

Received 29 November 2024, accepted 6 December 2024, date of publication 12 December 2024,
date of current version 23 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3516525

RESEARCH ARTICLE

A Simple but Effective Way to Handle Rotating Machine Fault Diagnosis With Imbalanced-Class Data: Repetitive Learning Using an Advanced Domain Adaptation Model

DONGHWI YOO¹, MINSEOK CHOI¹, HYUNSEOK OH¹,
AND BONGTAE HAN², (Member, IEEE)

¹School of Mechanical and Robotics Engineering, Gwangju Institute of Science and Technology, Gwangju 61005, Republic of Korea

²Department of Mechanical Engineering, University of Maryland, College Park, MD 20742, USA

Corresponding author: Hyunseok Oh (hsoh@gist.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT; MSIT) (No. 2021R1A2C1008143 and No. RS-2022-00144441).

ABSTRACT Fault data from in-service rotating machines are extremely scarce. This is usually true even when healthy data are abundant, leading to the problem of class imbalance. Numerous solutions have been proposed to cope with the problem of class imbalance; each solution has its own advantages and disadvantages in implementation. This paper proposes a much simpler and efficient method for fault diagnosis of rotating machines. By employing pseudo-labeling, weighted random sampling, and time-shifting, the proposed repetitive learning method generates pseudo-augmented source and target fault data. Deep convolutional domain adaptation networks are followed to extract features by minimizing different losses. The evaluation results demonstrate the effectiveness of the proposed method, achieving accuracy rates of 90.79% (CWRU), 76.26% (XJTU), and 86.45% (GIST) under extreme imbalance conditions ($\rho = 0.01$), outperforming existing methods by 10-30% while maintaining computational efficiency. The evaluation results show that repetitive learning produces accurate prediction performance even in situations with extremely imbalanced data, which corroborates the effectiveness offered by the proposed method, despite its simplicity.

INDEX TERMS Class imbalance, data augmentation, domain adaptation, intelligent fault diagnosis, rotating machines.

I. INTRODUCTION

Deep learning-based approaches for fault diagnosis of rotating machines have been studied extensively, often using vibration signals [1], [2], [3], [4], [5]. It has been observed in many fields that healthy data are abundant and fault data are scarce [6]; this is known as the class-imbalance problem. Several data-driven methods have been proposed to handle the class-imbalance problem for fault diagnostics [7], [8], [9].

One way to cope with the scarce fault data is to augment it. Data augmentation techniques, such as rotation,

The associate editor coordinating the review of this manuscript and approving it for publication was Guillermo Valencia-Palomo¹.

reversing, permutation, scaling, magnitude-warping, jittering, etc., are straightforward to implement. For example, bearing fault diagnosis performance was improved when reversing of vibration signals was incorporated with limited labeled samples [10]. Applying data augmentation to one-dimensional (1D) signals, such as vibration waveforms, is not always possible. For “reversing” data augmentation, the maximum amount of augmented data is only twice that of the original 1D signal. Recently, more sophisticated approaches have emerged to address these limitations. For instance, system identification techniques have been proposed for vibration signal augmentation, where models are created using trigger signals as input and vibration

signals as output, showing promising results in generating realistic synthetic data while preserving the physical characteristics of mechanical systems [11]. Additionally, statistical feature-based augmentation methods combined with deep neural networks have demonstrated significant improvements in prediction accuracy, achieving up accuracy through the synthesis of fault-specific patterns in electric machinery vibration signals [12]. Furthermore, recent studies have introduced sophisticated methods such as purposive data augmentation strategies using minority class imbalance rate (MiCIR) to determine optimal augmentation targets and amounts, while preserving the inherent characteristics of defect samples [13]. Other research has explored semantic-discriminative augmentation-driven approaches that carefully control mixed sample sources from identical classes to maintain semantic consistency while addressing class imbalance problems in fault diagnosis [14].

Advanced, but complicated, generative model-based techniques have been presented to relieve the class-imbalance problem, based on normalized variational autoencoder (VAE) [15], generative adversarial network (GAN) [16], [17], [18], and diffusion-based model [19] approaches. For example, Zhao et al. [15] proposed a neural architecture that combined the normalized conditional VAE with the adaptive focus loss. Wang et al. [20] presented a gradient flow-based meta-GAN for fault diagnosis of rotating machines. This method achieved a high diagnostic accuracy with a low standard deviation; however, the computational cost is high at the inference stage due to the pyramid structure of GANs.

The attention-enhanced conditioning-guided diffusion-based approach [19] achieved a remarkably-high accuracy of 98.96% when evaluated with real inverter signals, demonstrating the value of the synthetic samples for bearing fault detection. It should be noted that the diffusion probabilistic model-based approach can be challenging to implement since the inference speed is low and the training cost is extremely high [21]. Most VAE- and GAN-based approaches lack the physical interpretability of the data augmentation methodologies, as they are regarded as black boxes.

The improvements in the previous studies are evident. Yet, there still exist some critical issues to resolve the class-imbalance problem in fault diagnosis of rotating machines in practice, including but not limited to:

- (1) A sufficient number of diverse data should be generated with an affordable computational cost.
- (2) Fault diagnosis results should be accurate and reliable.
- (3) The data generation processes should be physically interpretable.

Given the challenges remaining in the existing methods related to the class-imbalance and cross-domain problems, this paper proposes a simple but effective way to handle the problem, called the Repetitive Learning Method. When the AI model is trained on class-imbalanced datasets, mini-batch learning often repeatedly samples the same fault data, reducing diversity. The repetitive learning method ensures diverse data generation in each iteration of model training even

when having few faults data sample in the original dataset. Time-shifting data augmentation, a key element of repetitive learning, leverages the repeated fault-specific patterns in rotating machines. The method can preserve the physical characteristics of the data, improving the model's ability to handle class imbalance. The proposed method is combined with a deep convolutional domain adaptation model and implemented for fault diagnosis of rotating machines subjected to different operating conditions. A sufficient number of synthetic data can be augmented with low computational cost, while achieving acceptable accuracy, when they are trained with severely imbalanced class data and then forced to diagnose rotating machines under different operating conditions. The effectiveness of the proposed method is verified by applying it to several case studies of extreme situations with only a single fault data sample available for model training.

The remaining sections of this paper are organized as follows: Section II presents the theoretical background; Section III describes the details of the proposed method; Section IV reports the results of three case studies examined to evaluate the effectiveness of the proposed method; Section V summarizes the conclusions of this research and offers suggestions for future work.

II. TECHNICAL BACKGROUND

This section presents a brief overview of unsupervised domain adaptation and class-imbalanced learning. The challenge attributed to multiple operating conditions in rotating machines was often addressed by using unsupervised domain adaptation, while the difficulty due to class-imbalance could be addressed by adopting class-imbalanced learning.

A. UNSUPERVISED DOMAIN ADAPTATION

Conventional deep-learning models have been shown to offer a fault-diagnosis accuracy of almost 100% when particular vibration signals datasets are used [22], [23]. The validity of these existing models was evaluated by applying them to training data and test data drawn from identical probability distributions [24]. However, unlabeled data, which is common in field settings, can lead to dissimilar distributions between the training and test (operating) data [25], [26]. The discrepancy between the training and test data distributions often poses a challenge for fault diagnosis of rotating machines. When test data, which has a unique distribution due to the different operating conditions, are put into a deep-learning model trained initially with data from the training distribution, the predictability of the trained model is decreased due to the discrepancy between the training and test data distributions.

In prior work, domain adaptation has been used to address this issue by minimizing the distribution discrepancy in the feature space. Domain adaptation becomes unsupervised when the following two conditions are met: (1) the domain of the source data is different from that of the target data, and (2) the target data are unlabeled [27]. In this case, labeled source data from source domain information to unlabeled

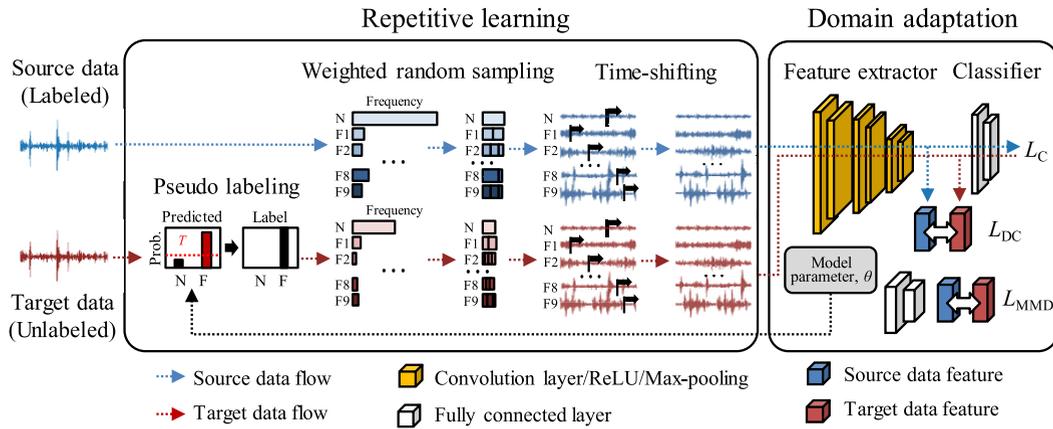


FIGURE 1. Proposed repetitive learning framework.

target data from target domain. Different domains, such as mechanical loads and rotational speeds, can be considered different operating conditions.

Unsupervised domain adaptation was employed to address the distribution discrepancy for enhanced fault diagnostics [28], [29], [30]. As an example, Guo et al. proposed a new model, called a deep convolutional transfer learning network (DCTLN), for fault diagnosis of machines with unlabeled data [31]. Several other transfer-learning-based fault-diagnosis methods, including Wasserstein distance [32], maximum mean discrepancy (MMD) [33], correlation alignment (CORAL) [34], maximum mean square discrepancy [35], variance discrepancy representation [36], and heterogeneous federated domain generalization network [37] have also been proposed.

B. CLASS-IMBALANCED LEARNING

Deep-learning models are generally trained under the assumption that the training samples in the classes are balanced. When the number of training samples is imbalanced, the decision boundary is prone to shift towards the majority class. As a result, the trained model will suffer from “label bias.” In this situation, the model will have poor predictive capability for the minority classes. Thus, a model trained in this circumstance usually underperforms during testing due to its low generalization capability. According to a survey in prior work [38], possible solutions to address this situation can be categorized into class rebalancing, module improvement, and information augmentation.

First, class rebalancing, through re-sampling and re-weighting, addresses the class imbalance effectively. However, it can reduce performance on the majority classes. Module improvement enhances network components such as representation learning and classifier design. Although effective, it can be complex and resource-intensive. Finally, information augmentation, particularly data augmentation, produces synthetic data to enhance model generalization and is practical because it can improve performance without sacrificing majority-class accuracy. Among these options, the

information augmentation approach provides the model with additional information, which makes it an ideal choice for this research.

Data augmentation is an example of information augmentation, which increases the size and quality of the data used for decision making. Data augmentation is simple and effective compared to other methods. In prior work, this approach was demonstrated for computer vision, showing that data augmentation techniques helped to significantly improve a model’s performance [39]. Thus, the class-imbalanced data problem can be mitigated by augmenting data through the affine transformation of available images. Nonetheless, the validity of the existing data augmentation methods was not fully investigated for rotating machine fault diagnostics using 1D signals such as vibration signals. To this end, we proposed a fault diagnosis method using a repetitive learning.

III. PROPOSED REPETITIVE LEARNING METHOD FOR FAULT DIAGNOSIS

This section presents a novel rotating machine fault-diagnosis framework with repetitive learning and domain adaptation. Three conditions are taken into consideration for implementation of the proposed method. The conditions are determined from careful observations of the operating conditions and vibration data found in in-use field conditions. First, it is assumed that there is a sufficient amount of normal data available to train the deep-learning model, while fault data are insufficient. Second, source data (e.g., testbed or simulation data) are labeled whereas, target data (e.g., data from machines in the field) are unlabeled. Finally, it was hypothesized that the predictive performance of artificial intelligence models could be improved through iterative training using synthetic data generated from the same source data distribution.

A. OVERVIEW

The proposed method in this study consists of a repetitive learning module and a domain adaptation module, as depicted in Fig. 1. The input to the AI model includes source and target

data collected from different domains. It is generally known that batch learning, which trains the model using all the data at once, is inefficient. To overcome this, previous studies [40], [41] improved learning efficiency by using the mini-batch learning method, which randomly samples small amounts of data for AI model training. However, when there is a class imbalance between normal and fault data in the source dataset, a small portion of the data is sampled repeatedly. For example, if there are 304 normal data and 16 fault data, and the mini-batch size is set to 32, approximately 30 normal data and 2 fault data will be sampled each time. To balance the number of normal and fault data, if 16 samples are uniformly selected from each class, the fault data will always consist of the same 16 samples. As a result, the fault data in the mini-batch will have much less diversity than the normal data, potentially causing problems during AI model training. Therefore, when there is an imbalance between data classes, the traditional mini-batch learning approach poses potential risks, and the new repetitive learning method is proposed to overcome this.

The repetitive learning module constructs a refined dataset to enable the AI model to repeatedly learn data with an identical probability distribution and then feeds this dataset into the subsequent domain adaptation module. To achieve this, the repetitive learning module consists of three components. First, pseudo-labeling addresses the issue of missing labels by assigning appropriate labels to the target data. Second, weighted random sampling mitigates the quantitative imbalance between normal and fault data in both the source and target datasets. Finally, time-shifting generates a sufficient amount of synthetic data and constructs the dataset for AI model training. This is a key element for repetitive learning, as it generates data that reflects the diversity encountered during the operation of rotating machinery while preserving the inherent physical characteristics of the source vibration data.

In the domain adaptation module, the AI model is trained to minimize the probability distribution gap between the source and target data, allowing the extraction of common features necessary for fault diagnosis. As domain adaptation is iteratively trained and the weights and biases are updated, the accuracy of pseudo-labeling improves. Through this iterative process between the repetitive learning and domain adaptation modules, the proposed method is expected to enhance the fault diagnosis performance. The details are described below.

B. REPETITIVE LEARNING

Normal data are abundant for most rotating machines in the field, while fault data are typically scarce. The imbalance ratio, ρ , is defined as:

$$\rho = \frac{N_f}{N_n} \quad (1)$$

where N_f and N_n are the number of fault and normal data samples in each class (fault and normal), respectively. For

example, ρ is 0.01 when the number of fault and normal data samples in each class are seven and 700, respectively.

The challenges associated with low imbalance ratios can be alleviated if fault data can be augmented. In the source domain (e.g., testbeds or simulations), fault data can be augmented, as the true label of the fault data is already known. However, fault data cannot be augmented for target data since the true label is unknown. Thus, it is impossible to conduct data augmentation unless the true labels of the target data are known. To cope with this challenge, the pseudo-label method is incorporated in this study. Pseudo-labeling artificially assigns labels predicted from the trained models to the unlabeled data [42]. Among various pseudo-label methods, the hard pseudo-label method is adopted in this study [43]. This method was chosen because it is simple and easy to implement, because it directly uses the prediction results from the network. Let \tilde{y}_i and p_i be the pseudo labels and the probability output of a trained model on a sample x_i , respectively. Then, the pseudo-label can be generated as:

$$\tilde{y}_i = \begin{cases} 1 & \text{if } p_i \geq T \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where $T \in (0, 1)$ is the threshold that filters out the pseudo-labels. The pseudo-label allows determination of which data are faulty among the unlabeled data in the target domain. In this way, the faulty labeled data are augmented in the source domain, while the pseudo-labeled fault data are in the target domain.

A weighted random sampler technique can also be incorporated to mitigate the adverse effects of class imbalances during training. The weighted random sampler (WRS) assigns weights to each sample in the dataset that correspond to the inverse of the class frequency. This approach ensures that samples from underrepresented classes are assigned higher weights, increasing their likelihood of being selected during training. Consequently, WRS counteracts the challenges of class imbalance, granting equal representation to all classes and preventing the model from being overly biased towards the majority class. By fostering balanced training data, the model is expected to become more adaptive, while generalizing patterns across all classes, thus ultimately enhancing the classification performance in both the source and target domains.

Using the time-shifting data augmentation method, a sufficient amount of data required for AI model training is generated. The time-shifting method creates numerous synthetic data samples from a single source, producing the necessary amount of data for repetitive learning. Vibration data generated by rotating components such as rotors and bearings exhibit a characteristic repetitive pattern in specific cycles (e.g., one revolution and harmonics, $1 / \text{bearing defect frequency}$). Therefore, if the length of a single vibration waveform is long enough to include this repetitive vibration pattern, splitting the waveform at a specific point will still preserve the repeating pattern. Even if the order of the split

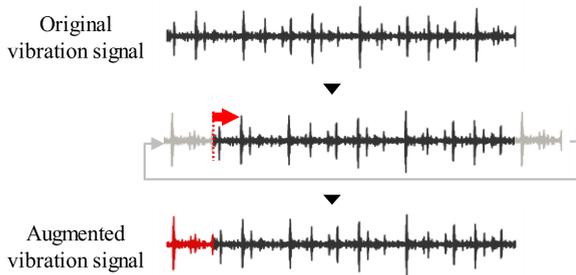


FIGURE 2. Illustration of the concept of time-shifting.

vibration waveforms is rearranged, the repetitive vibration pattern remains intact, except at the connection point. Fig. 2 illustrates an example of time-shifting based on the concept above. First, the whole vibration waveform is shifted to the right. Then, the vibration waveform that exceeds the original time domain is added as the beginning part of the shifted vibration waveform. The small amount of fault data can be augmented by randomly changing the size of the time shifts. It needs to be confirmed that the vibration waveform should contain two transient impulses at least. For example, the ball pass frequency at the inner race (BPFI) is 254 Hz when the rotational speed is 3,000 rpm. Then, time duration of a single transient impulse is seconds 3.937 msec ($= 1/254$ Hz). Therefore, the minimum length of the vibration waveform measured should be larger than 3.937 msec at least. To account for the safety margin, it is recommended that the length of the vibration waveform should be ten times larger than the minimum time interval value of the transient impulses. The adverse effect of improper selection of time shifts is shown in Fig. 3, which can cause noticeable distortion and information loss.

The CNN architecture includes locally connected and max pooling layers. Unlike fully connected layers, the locally connected layer extracts bearing defect characteristic factors through convolution operations between the kernel and the receptive field. Given a vibration waveform, convolution operations are iteratively performed across multiple receptive fields in the vibration waveform. In this context, the particular receptive field that is compromised by time-shifting is likely to exhibit a significantly low convolution value due to the disruption of time-dependent information, and the value is subsequently disregarded in the max pooling layer. Consequently, it is anticipated that time-shifting can extract bearing defect characteristics effectively when integrated with CNNs.

C. DOMAIN ADAPTATION

As stated earlier, several transfer-learning models, including DCTLN [31] and CMD [44], were developed to diagnose REB faults. Any CNN-based model listed above can be combined with the proposed repetitive learning method. This study employs one of the state-of-the-art domain adaptation models (i.e., DCTLN) as a representative and effective example.

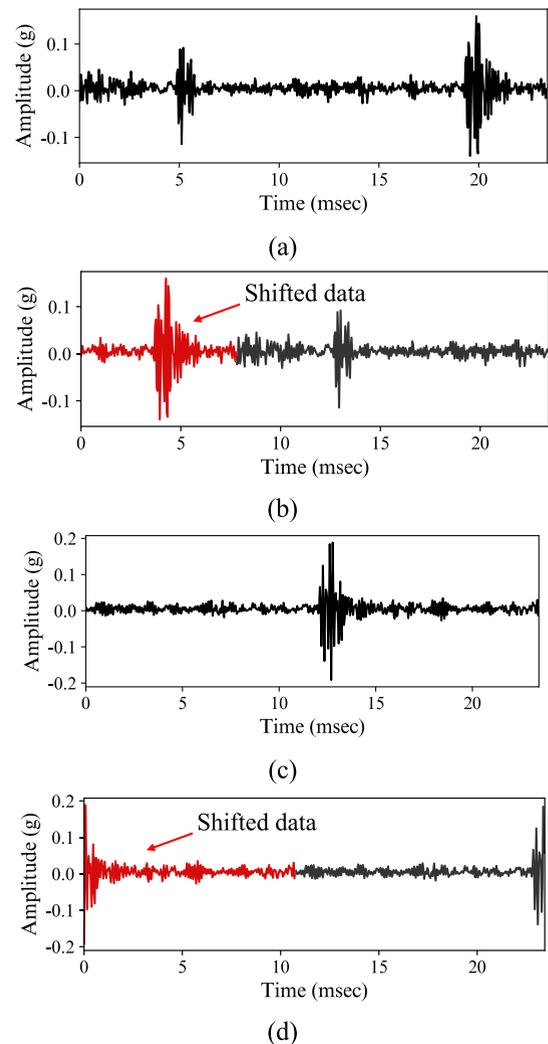


FIGURE 3. Representative vibration signals containing inner race faults: (a) vibration signal with two or more pulses; (b) time-shifted signal of (a); (c) vibration signal with a single pulse; and (d) the time-shifted signal of (c), where the pulse is fragmented resulting in the degradation of time-dependent information.

DCTLN aims to minimize the distribution discrepancy between the source and target domains in transfer-learning scenarios. To achieve this goal, the DCTLN architecture includes two key components. The first component is a deep convolutional neural network (CNN) model that extracts patterns from 1D signals, such as vibration signals, and assigns a corresponding probability to the target. The features extracted from the source and target data are fed into the following layers that quantify the distribution discrepancy between the source and target domains. The feature extractor captures high-level features from vibration signals, which subsequently serve as the foundation for the classifier's training process. The features extracted from the source data are fed into the classifier to determine the health conditions of the bearings. The loss function for the health classifier

is defined as:

$$L_C = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k I[y_i = k] \log \frac{e^{x_j}}{\sum_{l=1}^k e^{x_l}} \right] \quad (3)$$

where m is the batch size of the training samples; k represents the fault category; y is the target label; x is the output probability; and $I[\cdot]$ is an indicator function.

Domain adaptation extracts the common features shared between the two domains, while preserving domain-specific information. The domain classifier is the binary classifier that maximizes domain recognition errors (i.e., domain adversarial adaptation). If the domain classifier fails to distinguish between features originating from the source and target domains, the features are considered to be domain-invariant. The loss function for the domain classifier is defined as:

$$L_{DC} = \frac{1}{m} \sum_{i=1}^m (g_i \log d(x_i) + (1 - g_i) \log(1 - d(x_i))) \quad (4)$$

where m is the batch size of the training samples; g_i is the ground-truth domain label; and $d(x_i)$ denotes the output domain for the i th sample, which indicates whether x_i comes from the source domain or the target domain.

The domain distribution discrepancy is expressed by the maximum mean discrepancy (MMD) loss. To alleviate the distribution discrepancy between features acquired from distinct domains, the MMD distribution distance between the source and target data high-level features is calculated. The loss function for the domain distribution discrepancy is expressed as:

$$L_{MMD} = \frac{1}{n_s^2} \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} k(x_i^{(S)}, x_j^{(S)}) + \frac{1}{n_t^2} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} k(x_i^{(T)}, x_j^{(T)}) - \frac{1}{n_s n_t} \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} k(x_i^{(S)}, x_j^{(T)}) \quad (5)$$

where n is the number of domain samples; x is the output vector for domain discrimination; and $k(\cdot, \cdot)$ is the Gaussian radial basis function (RBF).

DCTLN has three optimization objectives: 1) minimize the health condition classification error for the source-domain dataset, 2) maximize the domain classification error for the source and target domain datasets, and 3) minimize the MMD distance between the source and target domain datasets. In summary, DCTLN has three losses: the cross-entropy loss (L_C), domain classifier loss (L_{DC}), and domain discrepancy loss (L_{MMD}). Specifically,

$$L = L_C - \lambda L_{DC} + \mu L_{MMD} \quad (6)$$

where λ and μ are the weight constants that control the strength of the losses. The domain-adaptation model is trained to minimize the total loss.

The overall procedure of the proposed method, including the data augmentation and the domain adaptation steps,

is presented in Algorithm 1. After initializing weights and biases, mini batches are sampled from the source and target data. Pseudo-labels are generated for the mini-batch of the unlabeled target data with the threshold T . For fault data, time-shifting is conducted with WRS. Using the augmented source and target data, a forward propagation is undertaken. The classification loss is calculated using the augmented source data, while the domain discrepancy loss is calculated using the augmented source and target data. This process is repeated until the convergence criteria are met.

Algorithm 1 Class Imbalanced Domain Adaptation

Input: the labeled source dataset $D_S = \{(x_i, y_i)\}_{i=1}^{N_S}$, the unlabeled target dataset $D_T = \{(x_j)\}_{j=1}^{N_T}$, number of batches $num_batches$, augment number N_{aug} , threshold T , Domain classifier weight λ , MMD weight μ , and learning rate l .

Output: the feature extractor f_{FE} and the label classifier f_{LC}

1. Initialize weights and biases for f_{FE} and f_{LC}
2. For epoch = 1 to num_epochs :
3. For batch = 1 to $num_batches$:
4. Sample the mini-batch from D_S and D_T with WRS
5. Predict the pseudo-label \tilde{y}_j of unlabeled target data x_j with the threshold T
6. Augment fault data with time-shifting for N_{aug} times
7. Forward propagation of D_S and D_T
8. Calculate the classification loss L_C
9. Calculate the domain discrepancy loss L_{DC}
10. Calculate the MMD loss L_{MMD}
11. Calculate the total loss $L = L_C - \lambda L_{DC} + \mu L_{MMD}$
12. Backpropagation
13. Update the weights and biases of f_{FE} and f_{LC} with the learning rate l

End

End

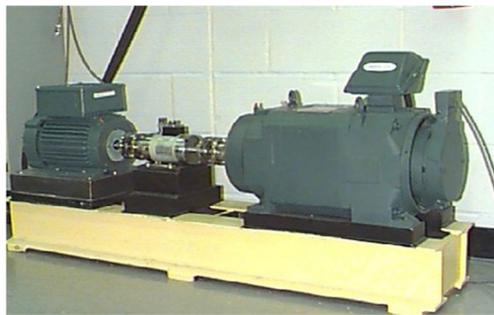
IV. CASE STUDIES

A. DATASETS

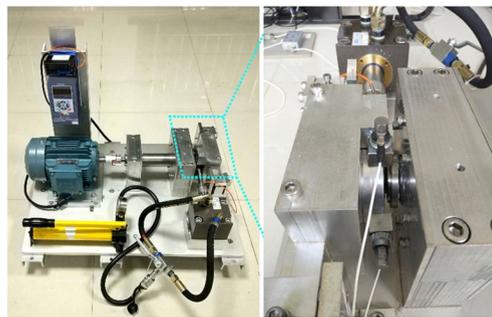
The performance of the proposed method was evaluated by applying it to three datasets: the first from Case Western Reserve University (CWRU), the second from Xi'an Jiaotong University (XJTU), and the third from Gwangju Institute of Science and Technology (GIST). The testbeds employed are depicted in Fig. 4. Vibration signals were measured from bearings for all datasets. The sampling rate, rotational speed, loading conditions, and health conditions are summarized in Table 1. Further details about the datasets are available in [45], [46], and [47].

The CWRU dataset contains four types of health conditions, including normal (N), ball fault (B), inner raceway fault (IR), and outer raceway fault (OR). Defects with 0.007, 0.014, and 0.021-inch diameters were artificially seeded. The XJTU dataset also contains four types of health conditions. However, only N, IR, and OR were used in this study because these health conditions are relevant to domain-adaptation tasks. The defects in the XJTU dataset were generated naturally through run-to-failure tests. Therefore, the fault-diagnostic task using the XJTU dataset is more challenging. In addition to the CWRU and XJTU datasets, which have been

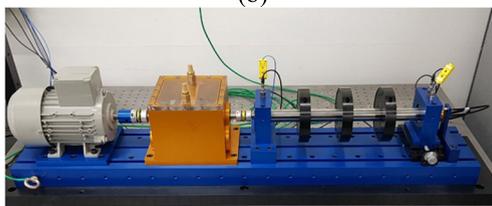
extensively analyzed, the new GIST dataset was employed to evaluate additional fault types in rotating machines, specifically misalignment and unbalance. The GIST dataset features four conditions: normal (N), outer raceway fault (OR), misalignment (M), and unbalance (U). Artificial defects of 0.6, 1.0, and 1.6 mm were artificially seeded for outer raceway faults. In addition, defects with severities of 0.4 and 0.8 mm for misalignment and 1.5 and 3.0 grams for unbalance were artificially seeded.



(a)



(b)



(c)

FIGURE 4. Testbeds from (a) CWRU, (b) XJTU, and (c) GIST.

B. EXPERIMENTAL SETUP

The original CWRU dataset consists of multiple sets of vibration signals with different labels. In this study, a single datum was acquired by imposing a sliding window to the vibration signals of the individual sets in the original CWRU dataset. A single datum was defined as 1,200 sampling points. The ten classes each have 1,000 individual data from the vibration signals. Table 2 summarizes the experimental setup for the CWRU dataset. The quantities of samples in the normal training datasets for both the source and target domains were 700 and 700, respectively. The amounts of faulty samples in the training datasets (B, IR, and OR with three defect lengths) in the source and target domains were determined by the

TABLE 1. Datasets Used in the Case Studies.

Dataset	Type	Sampling rate	Health conditions	Operating condition	
				Rotational speed	Loading condition
CWRU A	Artificially seeded defect	12 kHz	N/B/IR/OR	1797 rpm	0 hp
CWRU B		12 kHz	N/B/IR/OR	1772 rpm	1 hp
CWRU C		12 kHz	N/B/IR/OR	1750 rpm	2 hp
CWRU D		12 kHz	N/B/IR/OR	1730 rpm	3 hp
XJTU 1	Run-to-failure	25.6 kHz	N/IR/OR	2250 rpm	11 kN
XJTU 2		25.6 kHz	N/IR/OR	2400 rpm	10 kN
GIST 1	Artificially seeded defect	25.6 kHz	N/OR/M/U	450 rpm	0 kN
GIST 2		25.6 kHz	N/OR/M/U	600 rpm	0 kN
GIST 3		25.6 kHz	N/OR/M/U	750 rpm	0 kN

imbalance ratio of ρ . For example, when the imbalance ratio is 0.01, the inner race fault training samples in the source and target domains are seven and seven, respectively. These seven fault data were randomly sampled from the original 1,000 fault data. The quantity of test data samples in the target domain was 300 for data with any health condition.

TABLE 2. Experimental Setup for the CWRU Dataset.

Health condition	Training data		Test data
	Source domain	Target domain	Target domain
Normal	700	700	300
Ball fault (0.007 inch)	$700 \times \rho$	$700 \times \rho$	300
Ball fault (0.014 inch)	$700 \times \rho$	$700 \times \rho$	300
Ball fault (0.021 inch)	$700 \times \rho$	$700 \times \rho$	300
Inner race fault (0.007 inch)	$700 \times \rho$	$700 \times \rho$	300
Inner race fault (0.014 inch)	$700 \times \rho$	$700 \times \rho$	300
Inner race fault (0.021 inch)	$700 \times \rho$	$700 \times \rho$	300
Outer race fault (0.007 inch)	$700 \times \rho$	$700 \times \rho$	300
Outer race fault (0.014 inch)	$700 \times \rho$	$700 \times \rho$	300
Outer race fault (0.021 inch)	$700 \times \rho$	$700 \times \rho$	300

* ρ : imbalance ratio

Domain adaptation tasks were conducted using data from two different operating conditions. For example, the source domain corresponded to the operating condition of 1,797 rpm and zero hp, while the target domain was 1,772 rpm and one hp. Twelve domain-adaptation tasks were established with different source and target domains, i.e., CWRU A \rightarrow B/C/D, B \rightarrow A/C/D, C \rightarrow A/B/D, D \rightarrow A/B/C. The proposed domain-adaptation model was trained to classify ten classes with different types and severities of defects. The details of the architecture of the proposed method are shown in Table 3. In our study, the hyperparameters of the proposed method include 200 training epochs, a batch size of 256, a learning rate of 10^{-3} , an MMD weight (μ) of one, and a domain classifier weight (λ) of one. The pseudo-label threshold (T) was established at 0.95, guided by referencing the work from prior case studies [48], [49], [50].

The XJTU data are labeled, as the vibration data were collected from accelerated life testing of the bearings. In this study, the initial one-second data's maximum amplitude (A_n)

TABLE 3. Model Architecture.

Layer	Output size	Parameter	Number of trainable parameters
Input	1-[1200×1]	-	-
Conv1	8-[593×1]	Kernel 16×1, Stride 2, ReLU	136
Pool1	8-[295×1]	Kernel 4×1, Stride 2	-
Conv2	16-[146×1]	Kernel 4×1, Stride 2, ReLU	828
Pool2	16-[72×1]	Kernel 4×1, Stride 2	-
Conv3	32-[35×1]	Kernel 4×1, Stride 2, ReLU	2,080
Pool3	32-[16×1]	Kernel 4×1, Stride 2	-
Conv4	64-[7×1]	Kernel 4×1, Stride 2, ReLU	8,256
Pool4	64-[2×1]	Kernel 4×1, Stride 2	-
Flatten	128×1	-	-
FC1	128×1	-	16,512
FC2	C×1	-	1,290
Softmax	C	-	-
FC_DC1	128×1	-	16,512
FC_DC2	2×1	-	258

had to be calculated to determine a baseline whose vibration level remained low and stable. From the literature, it was found that there was no agreement for setting the threshold level. Therefore, in this study, vibration signals with an amplitude below $3 \times A_n$ were labeled as healthy data according to the Nelson rules in statistical process control. Vibration signals with an amplitude over $3 \times A_n$ were labeled as fault data.

The quantities of training and target datasets in the source and target domains are shown in Table 4. Similar to the CWRU dataset, the source and target domain normal data were 700 and 700, respectively. The imbalance ratio also determines the training source and target data for fault data IR and OR ρ . Two domain-adaptation tasks were established with different source and target domains, i.e., XJTU 1→2, 2→1. The proposed domain-adaptation model was identical to that described for the case study of the CWRU dataset. However, the proposed method was trained to classify three classes with different types and severities of defects.

The quantities of the training and target GIST datasets in the source and target domains are shown in Table 5. The quantities of the source and target domain's normal data were 700 and 700, respectively. The training source and target data for fault data OR, U, and M were also determined by the imbalance ratio ρ . Three domain-adaptation tasks were established with different source and target domains, i.e., GIST 1→2/3, 2→1/3, 3→1/2.

TABLE 4. Experimental Setup for the XJTU Dataset.

Health condition	Training data		Test data
	Source domain	Target domain	Target domain
Normal	700	700	300
Inner race fault	$700 \times \rho$	$700 \times \rho$	300
Outer race fault	$700 \times \rho$	$700 \times \rho$	300

* ρ : imbalance ratio

The deep-learning models were coded with Python 3.8.8 and Pytorch 1.14.0. Calculations were performed on a

TABLE 5. Experimental Setup for the GIST Dataset.

Health condition	Training data		Test data
	Source domain	Target domain	Target domain
Normal	700	700	300
Outer race fault (0.6 mm)	$700 \times \rho$	$700 \times \rho$	300
Outer race fault (1.0 mm)	$700 \times \rho$	$700 \times \rho$	300
Outer race fault (1.6 mm)	$700 \times \rho$	$700 \times \rho$	300
Misalignment (0.4 mm)	$700 \times \rho$	$700 \times \rho$	300
Misalignment (0.8 mm)	$700 \times \rho$	$700 \times \rho$	300
Unbalance (1.5 g)	$700 \times \rho$	$700 \times \rho$	300
Unbalance (3.0 g)	$700 \times \rho$	$700 \times \rho$	300

* ρ : imbalance ratio

desktop computer with an Intel Core i7-9800X (3.80 GHz) processor, 128 gigabytes of DDR4 RAM, and an NVIDIA GeForce RTX 2080 Ti graphics card.

C. RESULTS AND DISCUSSION

1) CWRU DATASET

The performance of the proposed method was evaluated as a function of the imbalance ratio. The results are shown in Fig. 5, where the twelve domain-adaptation tasks were repeated ten times. The performance of the proposed method was compared with those of the existing models previously proposed for fault diagnostics. Domain Adversarial Neural Network (DANN) [51] and Maximum Mean Discrepancy (MMD) [52] are fundamental domain adaptation approaches that have been successfully applied to rotating machinery fault diagnosis. DCTLN is an adversarial domain classifier with MMD loss that is used to address the challenge induced by variable operating conditions. Class-imbalance adversarial transfer learning (CIATL) [49] has global and local adversarial domain discriminators with class-balanced focal loss. The latter was proposed to address the challenges associated with both the class-imbalance problem and those from variable operating conditions.

As shown in Fig. 5(a), the proposed method and the existing models show similar performance if the imbalance ratio remains larger than 0.5. As the imbalance ratio decreases, the proposed method begins to outperform the existing models. As seen in Fig. 5(b), the accuracy of existing models decreases much more rapidly than that of the proposed method with low imbalance ratios. In the most extremely imbalanced case ($\rho = 0.0014$), the proposed method still shows an accuracy of over 80%, while the existing models offer only around 30% accuracy in this condition.

The accuracy of the twelve domain-adaptation tasks was examined further for the case of severely imbalanced training data (i.e., $\rho \leq 0.1$). The results are summarized in Table 6. The means of the accuracies for the imbalance ratios of 0.1, 0.05, and 0.01 are 94.27%, 93.62%, and 90.79%, respectively. The accuracy details of different models for the severely imbalanced dataset ($\rho = 0.01$) are summarized in Table 7.

These results were used to analyze the effect of domain adaptation on the model performance. The bold numbers in the table indicate the domain-adaptation task with the highest

TABLE 6. Accuracy Results of the Proposed Method for the CWRU Dataset.

Imbalance ratio (ρ)	Domain adaptation task using the CWRU dataset												Mean
	A→B	A→C	A→D	B→A	B→C	B→D	C→A	C→B	C→D	D→A	D→B	D→C	
0.1	97.75	92.93	87.89	96.29	99.26	92.11	92.05	97.29	96.39	90.55	92.74	95.96	94.27
	±1.48	±6.78	±8.20	±0.97	±0.76	±6.79	±2.10	±0.78	±2.68	±4.22	±6.92	±2.77	
0.05	96.23	92.22	90.05	93.91	98.21	94.94	89.19	96.73	97.70	90.68	92.05	91.53	93.62
	±2.08	±6.07	±6.00	±3.38	±1.79	±3.23	±2.96	±0.62	±1.44	±4.60	±5.26	±7.87	
0.01	88.45	93.31	90.49	90.09	97.86	90.19	86.39	94.57	94.65	82.82	90.64	90.02	90.79
	±13.0	±4.04	±3.45	±5.30	±2.17	±4.49	±4.81	±1.24	±3.46	±4.59	±7.44	±4.54	

TABLE 7. Comparison of Accuracy Results for Different Models Using the CWRU Dataset ($\rho = 0.01$).

Imbalance ratio (ρ)	Domain adaptation task using the CWRU dataset												Mean
	A→B	A→C	A→D	B→A	B→C	B→D	C→A	C→B	C→D	D→A	D→B	D→C	
DCTLN	59.78	60.46	57.27	58.44	58.87	60.43	56.51	61.01	62.18	54.74	56.80	59.60	58.84
	±3.27	±3.70	±2.24	±3.69	±4.63	±3.46	±4.27	±4.84	±2.87	±4.70	±4.92	±5.31	
DANN	48.33	55.71	61.06	67.29	68.51	62.47	73.16	71.57	56.46	57.66	73.06	69.76	63.75
	±16.09	±13.82	±13.13	±8.98	±11.85	±9.42	±5.00	±3.41	±14.46	±14.43	±5.97	±4.27	
DCTLN	59.78	60.46	57.27	58.44	58.87	60.43	56.51	61.01	62.18	54.74	56.80	59.60	58.84
	±3.27	±3.70	±2.24	±3.69	±4.63	±3.46	±4.27	±4.84	±2.87	±4.70	±4.92	±5.31	
CIATL	59.22	58.17	57.16	57.97	61.71	61.75	58.82	63.49	61.31	56.38	53.36	59.33	59.06
	±3.30	±3.57	±3.52	±5.35	±4.53	±5.29	±5.97	±5.74	±6.54	±5.05	±15.4	±5.11	
Proposed	88.45	93.31	90.49	90.09	97.86	90.19	86.39	94.57	94.65	82.82	90.64	90.02	90.79
	±13.0	±4.04	±3.45	±5.30	±2.17	±4.49	±4.81	±1.24	±3.46	±4.59	±7.44	±4.54	

TABLE 8. Comparison of Computation Time for Different Models Using the CWRU Dataset ($\rho = 0.01$).

Model	Training time per epoch (s)	Inference time per epoch (s)
MMD	0.4111	0.1156
DANN	0.3492	0.0974
DCTLN	0.4211	0.1233
CIATL	0.5869	0.1555
Proposed	0.6081	0.1600

performance. The proposed method offers the highest accuracy of 90.79%, on average.

Table 8 compares the computational performance of different methods on the CWRU dataset with $\rho = 0.01$. The training time per epoch and inference time per epoch are measured in seconds. As shown in the table, the MMD method has the shortest training time of 0.4111s per epoch, followed by DANN (0.3492s), DCTLN (0.4211s), CIATL (0.5869s), and our proposed method (0.6081s). For inference time, all methods show comparable performance ranging from 0.09 to 0.1600s per epoch. Although our method requires slightly more computational time, the difference is negligible considering the significant improvement in classification accuracy.

Confusion matrices of all the models are plotted in Fig. 6 for the domain-adaptation task of the CWRU dataset A→B, where the x- and y- axes are the label predicted by the model and the true label, respectively. All of the testbed models classify the normal health condition correctly. It should be noted that the number of healthy data was sufficient. All models without repetitive learning show low accuracy for the fault data. This can be attributed to the fact that the

decision boundary of the model was improperly generated due to label bias. The healthy data of 700 samples and the fault data of only seven in individual classes were used to train the models without repetitive learning, which forced the decision boundary to be generated towards the healthy samples. This analysis offers a possible explanation for why the performance of other models was poor in the cases of an extremely severely imbalanced dataset (i.e., $\rho = 0.01$).

As discussed earlier, CNN with locality and pooling can help a model to ignore the vibration signal segments impaired by time shifting. The feasibility of integrating the repetitive learning method into convolution blocks is evidenced by examining the feature map. The findings are illustrated in Fig. 7, which depicts the CWRU dataset’s A→B domain-adaptation task outcomes. Convolution blocks, trained with both normal and defective data, received input signals subjected to time shifts. Fig. 7(a) demonstrates the time-shifted vibration signal and the modified segments. Upon introducing the time-shifted signal depicted in Fig. 7(a), the CNN’s first and second feature maps are presented in Fig. 7(b) and (c), respectively. Thus, in the CNN context, it was shown that the disrupted segment of the time-shifted vibration signal is minimally impacted by max pooling.

Fig. 8 visualizes the healthy and faulty samples through t-distributed stochastic neighbor embedding (t-SNE) [53]. The points with a black-circular border line are the source data used for training, while the points without the border line are the target test data. For the proposed method’s t-SNE, the two domain data for individual classes are aligned consistently, as compared to the other models. The feature space of the proposed method for individual classes is more discriminative than that observed for the other models.

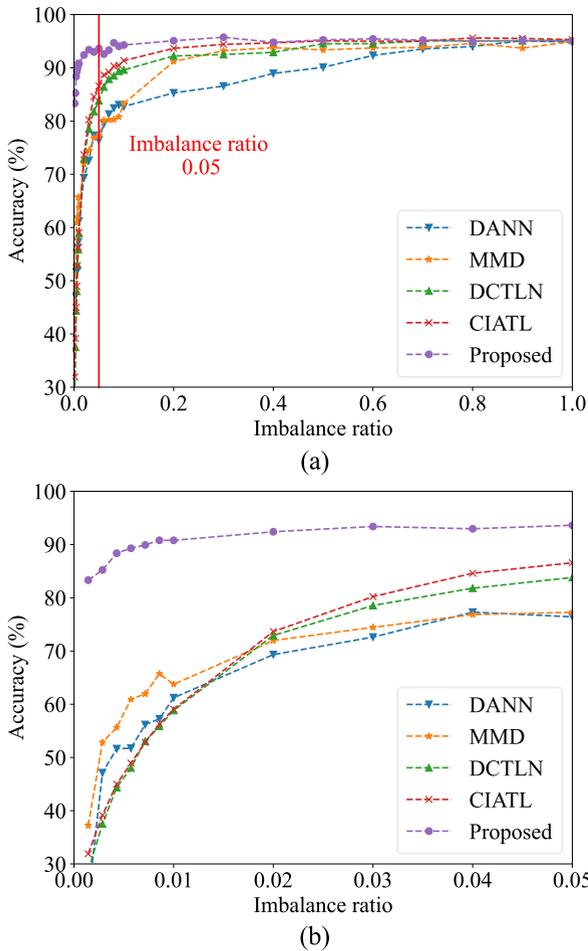


FIGURE 5. Comparison of the accuracy results of the proposed method with four existing methods for the CWRU dataset: (a) zero to one imbalance ratio and (b) zero to 0.05 imbalance ratio.

To evaluate the effectiveness of the proposed repetitive learning method, comprehensive ablation experiments were conducted comparing different augmentation techniques. We specifically examined conventional 1D signal augmentation methods including flipping, jittering, time reversal, and scaling [54], along with our time-shift domain augmentation (TSDA) approach. Fig. 9 presents the comparative results across different imbalance ratios ($\rho = 0.00$ to 0.05), demonstrating that TSDA consistently achieves superior performance. At $\rho = 0.01$, TSDA achieved 92.5% accuracy, significantly outperforming other methods such as flipping (78.3%), jittering (76.8%), time reversal (75.4%), and scaling (74.2%). The baseline case without augmentation achieved only 71.5% accuracy. This superior performance can be attributed to TSDA’s ability to preserve the physical characteristics of the fault signals while generating diverse training samples. Unlike conventional methods that may distort the fundamental frequency components, TSDA maintains the essential fault-specific patterns necessary for accurate diagnosis.

To investigate the effect of the pseudo-label threshold T on the performance of the proposed method, this paper was

analyzed the expected accuracy for different pseudo-label threshold values. Table 9 shows the predicted accuracy results with various pseudo-label threshold values ranging from 0.80 to 0.97 for the CWRU dataset with $\rho = 0.01$. The pseudo-label threshold value of 0.95 demonstrated the highest accuracy of $90.79\% \pm 5.30\%$ among all settings, validating our choice based on the prior studies [48]. Lower pseudo-label threshold values ($T \leq 0.90$) were predicted to yield reduced accuracy due to increased false pseudo-labels, despite enabling the utilization of more target domain data. For instance, at $T = 0.80$, the accuracy was expected to decrease to $85.32\% \pm 7.23\%$ due to the incorporation of less reliable pseudo-labels. Conversely, a higher pseudo-label threshold value ($T = 0.97$) was also predicted to result in performance degradation, likely due to the limited availability of pseudo-labeled data for training, with an expected accuracy of $88.54\% \pm 5.73\%$. These predictions indicate that $T = 0.95$ provides an optimal balance between pseudo-label reliability and data utilization for the proposed method.

TABLE 9. Results with Different Pseudo-Label Threshold Values for the CWRU Dataset ($\rho = 0.01$).

Pseudo-label threshold	Accuracy (%)
0.80	85.32 ± 7.23
0.85	87.41 ± 6.48
0.90	89.23 ± 5.82
0.95	90.79 ± 5.30
0.97	88.54 ± 5.73

2) XJTU DATASET

The accuracy of two domain-adaptation tasks was re-evaluated for the three severely imbalanced datasets. The results are summarized in Table 10. The means of the accuracy results for the imbalance ratios of 0.1, 0.05, and 0.01 are 82.91%, 80.22%, and 76.26%, respectively. As seen in the results from the CWRU dataset, the accuracy decreases as the imbalance ratio decreases. As expected, the mean accuracy using the XJTU dataset ($\approx 80\%$) was lower than that found from the CWRU dataset ($\approx 90\%$). It is important to recall that the defects in the XJTU dataset were generated naturally through run-to-failure tests. This confirms that bearing fault signals from the XJTU dataset had additional uncertainty and variability that is larger than that observed for the CWRU dataset.

TABLE 10. Accuracy Results of the Proposed Method for the XJTU Dataset.

Imbalance ratio (ρ)	Domain adaptation task using XJTU dataset		Mean
	1→2	2→1	
0.1	84.94 ± 5.27	80.88 ± 3.76	82.91
0.05	84.19 ± 6.59	76.24 ± 8.21	80.22
0.01	83.12 ± 9.44	69.40 ± 6.91	76.26

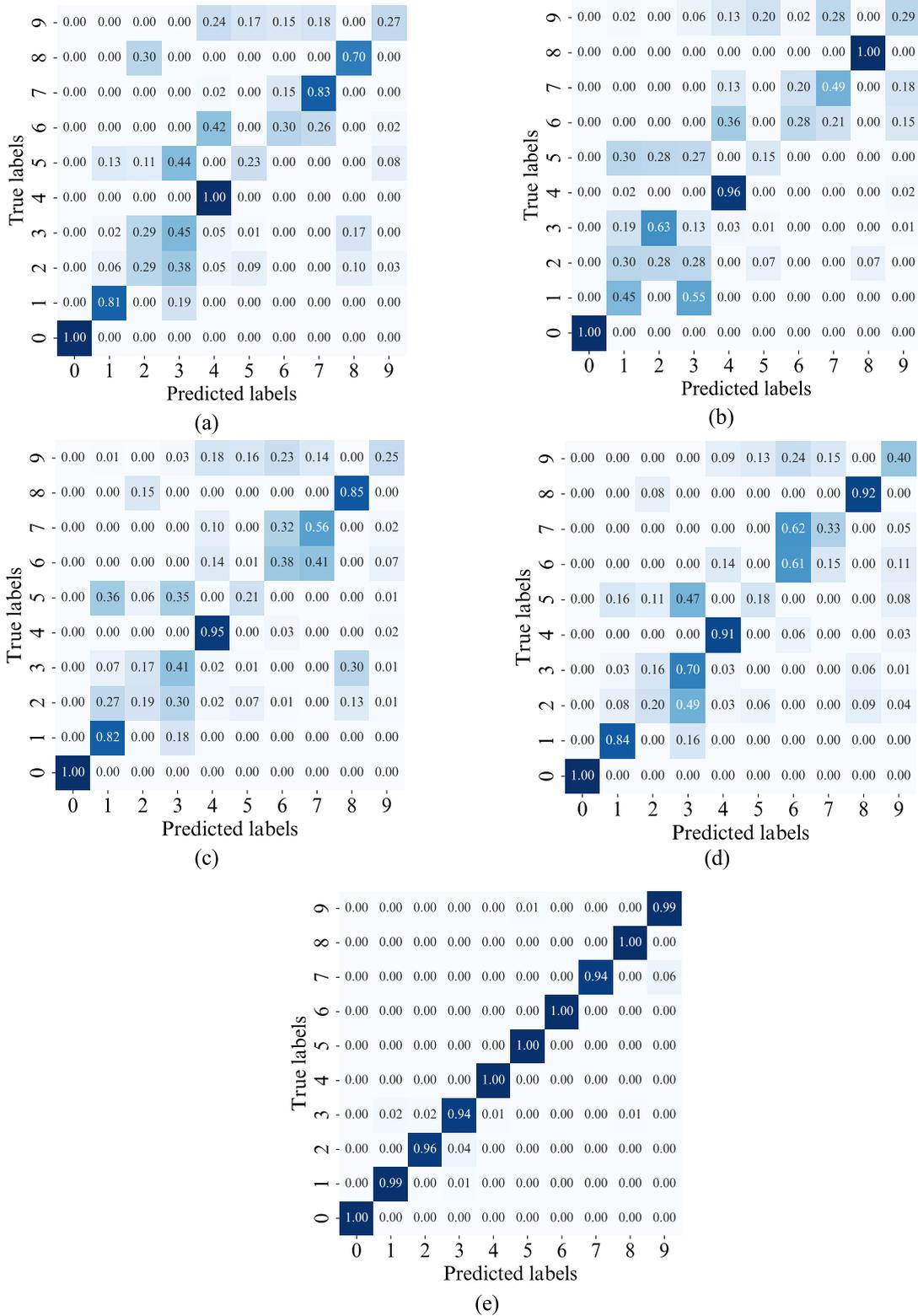


FIGURE 6. Confusion matrices of the domain-adaptation task for the CWRU dataset A→B with the imbalance ratio of 0.01: (a) DANN, (b) MMD, (c) DCTLN, (d) CIATL, and (e) the proposed method.

The performance of the proposed method is compared with that of different prior models in Fig. 10. The results are con-

sistent; i.e., the proposed method outperforms other models for all three severely imbalanced datasets, while maintaining

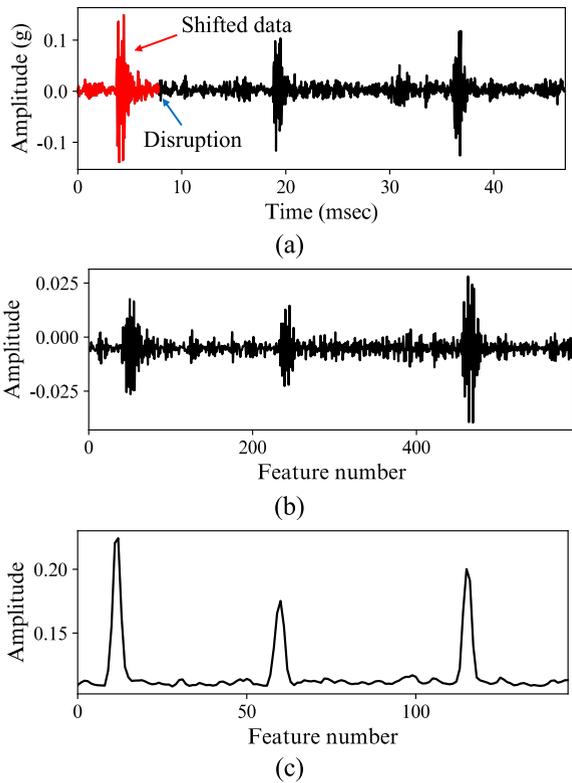


FIGURE 7. Time-shifting for repetitive learning: (a) time-shifted vibration signal inputted to the CNN, (b), and (c) visualization of the first and second feature maps, respectively.

accuracies higher than 75% even for the imbalance ratio of 0.01. It is worth noting that the models with use of the repetitive learning method again perform much better than the models without repetitive learning. The proposed method shows an accuracy of 76.26%, higher than all other models. These results confirm that the proposed method performs reliably even with a severely imbalanced dataset.

3) GIST DATASET

Six domain-adaptation tasks were conducted to evaluate the performance when severely imbalanced datasets were given for the rotating machine’s bearing fault, misalignment, and unbalance conditions. The results are summarized in Table 11. The means of the accuracy results for the imbalance ratios of 0.1, 0.05, and 0.01 were 91.90%, 89.10%, and 86.45%, respectively. Detecting energy changes caused by defects in the rotor at low speeds (600 RPM or less) presented a challenge for existing accelerometer-based monitoring systems [55]. The GIST dataset included low rotation speeds of 450 and 600 RPM. Despite the inherent difficulties in diagnosing low-speed bearings, the proposed method achieved a fault-diagnosis accuracy greater than 80%. This substantiates the robust performance of the proposed method, effectively handling additional uncertainty and variability in the fault signal.

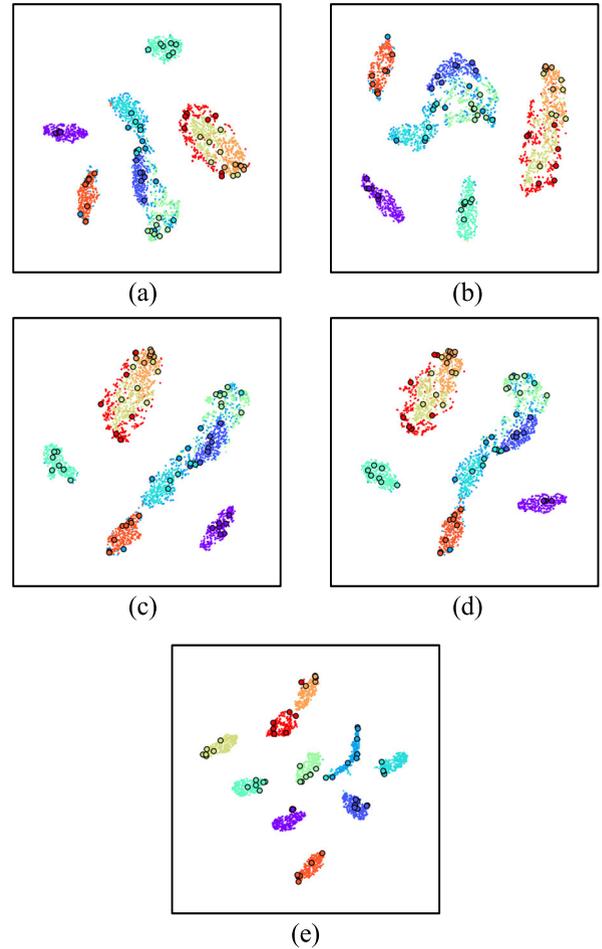


FIGURE 8. Visualization of the source- and target-domain features for the domain-adaptation task using the CWRU dataset A→B. The imbalance ratio was 0.01: (a) DANN, (b) MMD, (c) DCTLN, (d) CIATL, and (e) the proposed method.

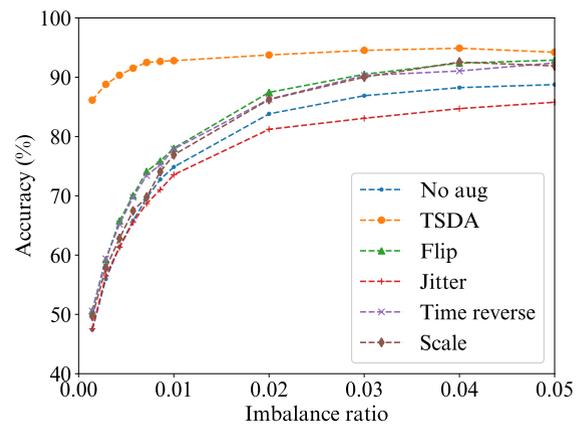
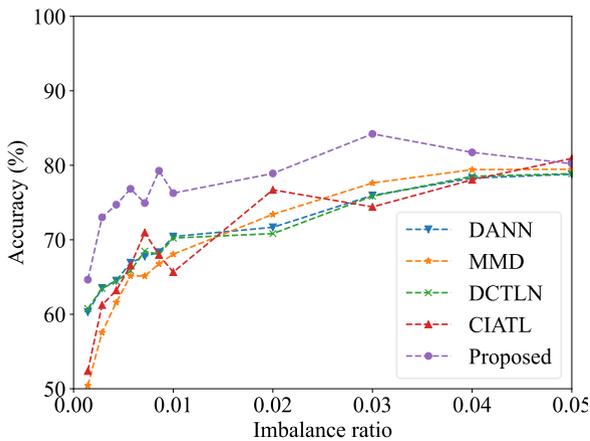
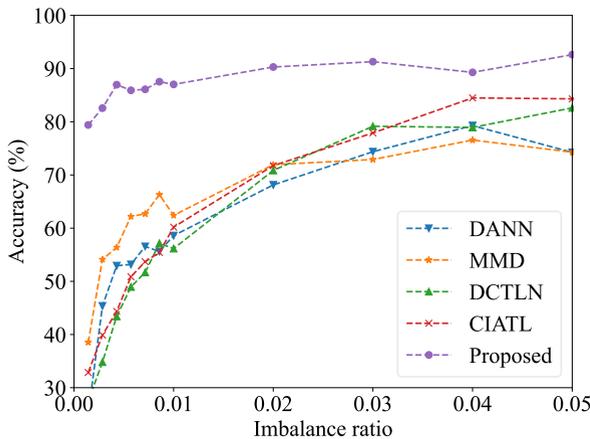


FIGURE 9. Comparison of the accuracy results for different augmentation methods.

The performance of proposed method is compared with existing models in Fig. 11. The results clearly show that the proposed method surpassed the others across all three severely imbalanced datasets. The accuracies of the proposed

TABLE 11. Accuracy Results of the Proposed Method for the GIST Dataset.

Imbalance ratio (ρ)	Domain adaptation task using the GIST dataset						Mean
	1→2	1→3	2→1	2→3	3→1	3→2	
0.1	90.15±6.59	87.32±8.37	92.05±3.09	94.39±2.70	94.59±2.70	92.92±3.99	91.90
0.05	87.07±6.12	87.83±7.75	92.14±2.96	86.39±7.83	90.87±3.58	90.29±3.74	89.10
0.01	90.38±4.61	82.39±5.90	86.74±4.50	84.86±7.16	88.44±2.28	85.87±4.96	86.45

**FIGURE 10. Comparison of the accuracy results of the proposed method with those of existing methods for the XJTU dataset.****FIGURE 11. Comparison of the accuracy results of the proposed method with those of existing methods for the GIST dataset.**

method were above 75%, even with only a single fault datum. Notably, the proposed method incorporating the proposed repetitive learning method significantly outperformed those without repetitive learning. These findings demonstrate the proposed method's reliability, even when severely imbalanced datasets are employed.

V. CONCLUSION

This paper presented a new repetitive learning method with domain adaptation techniques to address both class imbalance and cross-domain issues in rotating machine fault diagnosis. The proposed repetitive learning method was implemented by pseudo-labeling, weighted random sampling

augmentations, and time-shifting data augmentation. The proposed method was tested on three vibration datasets (CWRU, XJTU, and GIST), achieving accuracy rates of 90.79%, 76.26%, and 86.45%, even at an extreme imbalance ratio ($\rho = 0.01$), outperforming existing methods by 10% to 30%. The ablation study showed that the performance enhancement was attributed to the inclusion of the proposed repetitive learning module.

For the three data sets used in the study, the proposed repetitive learning method outperformed the existing methods even when the class imbalance and cross-domain problems co-existed. However, the results should be interpreted cautiously because the time-shifting operation changes the periodicity of vibration signals. This study aimed to only enhance diagnostic accuracy from the data-science perspective, not accounting for physics consistency. Physics-informed machine learning approaches, such as angle synchronous shifting, will be studied in the future to extend the applicability of the current study.

ACKNOWLEDGMENT

(Donghwi Yoo and Minseok Choi contributed equally to this work.)

REFERENCES

- [1] E. Iunusova, M. K. Gonzalez, K. Szipka, and A. Archenti, "Early fault diagnosis in rolling element bearings: Comparative analysis of a knowledge-based and a data-driven approach," *J. Intell. Manuf.*, vol. 35, no. 5, pp. 2327–2347, Jun. 2024.
- [2] H. Oh, J. H. Jung, B. C. Jeon, and B. D. Youn, "Scalable and unsupervised feature engineering using vibration-imaging and deep learning for rotor system diagnosis," *IEEE Trans. Ind. Electron.*, vol. 65, no. 4, pp. 3539–3549, Apr. 2018.
- [3] C. Wang, Y. Sun, and X. Wang, "Image deep learning in fault diagnosis of mechanical equipment," *J. Intell. Manuf.*, vol. 35, no. 6, pp. 2475–2515, Aug. 2024.
- [4] H. Oh, T. Shibutani, and M. Pecht, "Precursor monitoring approach for reliability assessment of cooling fans," *J. Intell. Manuf.*, vol. 23, no. 2, pp. 173–178, Apr. 2012.
- [5] H. Zhang, Y. Zhang, H. Meng, J. B. Lim, and W. Yang, "A novel global modelling strategy integrated dynamic kernel canonical variate analysis for the air handling unit fault detection via considering the two-directional dynamics," *J. Building Eng.*, vol. 96, Nov. 2024, Art. no. 110402.
- [6] T. Zhang, J. Chen, F. Li, K. Zhang, H. Lv, S. He, and E. Xu, "Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions," *ISA Trans.*, vol. 119, pp. 152–171, Jan. 2022.
- [7] S. Liu, H. Jiang, Z. Wu, and X. Li, "Data synthesis using deep feature enhanced generative adversarial networks for rolling bearing imbalanced fault diagnosis," *Mech. Syst. Signal Process.*, vol. 163, Jan. 2022, Art. no. 108139.
- [8] M. Kavianpour, A. Ramezani, and M. T. H. Beheshti, "A class alignment method based on graph convolution neural network for bearing fault diagnosis in presence of missing data and changing working conditions," *Measurement*, vol. 199, Aug. 2022, Art. no. 111536.

- [9] C. Jian and Y. Ao, "Imbalanced fault diagnosis based on semi-supervised ensemble learning," *J. Intell. Manuf.*, vol. 34, no. 7, pp. 3143–3158, Oct. 2023.
- [10] K. Yu, T. R. Lin, H. Ma, X. Li, and X. Li, "A multi-stage semi-supervised learning approach for intelligent fault diagnosis of rolling bearing using data augmentation and metric learning," *Mech. Syst. Signal Process.*, vol. 146, Jan. 2021, Art. no. 107043.
- [11] A. L. Marra, R. Juliani, and C. Garcia, "Data augmentation for vibration signals using system identification techniques," in *Proc. 5th Int. Conf. Syst. Rel. Saf. (ICSRS)*, Nov. 2021, pp. 281–285.
- [12] S. Kijpaiboonwat, W. Kongprawechnon, N. Chayopitak, W. Siriarporntham, C. Kingkan, and R. Pupadubsin, "Deep learning system with data augmentation for electric machinery fault diagnosis from vibration signals," in *Proc. 25th Int. Conf. Electr. Mach. Syst. (ICEMS)*, Nov. 2022, pp. 1–6.
- [13] L. Lin, S. Zhao, Y. Zhang, A. Wen, S. Zhang, J. Yan, Y. Wang, and Y. Zhou, "Purposeful data augmentation strategy and lightweight classification model for small sample industrial defect dataset," *IEEE Trans. Ind. Inform.*, vol. 20, no. 9, pp. 11475–11484, Sep. 2024.
- [14] C. Zhao and W. Shen, "Imbalanced domain generalization via semantic-discriminative augmentation for intelligent fault diagnosis," *Adv. Eng. Inform.*, vol. 59, Jan. 2024, Art. no. 102262.
- [15] X. Zhao, J. Yao, W. Deng, M. Jia, and Z. Liu, "Normalized conditional variational auto-encoder with adaptive focal loss for imbalanced fault diagnosis of bearing-rotor system," *Mech. Syst. Signal Process.*, vol. 170, May 2022, Art. no. 108826.
- [16] J. Zhang, S. Liu, W. Huang, F. Lyu, H. Xu, R. Yan, and B. Xu, "A cylinder block dynamic characteristics-based data augmentation method for wear state identification under data imbalance condition," *Mech. Syst. Signal Process.*, vol. 208, Feb. 2024, Art. no. 111036.
- [17] R. Wang, T. Li, Z. Gao, X. Yan, J. Wang, Z. Wang, and J. Gao, "A generative adversarial networks based methodology for imbalanced multidimensional time-series augmentation of complex electromechanical systems," *Appl. Soft Comput.*, vol. 153, Mar. 2024, Art. no. 111301.
- [18] W. Yang, H. Zhang, J. B. Lim, Y. Zhang, and H. Meng, "A new chiller fault diagnosis method under the imbalanced data environment via combining an improved generative adversarial network with an enhanced deep extreme learning machine," *Eng. Appl. Artif. Intell.*, vol. 137, Nov. 2024, Art. no. 109218.
- [19] P. N. Mueller, "Attention-enhanced conditional-diffusion-based data synthesis for data augmentation in machine fault diagnosis," *Eng. Appl. Artif. Intell.*, vol. 131, May 2024, Art. no. 107696.
- [20] R. Wang, Z. Chen, and W. Li, "Gradient flow-based meta generative adversarial network for data augmentation in fault diagnosis," *Appl. Soft Comput.*, vol. 142, Jul. 2023, Art. no. 110313.
- [21] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10684–10695.
- [22] S. Li, J. C. Ji, Y. Xu, K. Feng, K. Zhang, J. Feng, M. Beer, Q. Ni, and Y. Wang, "Dconformer: A denoising convolutional transformer with joint learning strategy for intelligent diagnosis of bearing faults," *Mech. Syst. Signal Process.*, vol. 210, Mar. 2024, Art. no. 111142.
- [23] Y. Xiao, X. Zhou, H. Zhou, and J. Wang, "Multi-label deep transfer learning method for coupling fault diagnosis," *Mech. Syst. Signal Process.*, vol. 212, Apr. 2024, Art. no. 111327.
- [24] H. Zheng, R. Wang, Y. Yang, J. Yin, Y. Li, Y. Li, and M. Xu, "Cross-domain fault diagnosis using knowledge transfer strategy: A review," *IEEE Access*, vol. 7, pp. 129260–129290, 2019.
- [25] J. Liu, F. Xie, Q. Zhang, Q. Lyu, X. Wang, and S. Wu, "A multisensory time-frequency features fusion method for rotating machinery fault diagnosis under nonstationary case," *J. Intell. Manuf.*, vol. 35, no. 7, pp. 3197–3217, Oct. 2024.
- [26] H. Zheng, Y. Yang, J. Yin, Y. Li, R. Wang, and M. Xu, "Deep domain generalization combining a priori diagnosis knowledge toward cross-domain fault diagnosis of rolling bearing," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.
- [27] R. Gopalan, R. Li, and R. Chellappa, "Unsupervised adaptation across domain shifts by generating intermediate data representations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2288–2302, Nov. 2014.
- [28] Z. Wang, X. He, B. Yang, and N. Li, "Subdomain adaptation transfer learning network for fault diagnosis of roller bearings," *IEEE Trans. Ind. Electron.*, vol. 69, no. 8, pp. 8430–8439, Aug. 2022.
- [29] K. Zhao, H. Jiang, Z. Wu, and T. Lu, "A novel transfer learning fault diagnosis method based on manifold embedded distribution alignment with a little labeled data," *J. Intell. Manuf.*, vol. 33, no. 1, pp. 151–165, Jan. 2022.
- [30] Q. Wang, Y. Xu, S. Yang, J. Chang, J. Zhang, and X. Kong, "A domain adaptation method for bearing fault diagnosis using multiple incomplete source data," *J. Intell. Manuf.*, vol. 35, no. 2, pp. 777–791, Feb. 2024.
- [31] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316–7325, Sep. 2019.
- [32] B. Han, X. Zhang, J. Wang, Z. An, S. Jia, and G. Zhang, "Hybrid distance-guided adversarial network for intelligent fault diagnosis under different working conditions," *Measurement*, vol. 176, May 2021, Art. no. 109197.
- [33] N. Lu, H. Xiao, Y. Sun, M. Han, and Y. Wang, "A new method for intelligent fault diagnosis of machines based on unsupervised domain adaptation," *Neurocomputing*, vol. 427, pp. 96–109, Feb. 2021.
- [34] R. Li, S. Li, K. Xu, X. Li, J. Lu, M. Zeng, M. Li, and J. Du, "Adversarial domain adaptation of asymmetric mapping with CORAL alignment for intelligent fault diagnosis," *Meas. Sci. Technol.*, vol. 33, no. 5, May 2022, Art. no. 055101.
- [35] Q. Qian, Y. Wang, T. Zhang, and Y. Qin, "Maximum mean square discrepancy: A new discrepancy representation metric for mechanical fault transfer diagnosis," *Knowl.-Based Syst.*, vol. 276, Sep. 2023, Art. no. 110748.
- [36] Q. Qian, H. Pu, T. Tu, and Y. Qin, "Variance discrepancy representation: A vibration characteristic-guided distribution alignment metric for fault transfer diagnosis," *Mech. Syst. Signal Process.*, vol. 217, Aug. 2024, Art. no. 111544.
- [37] Q. Qian, J. Luo, and Y. Qin, "Heterogeneous federated domain generalization network with common representation learning for cross-load machinery fault diagnosis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 54, no. 9, pp. 5704–5716, Sep. 2024.
- [38] Y. Zhang, B. Kang, B. Hooi, S. Yan, and J. Feng, "Deep long-tailed learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 9, pp. 10795–10816, Apr. 2023.
- [39] Z. Yang, R. O. Sinnott, J. Bailey, and Q. Ke, "A survey of automated data augmentation algorithms for deep learning-based image classification tasks," *Knowl. Inf. Syst.*, vol. 65, no. 7, pp. 2805–2861, Jul. 2023.
- [40] R. Shimizu, K. Asako, H. Ojima, S. Morinaga, M. Hamada, and T. Kuroda, "Balanced mini-batch training for imbalanced image data classification with neural network," in *Proc. 1st Int. Conf. Artif. Intell. Industries (AI4I)*, Sep. 2018, pp. 27–30.
- [41] S. Cha, S. Cho, D. Hwang, S. Hong, M. Lee, and T. Moon, "Rebalancing batch normalization for exemplar-based class-incremental learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 20127–20136.
- [42] C. Chen, J. Shi, M. Shen, N. Lu, H. Yu, Y. Chen, and C. Wang, "Pseudo-label guided sparse deep belief network learning method for fault diagnosis of radar critical components," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.
- [43] P. Cascante-Bonilla, F. Tan, Y. Qi, and V. Ordóñez, "Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning," in *Proc. AAAI Conf. Artif. Intell.*, May 2021, vol. 35, no. 8, pp. 6912–6920.
- [44] W. Zellinger, B. A. Moser, T. Grubinger, E. Lughofer, T. Natschläger, and S. Saminger-Platz, "Robust unsupervised domain adaptation for neural networks via moment alignment," *Inf. Sci.*, vol. 483, pp. 174–191, May 2019.
- [45] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mech. Syst. Signal Process.*, vols. 64–65, pp. 100–131, Dec. 2015.
- [46] B. Wang, Y. Lei, N. Li, and N. Li, "A hybrid prognostics approach for estimating remaining useful life of rolling element bearings," *IEEE Trans. Rel.*, vol. 69, no. 1, pp. 401–412, Mar. 2020.
- [47] D. Lee, C. Park, C.-W. Kim, J.-H. Jung, and H. Oh, "Digital twin finite element model for fault diagnosis of rotor system," *Trans. Korean Soc. Mech. Eng.-A*, vol. 47, no. 5, pp. 401–408, May 2023.

- [48] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li, "FixMatch: Simplifying semi-supervised learning with consistency and confidence," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 596–608.
- [49] J. Kuang, G. Xu, T. Tao, and Q. Wu, "Class-imbalance adversarial transfer learning network for cross-domain fault diagnosis with imbalanced data," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.
- [50] S. Zhou, S. Tian, L. Yu, W. Wu, D. Zhang, Z. Peng, and Z. Zhou, "Growth threshold for pseudo labeling and pseudo label dropout for semi-supervised medical image classification," *Eng. Appl. Artif. Intell.*, vol. 130, Apr. 2024, Art. no. 107777.
- [51] Z. Zheng, J. Fu, C. Lu, and Y. Zhu, "Research on rolling bearing fault diagnosis of small dataset based on a new optimal transfer learning network," *Measurement*, vol. 177, Jun. 2021, Art. no. 109285.
- [52] S. Schwendemann, Z. Amjad, and A. Sikora, "Bearing fault diagnosis with intermediate domain based layered maximum mean discrepancy: A new transfer learning approach," *Eng. Appl. Artif. Intell.*, vol. 105, Oct. 2021, Art. no. 104415.
- [53] L. Van der Maaten and G. E. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, Jan. 2008.
- [54] G. Iglesias, E. Talavera, Á. González-Prieto, A. Mozo, and S. Gómez-Canaval, "Data augmentation techniques in time series domain: A survey and taxonomy," *Neural Comput. Appl.*, vol. 35, no. 14, pp. 10123–10145, May 2023.
- [55] L. Song, H. Wang, and P. Chen, "Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 8, pp. 1887–1899, Aug. 2018.



MINSEOK CHOI received the B.S. and M.S. degrees in mechanical engineering from Gwangju Institute of Science and Technology, Gwangju, Republic of Korea, in 2021 and 2022, respectively, where he is currently pursuing the Ph.D. degree in mechanical engineering. His current research interest includes prognostics and health management for rotating machines.



HYUNSEOK OH received the B.S. degree in mechanical engineering from Korea University, Seoul, South Korea, in 2004, the M.S. degree in mechanical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2006, and the Ph.D. degree in mechanical engineering from the University of Maryland, College Park, MD, USA, in 2012. He is currently working as an Associate Professor with the School of the Mechanical Engineering, Gwangju Institute of Science and Technology, Gwangju, South Korea. His research interests include fault diagnostics, industrial artificial intelligence, and model verification and validation.



BONGTAE HAN (Member, IEEE) received the B.S. and M.S. degrees in mineral and petroleum engineering from Seoul National University, Seoul, South Korea, in 1981 and 1983, respectively, and the Ph.D. degree in engineering mechanics from Virginia Tech, Blacksburg, VA, USA, in 1991. He is currently working as a Keystone Professor of engineering with the Department of Mechanical Engineering, University of Maryland, College Park, MD, USA. He has co-authored a text book entitled *High Sensitivity Moiré: Experimental Analysis for Mechanics and Materials* (New York, NY, USA: Springer-Verlag, 1997) and edited two books. He has authored or co-authored 12 book chapters and more than 300 journal and conference papers in the field of microelectronics, photonics, and experimental mechanics. He was elected as a fellow of SEM and ASME, in 2006 and 2007, respectively. He is currently serving as an Editor-in-Chief for *Microelectronics Reliability*.



DONGHWI YOO received the B.S. and M.S. degrees in mechanical engineering from Gwangju Institute of Science and Technology, Gwangju, Republic of Korea, in 2021 and 2023, respectively. His current research interest includes prognostics and health management for rotating machines.