the target domain[7, 8].Li et.al proposed [9] a domain adversarial graph convolutional network to achieve unsupervised variable working condition diagnosis. Feng et.al[10] used meta-learning to achieve cross-domain transfer learning under few shot condition. Pan et.al[11] proposed a generating network to replenish imbalanced dataset. Yang et.al[2] designed a an adversarial network architecture named deep partial transfer learning network to alleviate model training problems caused by few fault sample.

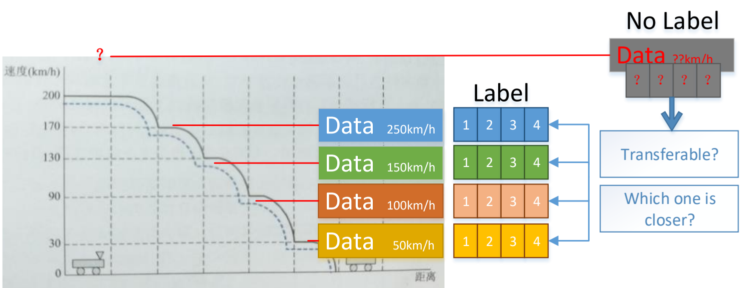


Figure 1 The unlabeled transfer learning challenge for higher speed work data

With the increase in the speed of high-speed trains, the number of working conditions has increased significantly, and the large-scale traversal transfer learning consumes a large amount of computational resources, resulting in poor real-time performance. the diagnostic model for lower speed is no longer applicable, and the lack of fault labels for high-speed data makes it more difficult to train deep learning diagnostic models[12,13]. The strong nonlinearity and lack of interpretability in feature extraction by neural networks, as well as the unknown transferability between domains, further complicate the issue. Moreover, for higher-speed trains/lines, the application time is short, and there is insufficient or even no fault data collected[14]. Therefore, how to guide the selection of source domain data and evaluate the credibility of the results before transfer learning becomes a major issue in the application of transfer learning methods.

In this paper, a novel transfer learning bearing fault diagnosis method is proposed for higher speed working condition, and its contributions are as follows:

1) A real high-speed train fault axelbox bearing experiment was conducted to obtain the vibration monitoring data under various speed working condition that is close to real train running scenarios.

2) By applying various data processing method and transfer learning method, the transfer learning diagnosis performance and pretrained models between different speed conditions were obtained, showcasing the uncertainty of transfer between operating speed conditions.

3) The proposed diagnostic framework is validated using experimental data and pre-trained models in each speed domain, and the experiments prove that the proposed method can predict the transfer performance, select the optimal speed conditions in the source domain, and complete the high-precision bearing fault diagnosis.

2 Methods

Based on problem definition of transfer learning, we give a definition for varying speed fault diagnosis of axle-box bearing in high-speed train: There are a dataset D^s with multiple speed working conditions, where the sub-dataset of the ith working condition is $D_i^s = \{x_k, y_k\}_{k=1}^{N_s}$. Using dataset D_i^s, an effective classifier net M_i can be obtained. There is an unknown rotational speed dataset $D_j^t = \{x_l\}_{l=1}^{N_t}$ and does not contain labels, how to select the pre-trained model and transfer learning method to achieve accurately classification?

In view of the proposed problems, the adaptive transfer learning strategy model is proposed, which is divided into three parts: feature distribution difference calculation, cross-transfer learning and ensemble learning mapping relation fitting. Among them, the feature distribution difference calculation is to use the existing pre-training for cross-computing, use the current working condition pre-training model to extract features of other working condition data, and calculate the distribution difference between the features of the two working condition data. The cross-transfer learning part is to use the pre-training model of the current rotational speed condition data to transfer learning other rotational speed conditions (assuming no labels) to obtain the diagnostic accuracy in each cross-task. Finally, the distribution difference is fitted to the diagnostic accuracy of transfer learning, and the feature-transfer learning effect mapping model is obtained. When there is unlabeled data for unknown rotational speed conditions, the existing pre-trained model can be used to extract features and predict the classification accuracy of transfer learning. By predicting performance and selecting source domain data that can best achieve successful transfer learning, the diagnosis accuracy of the target domain can be improved and guaranteed.

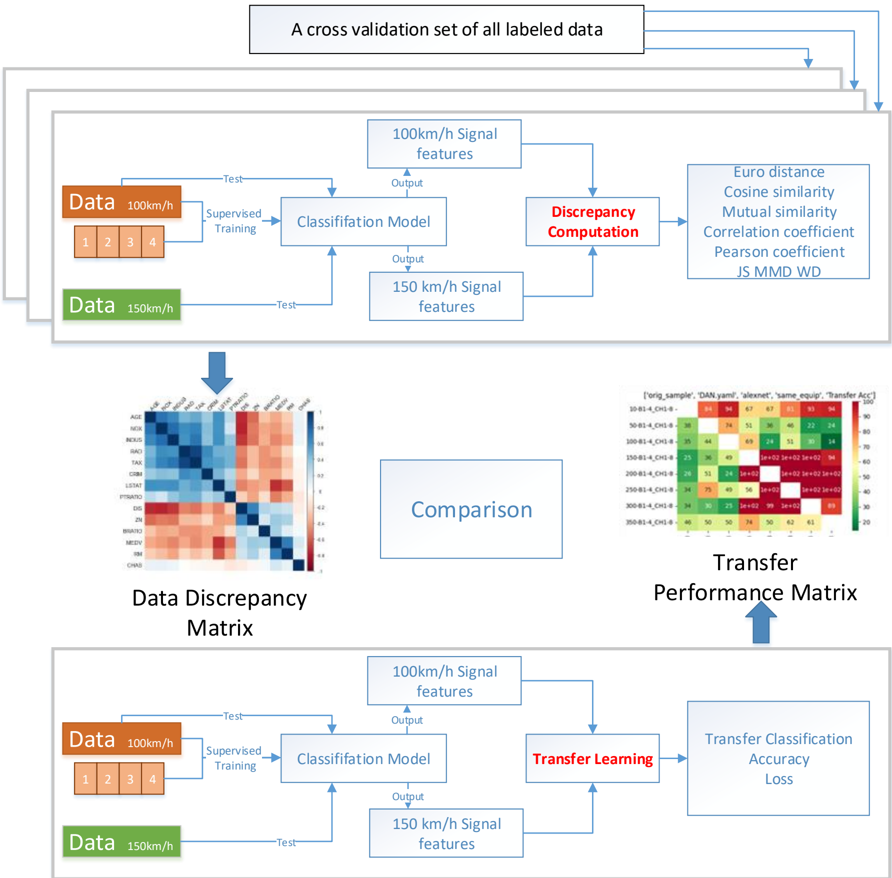


Figure 1 the Framework of proposed method

1) Cross domain transfer learning

The monitoring system obtains vibration data at constant rotational speed for each speed condition with sufficient labeling information. In order to evaluate and investigate the transferability of data between individual speed conditions, it is necessary to pre-train a classification model with data from one condition as a labeled source domain, and subsequently train the pre-trained model with data from one of the other conditions as an unlabeled target domain for transfer learning. After cross domain traversing the entire dataset containing n working conditions using this process, n pre-trained models, $n \times (n-1)$ transferred models and classification accuracies are obtained. From the classification accuracies of cross-domain transfer learning, it is possible to assess the transferability between different working conditions, and at the same time, the feature extraction models that can accurately classify under different working conditions are obtained.

2) Distribution discrepancy

From the distribution of data distribution differences, it can be observed that even if the monitoring data are affected by factors such as background noise masking and component response coupling, the distribution of operating conditions differences is still similar to the theory. However, after feature extraction by the nonlinear system of neural network, the distribution of classification accuracy has changed greatly. Therefore, the transferability between different working conditions is not intuitive and linearly related to the difference in speed. In the proposed method, the distribution difference between features extracted by the pre-trained model is calculated using the features extracted by the pre-trained model as input. The metrics used include similarity[15]: Pearson, correlation coefficient, cosine similarity, mutual information; Distance measures: Euclidean distance, Wasserstein distance; Other measures: JS divergence, MMD. Using multiple difference measures to construct the difference data set of different rotational speed conditions, you can show the differences in geometry, statistical distribution and other aspects of features, and contain more information to obtain better regression effect. For two probability distributions X, Y, there are two samples $x \in X, y \in Y$, and the difference vector of the samples is calculated as:

$$Discrepancy\ tensor = [d_{Euclidean}, cos, I, \rho, D_{KL}, JSD, MMD, WD]$$

$$d_{Euclidean} = \sqrt{(\boldsymbol{x}-\boldsymbol{y})^T(\boldsymbol{x}-\boldsymbol{y})}$$

$$\cos(\boldsymbol{x}-\boldsymbol{y}) = \frac{\boldsymbol{x}\cdot\boldsymbol{y}}{|\boldsymbol{x}|\cdot|\boldsymbol{y}|}$$

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} p(\boldsymbol{x},\boldsymbol{y})\log\frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{x})p(\boldsymbol{y})}$$

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

$$JSD(P||Q) = \frac{1}{2}D_{KL}(P||M) + \frac{1}{2}D_{KL}(Q||M), M = \frac{1}{2}(P+Q)$$

$$MMD^2(X,Y) = \left\|\frac{1}{n_1}\sum_{i=1}^{n_1}\phi(\boldsymbol{x}_i) - \frac{1}{n_2}\sum_{j=1}^{n_2}\phi(\boldsymbol{y}_j)\right\|_{\mathcal{H}}^2$$

$$WD_1(\mathbb{P},\mathbb{Q}) = \sup_{||f||_L\leq 1}\mathbb{E}_{x\sim\mathbb{P}}[f(\boldsymbol{x})] - \mathbb{E}_{x\sim\mathbb{Q}}[f(\boldsymbol{y})]$$

Which Cov stands for covariance, σ represents standard deviation, and $\varphi()$ represents the regenerative Hilbert spatial map.

3 Real Train Experiment Setup

The fault test system mainly consists of the whole vehicle rolling test bench and the fault bearing. The whole vehicle rolling test bench ensures the consistency of the test conditions and the real vehicle conditions, and the fault bearing disassembled by the real vehicle ensures the consistency of the test object and the real object, the monitoring data obtained by this test system is consistent with the vibration response of a real axle-box bearing failure in a real vehicle. The whole vehicle rolling test bench and its structure principle are shown in Figure 3.

In this test, the statistical high incidence and high risk of failure were studied by using NTN CRI-2692 double row tapered roller bearings of high speed train. The normal, indentation and stripping failure specimens were selected and mounted on two sets of bogies of the whole rolling test rig. The specific bearing specimen information is shown in Table 1

Table 1 Table of bearing positions and damage states

Bearing Type	No.	Installation Location	Damage State	No.	Installation Location	Damage State
NTN CRI-2692	B1	1st axel left	Normal	B5	3th axel left	Normal
	B2	2nd axel left	Outer ring indentation	B6	4th axel left	Outer ring indentation
	B3	1st axel right	Outer ring peeling	B7	3th axel right	Outer ring peeling
	B4	2nd axel left	Corrosion failure	B8	4th axel right	Corrosion failure

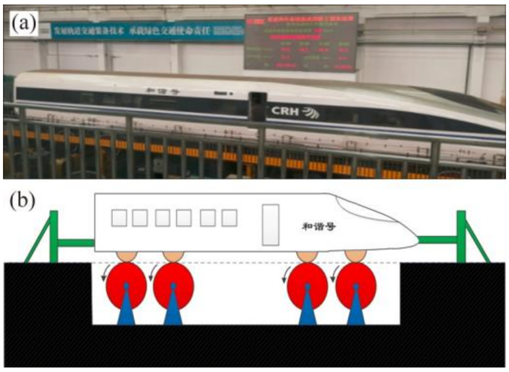


Figure 2 Structure of Complete train rolling test platform

The test object covers 8 bearings, and 2 measurement points are arranged on each bearing, totaling 16 sensors. In the test process, the vehicle wheel pairs are driven by the whole vehicle rolling test bench to uniformly accelerate to a number of set speeds and then maintain a period of uniform operation. The complete test design is shown in Table 2, including 7 test conditions with the speed range of 0~350 km/h, covering the speed of the mainline operation.

Table 2 Test working condition table

No.	Speed (km/h)	Sampling time	Sample rate	RPM of axel (r/s)
1	50	20s	25.6k	308.43
2	100			616.87
3	150			925.31
4	200			1233.75
5	250			1542.19
6	300			1850.63
7	350			2159.07

4 VALIDATION EXPERIMENTS

Due to the powerful feature extraction ability of convolutional neural networks, this paper constructs a neural network with five convolutional layers and three fully connected layers based on the classic neural network Le-Net. The feature extractor uses one-dimensional convolutional layers, and in order to better extract sparse features in the signal, larger convolutional kernels (kernel size=128, 64) are set in the first and second layers[16]. The pooling layer adopts maximum pooling. Dropout layers are set in odd layers to alleviate overfitting. For specific network parameters, please refer to Table 3.

Table 3 Network backbone structure and parameters

Feature Extractor(CNN)

Layer	Filters	Kernel size	Stride	Activation function
Conv1	16	1×128	1	ReLU
Pool1	-	4	4	
Dropout1		Rate: 0.5		
Conv2	32	1×64	1	ReLU
Pool2	-	4	4	
Conv3	64	1×16	1	ReLU
Pool3	-	4	4	
Dropout3		Rate: 0.5		
Conv4	128	1×4	1	ReLU
Pool4	-	2	2	
Conv5	256	1×2	1	ReLU
Pool5	-	2	2	
Dropout5		Rate: 0.5		
		Flatten		

Label Classifier			Domain Discriminator		
Layer	Number of neurons	Activation function	Layer	Number of neuron	Activation function
FC1	128	ReLU	FC1	128	ReLU
FC2	64	ReLU	FC2	64	ReLU
FC3	Class number	Softmax	FC3	1	-

In order to evaluate the performance of different deep transfer learning methods on multi-condition tasks, this paper selected classic transfer learning methods such as Deep Adaptation Network (DAN)[17, 18], Deep Correlation alignment for deep domain adaptation (DeepCoral)[19], Domain-adversarial neural network (DANN)[20], Deep Subdomain Adaptation Network (DSAN)[21], Dynamic Adversarial Adaptation Network (DANN)[22] and Batch Nuclear-norm Maximization (BNM)[23] for network training. Due to the different requirements of different transfer learning methods for network structure, this paper set up three network modules: feature extractor, label classifier, and domain discriminator. The specific parameters are detailed in Table 3. The ensemble learning module uses the Gradient Boosting Regress method to construct regression trees, with specific parameters detailed in **Hiba! A hivatkozási forrás nem található.**.

A. Transfer learning performance under working condition transversal

The data of seven different working conditions of the two bogies were cross-transfer learning, with a total of 42 transfer tasks, using six transfer learning methods and two datasets from different bogies. The average transfer accuracy, classification accuracy of pre-trained models, and improvement accuracy of 42 transfer tasks in each method are calculated in the following table, as shown in the Figure 3 and **Hiba! A hivatkozási forrás nem található.**.

From the perspective of data form, because the domain training classification accuracy of envelope spectral data is higher than that of time domain raw data, even if the transfer improvement effect of time domain raw data is better, the transfer accuracy of envelope spectrum data is still higher

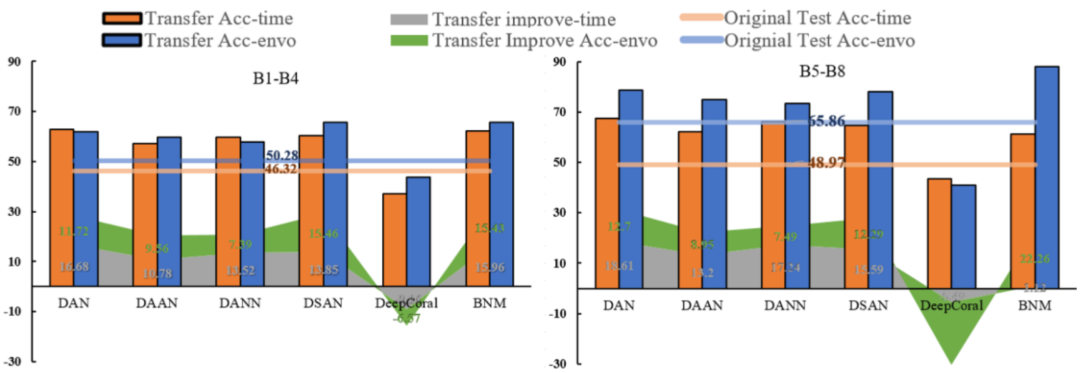
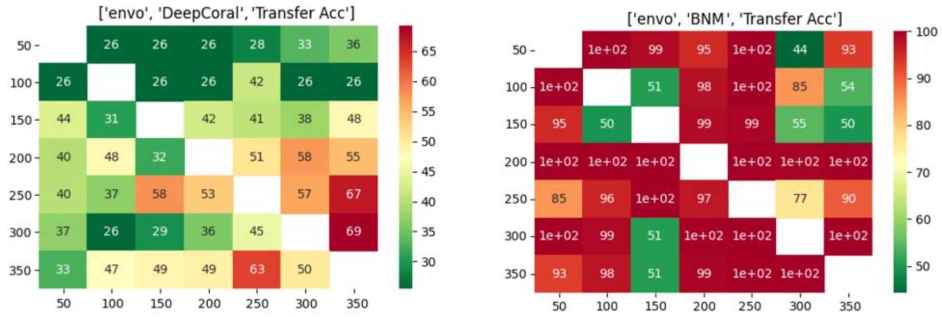


Figure 3 Comparisons of transfer learning accuracy average



(a)DeepCoral model trained by envelope dataset (b)BNM model trained by envelope dataset

Figure 4 Mesh plot of transferred model classification accuracy for various speed conditions

Table 4 Prediction results of BNM network trained with envelope spectrum data under different working conditions

Transfer Method/ Data type/	Channel	Target	Optimum Source Domain	ACC	Predict Source Domian	ACC	ACC (Real)
BNM Envo	B1-B4	50	300	72.8	200	68.8	55.4
BNM Envo	B1-B4	100	150	99.9	150	86.3	99.9
BNM Envo	B1-B4	150	100	100	350	100	100
BNM Envo	B1-B4	200	300	99.9	300	100.5	99.9
BNM Envo	B1-B4	250	200	100	300	100.5	100
BNM Envo	B1-B4	300	250	100	250	100.5	100
BNM Envo	B1-B4	350	300	100	300	95.5	100
BNM Envo	B5-B8	50	200	100	300	99.8	99.69
BNM Envo	B5-B8	100	50	100	200	95.9	100
BNM Envo	B5-B8	150	200	100	200	98.2	100
BNM Envo	B5-B8	200	300	99.6	300	97.2	99.6
BNM Envo	B5-B8	250	50	100	350	98.2	100
BNM Envo	B5-B8	300	350	100	350	97.54	100
BNM Envo	B5-B8	350	200	100	300	97.9	100

From the perspective of different bogies, the pre-training accuracy of bogies of the same structure is not similar, and the transfer accuracy is quite different, but the trend is still similar. From the perspective of different method comparisons, the most significant is the failure of the DeepCoral method, which exhibits severe negative transfer in various data types and bogies, as shown in Figure 6(a). In addition to this, the structure transfer method BNM works best in the diagnosis of envelope spectral data for B5-B8 bogies, as shown in Figure 6(b).

When comparing the transfer learning performance between different cross-speed tasks, several significant pattern can be discovered. For the overall trend, the efficient of transfer learning depends on the ratio of the cross speed (between source and target domain) to the source domain speed. It is same with the trend derived from mechanics analysis. But for specific tasks, classification accuracy shows a mutation among the same source domain, manifested as exceptionally high or low accuracy. So it is infeasible to predict transfer and transfer effects using linear speed condition differences. If new working condition data appears, it is necessary to retrain the model and achieve the optimal transfer effect by traversing the working condition transfer, which will waste a lot of time and computing resources. The data of different monitoring objects may be perturbed by uncertainty, but the laws of transferability between working conditions obtained by using different transfer learning methods are similar.

4 Conclusions and Contributions

this paper proposes a method for evaluating the transfer effect of source domain data by integrating weak learners to find the mapping relationship between data feature distribution differences and transfer learning effects, achieving source domain effect evaluation without the need for model retraining. To verify the proposed method, fault simulation experiments were conducted using a real car rolling test platform, a real high-speed train vehicle, and real bearing fault parts. Cross-validation transfer learning was performed using bearing monitoring data under different speed conditions, and the proposed method was validated. The results show that the proposed method can accurately evaluate the diagnostic effect of transfer learning and provide data processing methods and application recommendations for transfer learning models.

References

[1] H. Wang, J. Wang, Y. Zhao, Q. Liu, M. Liu, and W. Shen, "Few-Shot Learning for Fault Diagnosis With a Dual Graph Neural Network," *IEEE Transactions on Industrial Informatics,* vol. 19, no. 2, pp. 1559-1568, 2022.

[2] B. Yang, C.-G. Lee, Y. Lei, N. Li, and N. Lu, "Deep partial transfer learning network: A method to selectively transfer diagnostic knowledge across related machines," *Mechanical Systems and Signal Processing,* vol. 156, p. 107618, 2021/07/01/ 2021.

[3] H. Hu, B. Tang, X. Gong, W. Wei, and H. Wang, "Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks," *IEEE Transactions on Industrial Informatics,* vol. 13, no. 4, pp. 2106-2116, 2017.

[4] G. Xu, D. Hou, H. Qi, and L. Bo, "High-speed train wheel set bearing fault diagnosis and prognostics: A new prognostic model based on extendable useful life," *Mechanical Systems and Signal Processing,* vol. 146, p. 107050, 2021.

[5] T. Zhang *et al.,* "Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions," *ISA transactions,* vol. 119, pp. 152-171, 2022.

[6] C. Cheng, J. Wang, H. Chen, Z. Chen, H. Luo, and P. Xie, "A review of intelligent fault diagnosis for high-speed trains: Qualitative approaches," *Entropy,* vol. 23, no. 1, p. 1, 2020.
[7] C. Li, S. Zhang, Y. Qin, and E. Estupinan, "A systematic review of deep transfer learning for machinery fault diagnosis," *Neurocomputing,* vol. 407, pp. 121-135, 2020.
[8] A. Farahani, S. Voghoei, K. Rasheed, and H. R. Arabnia, "A brief review of domain adaptation," *Advances in data science and information engineering,* pp. 877-894, 2021.
[9] T. Li, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Domain Adversarial Graph Convolutional Network for Fault Diagnosis Under Variable Working Conditions," *Ieee Transactions on Instrumentation and Measurement,* vol. 70, 2021 2021, Art. no. 3515010.
[10] Y. Feng, J. Chen, J. Xie, T. Zhang, H. Lv, and T. Pan, "Meta-learning as a promising approach for few-shot cross-domain fault diagnosis: Algorithms, applications, and prospects," *Knowledge-Based Systems,* vol. 235, p. 107646, 2022/01/10/ 2022.
[11] T. Pan, J. Chen, J. Xie, Z. Zhou, and S. He, "Deep Feature Generating Network: A New Method for Intelligent Fault Detection of Mechanical Systems Under Class Imbalance," *IEEE Transactions on Industrial Informatics,* vol. 17, no. 9, pp. 6282-6293, 2021.
[12] Z. Chen, J. Wu, C. Deng, X. Wang, and Y. Wang, "Deep Attention Relation Network: A Zero-Shot Learning Method for Bearing Fault Diagnosis Under Unknown Domains," *IEEE Transactions on Reliability,* 2022.
[13] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mechanical Systems and Signal Processing,* vol. 138, p. 106587, 2020.
[14] L. Chen, Q. Li, C. Shen, J. Zhu, D. Wang, and M. Xia, "Adversarial domain-invariant generalization: A generic domain-regressive framework for bearing fault diagnosis under unseen conditions," *IEEE Transactions on Industrial Informatics,* vol. 18, no. 3, pp. 1790-1800, 2021.
[15] X. Liu and S. Zhang, "Who is closer: A computational method for domain gap evaluation," *Pattern Recognition,* vol. 122, p. 108293, 2022/02/01/ 2022.
[16] D. Yang, H. R. Karimi, and K. Sun, "Residual wide-kernel deep convolutional auto-encoder for intelligent rotating machinery fault diagnosis with limited samples," *Neural Networks,* vol. 141, pp. 133-144, 2021/09/01/ 2021.
[17] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," *arXiv preprint arXiv:1412.3474,* 2014.
[18] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *International conference on machine learning*, 2015, pp. 97-105: PMLR.
[19] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *Computer Vision–ECCV 2016 Workshops: Amsterdam, The*

Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14, 2016, pp. 443-450: Springer.

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

[21] Y. Zhu *et al.*, "Deep subdomain adaptation network for image classification," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 4, pp. 1713-1722, 2020.

[22] C. Yu, J. Wang, Y. Chen, and M. Huang, "Transfer learning with dynamic adversarial adaptation network," in *2019 IEEE international conference on data mining (ICDM)*, 2019, pp. 778-786: IEEE.

[23] S. Cui, S. Wang, J. Zhuo, L. Li, Q. Huang, and Q. Tian, "Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 3941-3950.