

Research Article

Unbalanced Fault Diagnosis Based on an Invariant Temporal-Spatial Attention Fusion Network

Jianhua Liu,¹ Haonan Yang,² Jing He ,² Zhenwen Sheng ,³ and Shou Chen⁴

¹College of Railway Transportation, Hunan University of Technology, Zhuzhou 412007, China

²College of Electrical and Information Engineering, Hunan University of Technology, Zhuzhou 412007, China

³Shandong Xiehe University, Jinan 250000, China

⁴Shenzhen Beauty Star Co., Ltd., Shenzhen 518000, China

Correspondence should be addressed to Zhenwen Sheng; xieheshengzhenwen@126.com

Received 15 September 2021; Accepted 10 March 2022; Published 30 March 2022

Academic Editor: Heng Liu

Copyright © 2022 Jianhua Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The health status of mechanical bearings concerns the safety of equipment usage. Therefore, it is of crucial importance to monitor mechanical bearings. Currently, deep learning is the mainstream approach for this task. However, in practical situations, the majority of fault samples have the issue of severe class unbalancing, which renders conventional deep learning inapplicable. Targeted at this issue, this paper proposes an invariant temporal-spatial attention fusion network called ITSA-FN for bearing fault diagnosis under unbalanced conditions. First, the proposed method utilizes the invariant temporal-spatial attention representation section, which consists of a pretrained convolutional auto-encoder model, a convolutional block attention module, and a long short-term memory network, to extract independent features and invariant features of spatial-temporal characteristics from input signals. Then, a multilayer perceptron is used to fuse and infer from the extracted features and design a new loss function from the focal loss for network training. Finally, this article validates proposed model's effectiveness through comparative experiments, ablation studies, and generalization performance experiments.

1. Introduction

Rolling bearings play a key role in mechanical equipment. Any tiny fault can affect the operation of the equipment or even produce latent threats to lives and property [1–4]. Therefore, it is necessary to monitor the health status of mechanical equipment. In recent years, deep learning has been used widely in the field of intelligent fault diagnosis of machinery [5–8]. Wang et al. proposed a deep learning similarity measure model for machine status classification [9]. Zheng et al. combined a capsule neural network and an abnormal module (XCN) and applied them to intelligent fault diagnosis [10]. Li et al. proposed a self-adaptive method in the field of mechanical fault diagnosis based on deep learning [11]. The abovementioned works have achieved good results in certain scenarios. However, these methods are mostly based on balanced mechanical datasets and disregard the situation of unbalanced mechanical statuses. In

reality, normal equipment samples are usually redundant, whereas the instances of faulty samples are rare. This unbalanced sample distribution renders conventional deep learning prone to biased attention (i.e., learning class features more from large populations while disregarding features of small populations), resulting in deteriorated identification accuracy.

Targeting the unbalanced data issue in fault diagnosis, researchers have proposed a variety of methods [12–14]. Currently, solutions to unbalanced classification mainly fall into two categories: data-level methods and algorithm-level methods. Data-level methods include oversampling for small population classes and undersampling for large population classes. State-of-the-art research favors the former [15]. Regarding the oversampling method, random oversampling (RAMO) is the simplest oversampling technique, but it is prone to overfitting [16]. Therefore, Chawla et al. proposed a synthetic minority oversampling technique

(SMOTE) to solve this problem by interpolating new synthetic samples [17]. However, SMOTE has the issues of overgeneralization and noise sensitivity. To address this issue, Bunk et al. proposed the Safe-level-SMOTE technology, which uses the K -nearest neighbor sample distribution to determine the “safety level” value and then synthesizes a new instance close to the maximum safety level [18]. Although Safe-level-SMOTE has certain advantages over SMOTE, its new instance usually has the problem of being far away from the decision boundary. In addition, Borderline-SMOTE (BSMOTE) [19], Cluster-SMOTE [20], Adaptive Synthetic Sampling (ADASYN) [21], Kmeans-SMOTE [22], and other cluster sampling methods are also achieved good results. However, these methods all rely excessively on the value of the parameter k . Thus, some weighting methods that do not rely on the nearest-neighbor method to obtain boundary instances are proposed, for example, majority weighted minority oversampling technology (MWMOTE), cluster-based majority weighted minority oversampling (Cluster-MWMOTE) [23], sample feature oversampling (SCOTE) [24], and expectation maximization local weighting minority oversampling (EM-LWMOTE) [25]. In addition to the abovementioned oversampling methods, deep generator models have also been widely used in the field. Li et al. proposed the assistant classifier Wasserstein generative adversarial network with gradient penalties (ACWGAN-GP) to generate high-quality samples for the minority class [26]. Yang et al. solved the unbalanced data issue through a data generation method based on a generative adversarial network and then utilized the multiscale convolutional neural network (MSCNN) for fault diagnosis on a harmonic drive [27]. Wang et al. designed the conditional variation auto-encoder generative adversarial network (CVAE-GAN) for unbalanced fault diagnosis [28]. The abovementioned methods achieved good results. However, on one hand, these methods may miss latent valuable information and, on the other hand, may produce additional computation loads to the networks. Therefore, for practical industrial cases, oversampling methods are usually not a good solution.

For algorithm-level methods, cost-sensitive learning is the mainstream way to deal with imbalanced classification problems [15]. It is a learning paradigm that allocates the cost of misclassification to the categories involved in the classification task. Existing cases include cost-sensitive neural networks [29], cost-sensitive decision trees [30], and cost-sensitive extreme learning machines [31]. However, most of these cases belong to the category of machine learning, and there are relatively few studies on deep neural networks. Thus, Peng et al. proposed a new type of bidirectional gated recurrent unit (BGRU) and combined it with a cost-sensitive active learning strategy for fault diagnosis [32]. Zhang et al. proposed the evolutionary cost-sensitive deep belief network (ECS-DBN) for unbalanced classification in which misclassification costs on training dataset are optimized through self-adaptive differential evolution [33]. In addition, in terms of cost-sensitive online learning, based on the cost-sensitive learning framework [34], cost-sensitive online gradient

descent (COG) was proposed. However, COG was limited to only extracting first-order information of the sample (weighted average of the gradient). To solve this problem, Zhao et al. proposed a series of cost-sensitive online classification algorithms with self-adaptive normalization and validated them [35]. It is worth mentioning that the key to these methods is to determine the cost for each class. However, in practical situations, it is usually difficult to determine the real misclassification costs for different classes. Focusing on this problem, Tsung et al. designed a focal loss function through reconstructing the standard cross-entropy loss in which adjustment factors were added to make adjustments to unbalanced samples. This method achieved a good performance in object detection tasks [36]. Based on this, He et al. proposed the spatial-temporal multiscale neural network (STMNN) and trained the model with focal losses, which effectively solved the unbalanced issue [37]. Although the abovementioned algorithm-level methods have achieved good results, these methods mostly instruct the network to adjust the unbalanced data during training and ignore the importance of the feature-extraction process.

It has been demonstrated by some research works [38–41] that transfer learning and information fusion both play very important roles in feature extraction. For transfer learning, it may be viable to transfer the features of balanced data to unbalanced data, thereby diminishing the negative impacts of the unbalanced issue. For information fusion, the amount of information acquired by the network from the features can be increased by fusion [42], thereby upgrading the network’s confidence. Motivated by the abovementioned methods, this paper focuses on both feature-extraction methods and algorithm-level learning strategies, and presents an invariant temporal-spatial attention fusion network (ITSA-FN) to solve the unbalanced data issue. The contributions are summarized as follows:

- (1) Proposed method designs an invariant temporal-spatial attention fusion network for feature extraction and inference representation of unbalanced data, which has achieved sound results.
- (2) A new joint constrained focal loss is designed, which is capable of facilitating the network’s learning of relevant features from the algorithm aspect and adjusting the negative impacts of unbalanced data.
- (3) This article designs comparative experiments, ablation studies, and generalization performance analysis on vibration and current datasets with a variety of unbalanced ratios, and validated the effectiveness and superiority of our proposed model.

The rest of the paper is organized as follows. Section 2 introduces the theoretical background, while Section 3 describes in detail the framework and learning strategy of our proposed method. Section 4 discusses the experiments, including the experimental datasets, experimental details, network parameters, and experimental result analysis. Section 5 presents the conclusions of this study and suggests future research directions.

2. Theoretical Background

In this section, we provide introductions to the theory of the network models, convolutional auto-encoder (CAE), and convolutional block attention module (CBAM) used in this paper.

2.1. CAE. A CAE is a combination of a convolutional neural network (CNN) and an auto-encoder (AE), which is a more efficient unsupervised feature extractor. Compared with an AE, a CAE replaces the fully connected layer with a convolutional layer and a pooling layer (downsampling layer) in the encoder and utilizes a deconvolution layer and an upsampling layer in the decoder. The computation process is as follows.

During encoding, let the input samples be x_m . After the convolution and pooling operations, the feature maps of the hidden layer can be written as [43]

$$h(x_m) = \text{pool}(\sigma_1(x_m * W^k + b^k)), \quad (1)$$

where “ W ” contains the initialized convolution kernels, the total number of which is k ; “ b ” denotes a bias of convolutional kernel; “ σ_1 ” is the activation function; “ $*$ ” denotes the convolution operation; and “pool” denotes the pooling operation. The convolution of the kernels and the input x_m generates k feature maps $h(x_m)$.

During decoding, an unreasonable kernel size may lead to matrix block overlapping during deconvolution. Therefore, this paper replaces the deconvolution step with upsampling and convolution, and recovers the input size through nearest-neighbor interpolation during upsampling. However, since upsampling has already recovered the feature map size, it is sufficient to conduct a simple convolution, as illustrated in the following [43]:

$$R = \sigma_1(r * \hat{W}^k + \hat{b}^k), \quad (2)$$

where “ σ_1 ” is the leaky ReLU activation function, “ \hat{W} ” is the convolution kernel of the decoder, “ \hat{b} ” is the bias of the decoder, “ r ” is the hidden layer feature map $h(x_m)$ after size recovering through upsampling interpolation, and “ R ” consists of the reconstructed data.

2.2. Convolutional Block Attention Module. CBAM is a simple yet effective feed-forward neural-network attention module. As illustrated in Figure 1, for a given intermediate feature mapping $F \in \mathbb{R}^{C \times H \times W}$ as input, the module infers the attention mappings along two independent dimensions (channel and spatial) in order. The computational equations can be written as [44]

$$\begin{aligned} F' &= M_c(F) = \sigma_2(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))), \\ F'' &= M_s(F') = \sigma_2(f^{7 \times 7}([\text{AvgPool}(F'); \text{MaxPool}(F')])), \end{aligned} \quad (3)$$

where “ \otimes ” denotes elemental multiplication, “CBAM” is the convolutional block attention module, “ F' ” is output of channel attention mapping, “ F'' ” is output of channel-spatial attention mapping, $M_c \in \mathbb{R}^{C \times 1 \times 1}$ is the 1D channel attention mapping, $M_s \in \mathbb{R}^{1 \times H \times W}$ is the 2D spatial attention mapping, “MLP” is the multilayer perceptron mapping representation, “AvgPool” is the average pooling operation, and “MaxPool” is the maximum pooling operation.

$f^{7 \times 7}$ denotes convolution with filters of size 7×7 . “ σ_2 ” is the sigmoid activation function.

3. The Proposed Method

Targeting the issue of unbalanced fault categories in actual situations, as illustrated in Figure 2, this paper presents an ITSA-FN model, which combines pretrained CAE modules, CBAM block, and LSTM for temporal-spatial feature extraction, and uses MLP to fuse and infer different modal features. Otherwise, we train the model with a newly designed joint normalized constrained loss function.

3.1. Invariant Temporal-Spatial Attention Fusion Neural Network

3.1.1. Transfer Feature Extraction. To relieve the biased learning issue (e.g., where the network leans toward classes with larger sample populations) of the feature-extraction network (i.e., CAE) due to unbalanced samples, this paper adopts a transfer-learning method. First, this article trains the feature CAE with balanced samples under ideal experimental conditions. Then, after learning the distributions of the balanced samples, the unbalanced vibration signal (x_v) and current signal (x_c) are fed into the network for fine-tuning, thereby obtaining the data’s low-level features. This is the encoding-decoding process, which can be expressed as follows:

$$\begin{aligned} h_m(x_m) &= \text{CAE}_m^{\text{encoder}}(x_m; \theta_m^{\text{encoder}}), m \in \{v, c\}, \\ R_m &= \text{CAE}_m^{\text{decoder}}(h_m(x_m); \theta_m^{\text{decoder}}), m \in \{v, c\}, \end{aligned} \quad (4)$$

where $h_m(x_m)$, R_m , $\theta_m^{\text{encoder}}$, $\theta_m^{\text{decoder}}$, $\text{CAE}_m^{\text{encoder}}$, and $\text{CAE}_m^{\text{decoder}}$ denote the pretrain CAE encoding representation, decoding (reconstruction) representation, encoder network parameter, decoder network parameter, encoder network section, and decoder network section of modal m , respectively. In addition, the structure of the pretrained model is illustrated in Figure 3, in which the module in green is the reserved model, and the section in gray is the classifier that guides the network during supervised training.

3.1.2. Temporal-Spatial Attention Feature Representation. It has been proved by some research [45–47] that fault signal sequences are spatially and temporally correlated, and fault symptoms concentrate in the vicinity of the triggering time. Applying an attention mechanism and adopting a network with memory functionalities can make the network focus actively on the vicinity around fault moments and capture the contextual information, respectively, thereby improving

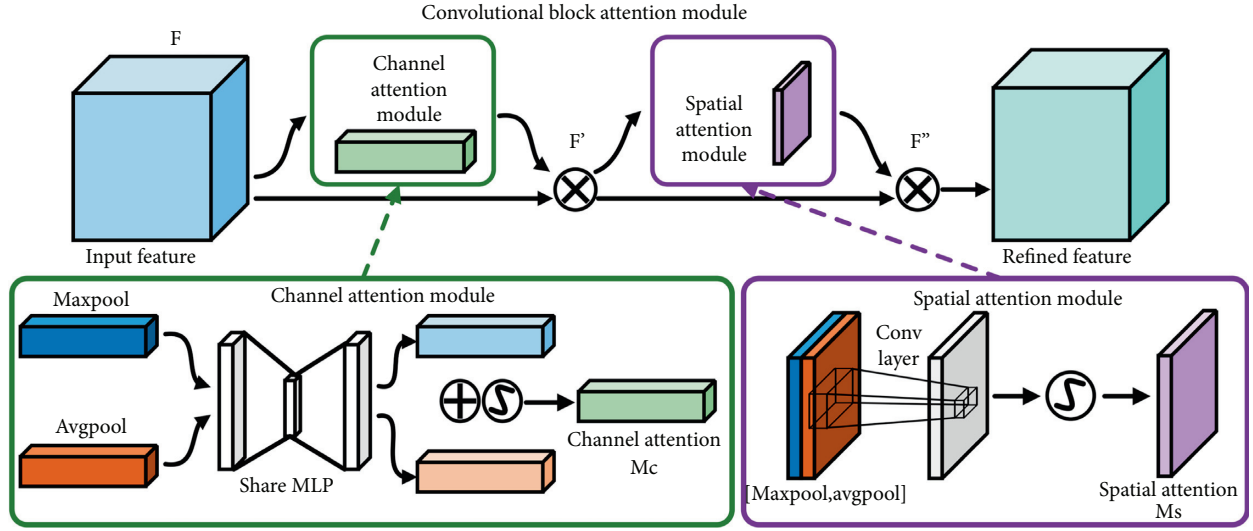


FIGURE 1: The framework of CBAM.

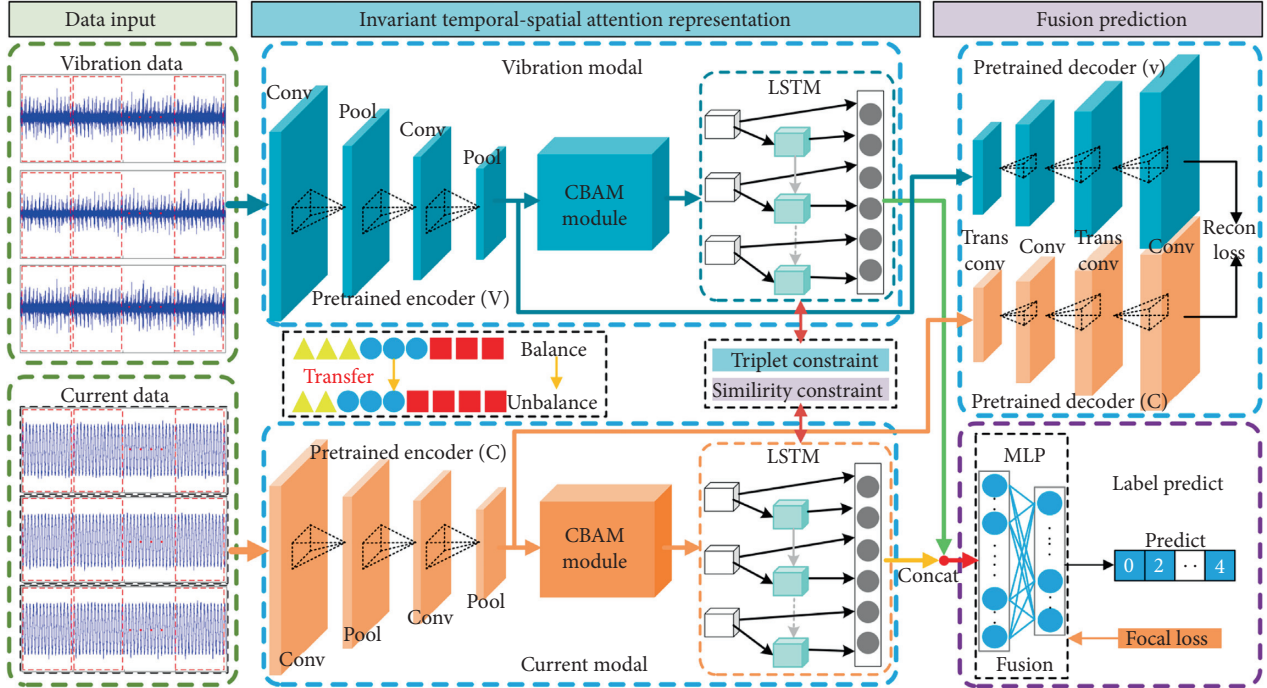


FIGURE 2: The framework of ITSA-FN.

the learning efficiency of fault features [48, 49]. Therefore, in this paper, we adopt the CBAM module, which is capable of capturing channel attention and spatial attention, and the LSTM to acquire the spatial-temporal characteristics of the faults and facilitate the network's ability to concentrate on learning fault information. The process is as follows.

First, CBAM is applied to the obtained encoding representations of the two modalities to obtain the spatial characteristics, which can be written as

$$Att_m = CBAM_m(h_m(x_m); \theta_m^{CBAM}), m \in \{v, c\}, \quad (5)$$

where Att_m , $CBAM_m$, and θ_m^{CBAM} denote the attention representation, the convolutional block attention network, and the attention network parameter of modal m , respectively.

Then, the attention representation is fed into the LSTM network to acquire the temporal characteristics, which can be written as

$$M_m = LSTM_m(Att_m; \theta_m^{LSTM}), m \in \{v, c\}, \quad (6)$$

where M_m , $LSTM_m$, and θ_m^{LSTM} denote the spatial-temporal attention representation, the LSTM network, and network parameter of modal m , respectively.

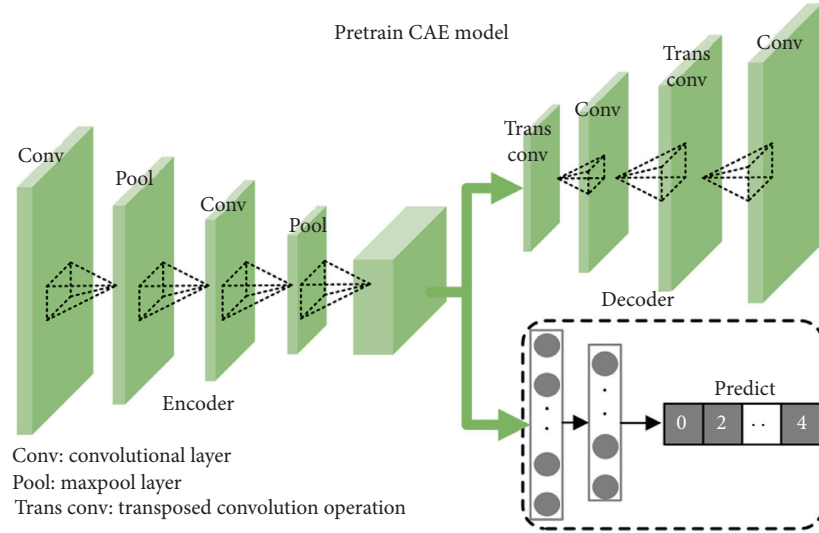


FIGURE 3: The framework of pretrain CAE.

3.1.3. Fusion Inference Representation. Information-fusion methods can improve the network's confidence by increasing the information volume. Therefore, from the point of view of information fusion, this paper fuses the data of two modalities (vibration and current) to optimize network performance, and the process is as follows.

After acquiring the spatial-temporal attention representation of the two modalities, we concatenate and fuse them into fusion representation

$$\text{Fusion} = M_v \oplus M_c, \quad (7)$$

where " \oplus " denotes vector concatenation, and "Fusion" denotes the fusion vector after concatenation. Then, the fusion spatial-temporal attention representation is fed into the inference network (MLP) for fusing and fault inferences, which can be written as

$$\text{Inf} = \text{MLP}(\text{Fusion}; \theta^{\text{MLP}}), \quad (8)$$

where "Inf" denotes the inference output, " θ^{MLP} " is the inference network parameter, and "MLP" is the multilayer perceptron with leaky ReLU and tanh activation functions.

3.2. Learning Strategy. This article designs a new normalized constrained loss function based on the imbalanced issue and the network structure, which will be described in detail in this section.

3.2.1. Task Loss. The focal loss can self-adaptively adjust the impacts from a variety of sample populations through the addition of adjustment factors to the cross-entropy loss. Therefore, in order to better solve the issue of unbalanced sample, this paper adopts the focal loss as the task loss to guide network's learning, the computation equation of which can be written as [36]

$$L_{\text{focal}}(P) = -\alpha(1-P)^\gamma \log(P), \quad (9)$$

where "P" denotes the probability of a sample's correct classification, " α " denotes a weighting factor, and $(1-P)^\gamma$ denotes the adjustment factor, where " γ " is the adjustable focal parameter. In addition, $P = \text{Inf}$ is the inference network output ($\alpha = 0.25$, and $\gamma = 2$).

3.2.2. Recon Loss. The recon loss aims to minimize the distance between the input data and the reconstructed data. It is used in the fine-tuning training of a pretrained CAE model. In addition, integrating the recon loss into learning can facilitate the network's learning of trivial representations and suppress input noise. This paper utilizes the mean squared error as the recon loss, the computation equation of which can be written as [50]

$$L_m^{\text{recon}} = \|x_m - R_m\|_2^2, m \in \{v, c\}, \quad (10)$$

where x_m and R_m denote the original input data and the reconstructed data, respectively, and $\|\cdot\|_2^2$ denotes the square of the L_2 norm.

3.2.3. Similarity Loss. When performing the fusion task, directly inputting the connected feature of each modality into the MLP fusion layer will make the network unable to effectively explore the interactions between the modalities, resulting in the problem of uncomprehensive fusion representation. Therefore, to obtain more comprehensive fusion representations, this paper adds a similarity constraint between the spatial-temporal attention representations of the two modalities to facilitate the network's capturing of invariant spatial-temporal attention representations. The central moment deviation (CMD) metric measures the distance between two distributions through matching differences between high-order moments. Its computation is simple and efficient, and capable of reducing the computation cost of the network [51]. Therefore, this article utilizes CMD as the similarity loss, the definition of which can be written as follows.

Let X and Y be bounded random samples having distributions of p and q in the tight interval of $[a, b]^N$, respectively. Then, the central moment deviation normalizer CMD_K can be defined as empirical estimation of CMD metric as follows [51]:

$$CMD_K(X, Y) = \frac{1}{b-a} \|E(X) - E(Y)\|_2 + \sum_{k=2}^K \|C_k(X) - C_k(Y)\|_2, \quad (11)$$

where C_k and $E(X)$ can be written as follows:

$$C_k(X) = E((x - E(X))^k)E(X) = \frac{1}{|X|} \sum_{x \in X} x, \quad (12)$$

where $E(X)$ is the empirical expectation vector of sample X , and $C_k(X)$ is a vector on the x -axis of the central moment of all samples of order k . The CMD similarity loss used in this paper can be written as follows:

$$L_{sim} = CMD_K(M_v, M_c). \quad (13)$$

3.2.4. Triplet Loss. Before the fusion of invariant spatial-temporal attention representations, to ensure high-level similarity relationships between the two modalities, this paper minimizes the distance among all samples with a similar semantic (class) from different modalities through the addition of a triplet boundary constraint L_{trip} while maximizing the distance among representations without similarities [52]. For the vibration modal, we construct a triplet representation of $(M_v, M_{c+}, \text{and } M_{c-})$ in which the current representation is semantically positively correlated with vibration representation M_v , while the current representation M_{c-} is negatively correlated with M_v . Then, the triplet boundary loss with the vibration modal as the anchor can be written as [52]

$$L_{trip}^v = \max(d(M_v, M_c^+) - d(M_v, M_c^-) + \text{margin}, 0). \quad (14)$$

Similarly, the triplet boundary loss with the current modal as the anchor can be written as follows:

$$L_{trip}^c = \max(d(M_c, M_v^+) - d(M_c, M_v^-) + \text{margin}, 0), \quad (15)$$

where “ d ” denotes the Euclidian distance $\|\cdot\|_2^2$, “margin = 1” denotes the boundary value (ensuring that the loss function $L_{trip} \geq 0$), and $M_{c/v}$ denotes the invariant spatial-temporal attention representations of the two modalities.

Combining (14) and (15), this paper obtains the overall triplet boundary loss as follows:

$$L_{trip} = L_{trip}^v + L_{trip}^c. \quad (16)$$

3.2.5. Total Loss. In summary, this paper combines the abovementioned losses and proposes a new joint constrained focal loss function as follows:

$$L_{total} = L_{focal} + \beta L_v^{recon} + \lambda L_c^{recon} + \eta L_{sim} + \delta L_{trip}, \quad (17)$$

where $\beta, \lambda, \eta,$ and δ denote the adjustment factors of the loss function that adjust the contributions from each loss. This paper trains the network by minimizing this objective loss. The experimental procedures and details will be described in Section 3.3.

3.3. Methodology Procedures. In this section, we introduce the procedures of the method. The flowchart is shown in Figure 4.

According to the content in Figure 4, the details of the method steps are as follows:

- (1) Step 1: Preprocess the data. Experiment normalizes the vibration and current input data into the interval $[-1, 1]$ and converts the normalized 1D data segments into 2D through reshaping, so that they could be fed into the 2D convolutional layer (i.e., the data are reshaped from 400×400 into $400 \times 1 \times 20 \times 20$).
- (2) Step 2: Acquire the pretrained CAE model. Experiment inputs the preprocessed source domain data (vibration and current) into our designed CAE network and then conducts the training and saves the resultant optimal model.
- (3) Step 3: Acquire the low-level feature representations. Experiment inputs the preprocessed target domain data into the pretrained model obtained in Step (2) and acquires the encoding and decoding representations of the CAE to guide the network in further learning.
- (4) Step 4: Acquire the spatial-temporal attention feature representation. Experiment inputs the encoding representation obtained in Step (3) into the CBAM module and LSTM network in order, thereby obtaining the spatial-temporal attention representations of the two modalities.
- (5) Step 5: Acquire the fusion inference representations. Experiment concatenates the spatial-temporal attention feature representations of the vibration and current datasets obtained in Step (4) and inputs them into the MLP for fusion, thereby obtaining the final inference representation.
- (6) Step 6: Compute the loss function. Experiment calculates the loss function for each component, respectively, according to the learning strategy and calculates the total objective loss L_{total} with the adjustment factors.
- (7) Step 7: Iteration and training. Experiment minimizes the total objective loss to train proposed network.
- (8) Step 8: Save the optimal model. Experiment saves the resultant optimal model for testing.

4. Experiment Results and Discussion

This section mainly discusses the experimental datasets, the network configuration, and the experimental procedures.

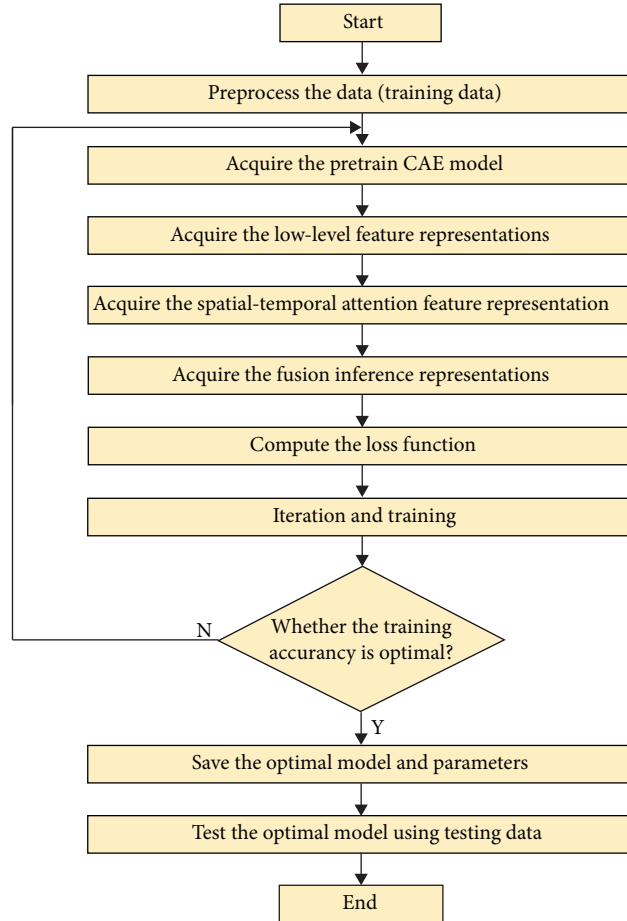


FIGURE 4: The flowchart of the method.

TABLE 1: Dataset interpretation.

Data domain	Code	State	Element	Combination	Extent	Speed	Torque	Radial force
Target domain	K001	Normal	—	—	—	900	0.7	1000
	KA04	Pitting	Outer ring	Single	1	900	0.7	1000
	KA15	Identification	Outer ring	Single	1	900	0.7	1000
	KB27	Pitting	Outer/inner ring	Multiple	2	900	0.7	1000
	KI21	Pitting	Inner ring	Single	1	900	0.7	1000
Source domain	K001	Normal	—	—	—	1500	0.1	1000
	KA04	Pitting	Outer ring	Single	1	1500	0.1	1000
	KA15	Identification	Outer ring	Single	1	1500	0.1	1000
	KB27	Pitting	Outer/inner ring	Multiple	2	1500	0.1	1000
	KI21	Pitting	Inner ring	Single	1	1500	0.1	1000

4.1. Datasets. To validate the effectiveness of proposed approach, this paper conducts experiments on bearing damage datasets released by the University of Paderborn, Germany [53]. The datasets consist of manual damages and accelerated lifetime damages in two sections. To be closer to practical situations, this paper adopts the accelerated lifetime damage data.

Specifically, the data experiments used are as follows. This paper considers two different operating conditions: the target domain (rotational speed $N=900$ rpm, load torque $T=0.7$ Nm, and radial force $F=1000$ N) and the source domain (rotational speed $N=1500$ rpm, load torque

$T=0.1$ Nm, and radial force $F=1000$ N). The measured vibration and current signals of the equipment assembled with bearings under five different health statuses are illustrated in Table 1, in which each operating condition (rotational speed, torque, and radial force) contains data files collected by 20 sensors. This article selects one from them for experiment. Furthermore, this paper extracts 160,000 (400×400) data points from each file for experiments (in which one sample has 400 data points; thus, there were a total of 400 samples) and partitions them with a ratio of 3:1 into the training set and testing set, respectively.

TABLE 2: Unbalanced data distribution.

Imbalance rate\code	K001	KA04	KA15	KB27	KI21
10:5	400	200	200	200	200
10:4	400	160	160	160	160
10:3	400	120	120	120	120
10:2	400	80	80	80	80
10:1	400	40	40	40	40
10:1:5:5:5	400	40	200	200	200
10:5:1:5:5	400	200	40	200	200
10:5:5:1:5	400	200	200	40	200
10:5:5:5:1	400	200	200	200	40
10:5:4:3:2	400	200	160	120	80

TABLE 3: Hyperparameter settings.

Hyperparameter	Value
β	0.7
λ	0.1
η	1.0
δ	0.1
K	3
lstm dim	64
mlp dim	128

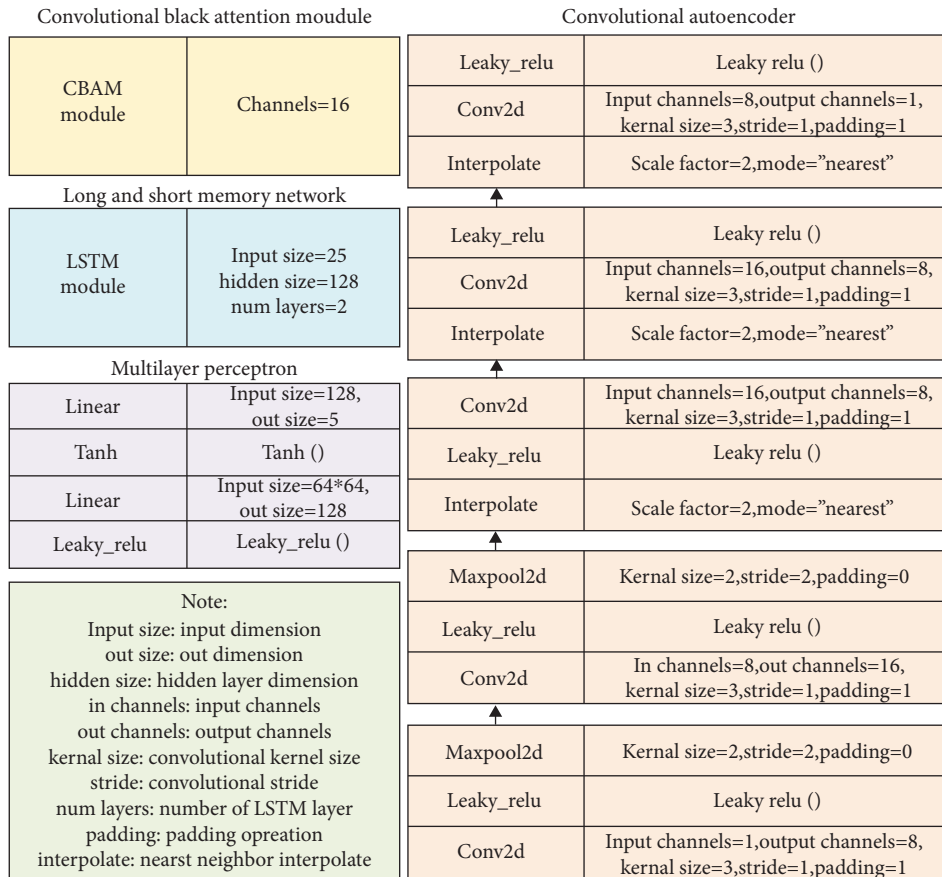


FIGURE 5: The structure setting of ITSA-FN.

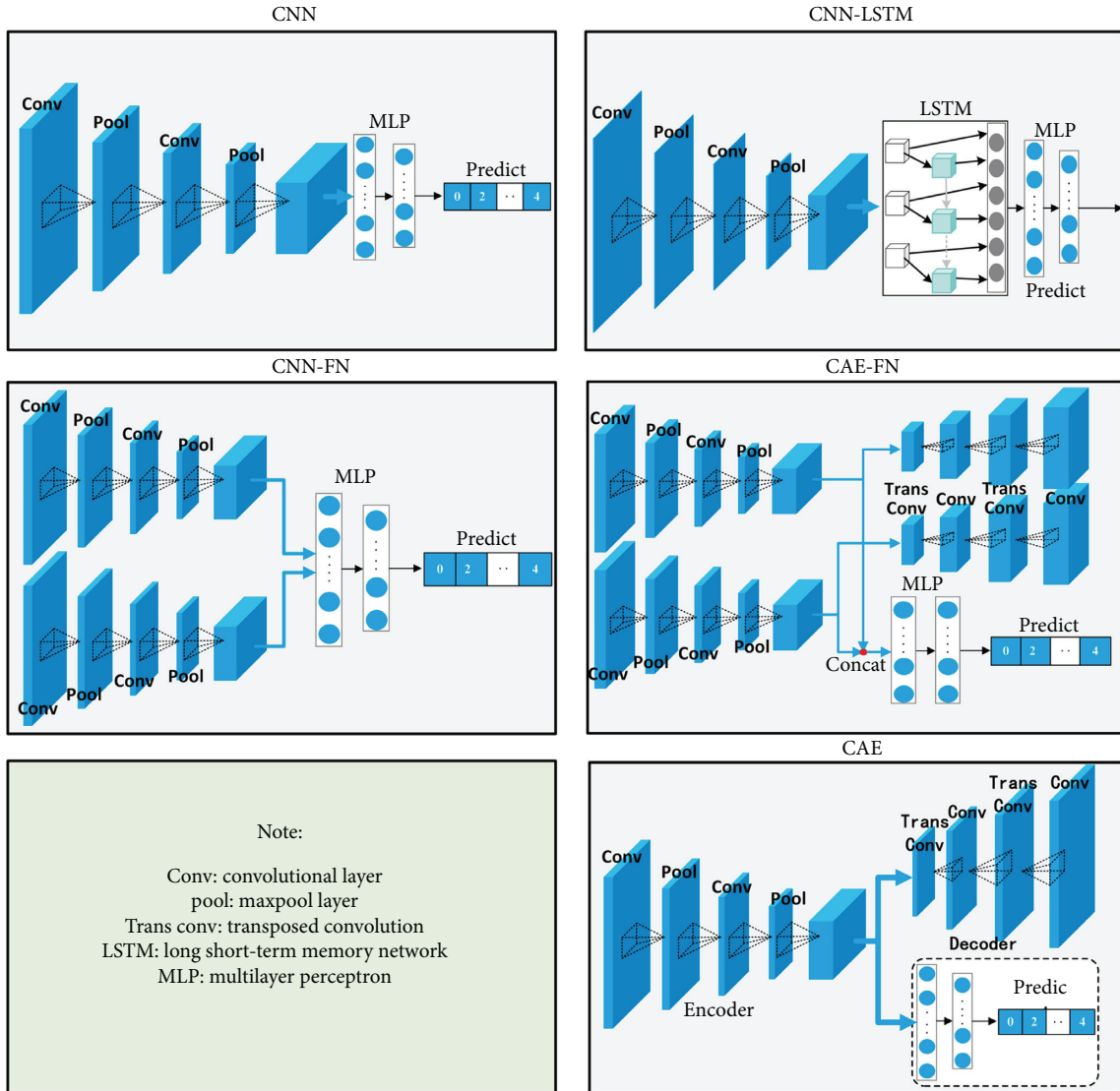


FIGURE 6: The structure of comparative model.

Furthermore, we configure the distribution of our experimental data in the manner illustrated in Table 2, which lists the sample population of each class under different unbalanced ratios.

4.2. Experimental Network Configuration. This paper conducts the experiments with the open-source machine-learning framework PyTorch. The environment configurations are shown as follows: (1) CPU (AMD Ryzen 5 2600X Six-Core Processor, 3.60 GHz), (2) 16 GB of RAM, (3) GPU (NVIDIA GeForce GTX 1660, 6G), and a (4) code library and environment (PyTorch = 1.2.0 and Python = 3.7.9, respectively).

In addition, this paper selects network hyperparameters using a grid search algorithm, in which the optimal parameters are selected from the following finite sets: $\beta \in \{0.6, 0.7, 0.8, 0.9, 1.0\}$, $\lambda \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$, $\eta \in \{0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$, $\delta \in \{0.01, 0.1, 0.2, 0.3\}$, $K \in \{1, 2, 3, 4, 5\}$, $\text{lstm dim} \in \{32, 50, 64\}$, and $\text{mlp dim} \in \{128, 256, 512, 1024\}$, where β , λ , η , and δ are the adjustment factors of the loss function; K

is the CMD order; “lstm dim” is the LSTM hidden layer dimension; and “mlp dim” is the MLP dimension. The grid searching saves the optimal model and returns the respective hyperparameters, as illustrated in Table 3.

This article configures the optimizer, learning rate, random seed, epoch, and batch size to be “Adam,” 0.01, 123, 300, and 100, respectively, for the training. In addition to these key hyperparameters, the structural details of our network components are illustrated in Figure 5.

4.3. Results and Discussion. For an effective validation of the proposed method, in this subsection, this paper designs comparative experiments, ablation studies, and generalization performance experiments.

4.3.1. Experiment Analysis. In this subsection, for unbalanced samples (vibration and current) of different class

TABLE 4: Experimental results I.

Method/ratio	10:5	10:4	10:3	10:2	10:1
CNN	93/76.33	92.69/73.85	90.91/70.91	89.44/68.88	87.14/71.43
CAE	92.33/78.33	90.77/70.38	92.73/73.64	91.67/71.11	89.29/67.14
CNN-LSTM	93.33/76	92.31/76.54	91.82/72.73	90.55/70	88.57/71.43
CNN-fusion	94.33	95	93.64	93.33	91.43
CAE-fusion	94	93.46	94.09	93.89	92.14
Our method	96.33	96.92	96.81	96.67	95

Index: accuracy: unit: %. Note: the first three are the test results of the single-mode method (the left is vibration, and the right is current).

TABLE 5: Experimental results II.

Method/ratio	10:1:5:5:5	10:5:1:5:5	10:5:5:1:5	10:5:5:5:1	10:5:4:3:2
CNN	92.31/70.77	91.54/70.38	91.54/72.69	91.92/73.46	91.66/73.33
CAE	92.69/71.92	88.85/71.15	89.61/70.38	89.23/75	87.5/75.42
CNN-LSTM	94.61/72.31	91.92/74.23	93.08/73.46	92.69/75.77	92.92/74.17
CNN-fusion	95.38	94.23	93.85	94.23	94.17
CAE-fusion	95.77	93.46	94.23	94.61	94.16
Our method	97.69	96.15	96.54	97.69	97.08

Index: accuracy: unit: %. Note: the first three are the test results of the single-mode method (the left is vibration, and the right is current).

population ratios, this paper designs a single modal (vibration or current) fault diagnosis model based on a CNN, CAE, a CNN with long short-term memory (CNN + LSTM), and two modal fusion models based on convolutional fusion neural network (CNN-fusion) and CAE fusion neural network (CAE-fusion) to compare with our proposed model. The structure of comparative model is illustrated in Figure 6, where the parameters of the convolutional layer and the MLP are identical to our network. The comparative experiments results are listed in Tables 4 and 5.

It can be observed from the results of the single modal methods that the vibration signal features are more prominent than those of the current signal, which is consistent with practical situations. Compared with CNN, the CNN-LSTM model achieves better performance under the different unbalanced conditions, and this benefits from the memory functionality of LSTM. In addition, compared with the single modal methods, it can be clearly observed that the fusion models show obvious improvements in diagnosis performance. This means that the network’s confidence and performance can be effectively upgraded from the richer information obtained through the fusion feature.

In order to further validate the superiority of our network’s classification performance, we visualize the confusion matrices of CNN-LSTM (vibration modal), CNN-fusion, CAE-fusion, and our method in Figure 7 with an unbalanced ratio of 10:5:4:3:2. It can be observed from the confusion matrices that our approach achieves the best classification performance in each fault category. This may be because our model integrates the spatial-temporal characteristics in the feature representations and adopts a reasonable constrained-learning strategy. Therefore, in order to validate the performance and rationality of our method, we carry out further analysis below.

4.3.2. Visualization Analysis. In this subsection, we validate the rationality and effectiveness of our method through T-distributed stochastic neighbor embedding (T-SNE) from an intuitive perspective.

To reduce the workloads, this paper carries out the experiments with a representative unbalanced dataset (with an unbalanced ratio of 10:5:4:3:2) and performs a T-SNE visualization analysis on our network’s feature representation process. As illustrated in Figure 8, the five plots show the distributions of the semantic features of the original input, encoder representation, spatial-temporal attention representation, MLP fusion representation, and the final inference representation. In each plot, the dots in different colors represent samples of different classes. In addition, the closer the dots of the same color and the farther away from the dots of different colors, the better the network’s performance. It can be observed from Figure 8 that the samples’ clustering effect becomes better in the order of network’s modules. This means that the process of our network is reasonable when it carries out a fault diagnosis under unbalanced conditions.

4.3.3. Ablation Studies Analysis. In the previous subsection, this paper visualizes the representation process of our network and validated the network’s classification performance through the confusion matrices. However, there is a lack of quantitative assessment indices that could assess the model’s performance. Therefore, in this subsection, this paper conducts an ablation study on the model’s structural design and learning strategy, and utilizes the inference accuracy as an assessment index (unbalanced ratio = 10:5:4:3:2). The results are illustrated in Table 6.

It can be observed from the loss function results in Table 6 that the network is most sensitive to the focal loss. Because when it is used as the task loss, it has decisive impacts on the network’s output, and its penalty mechanism

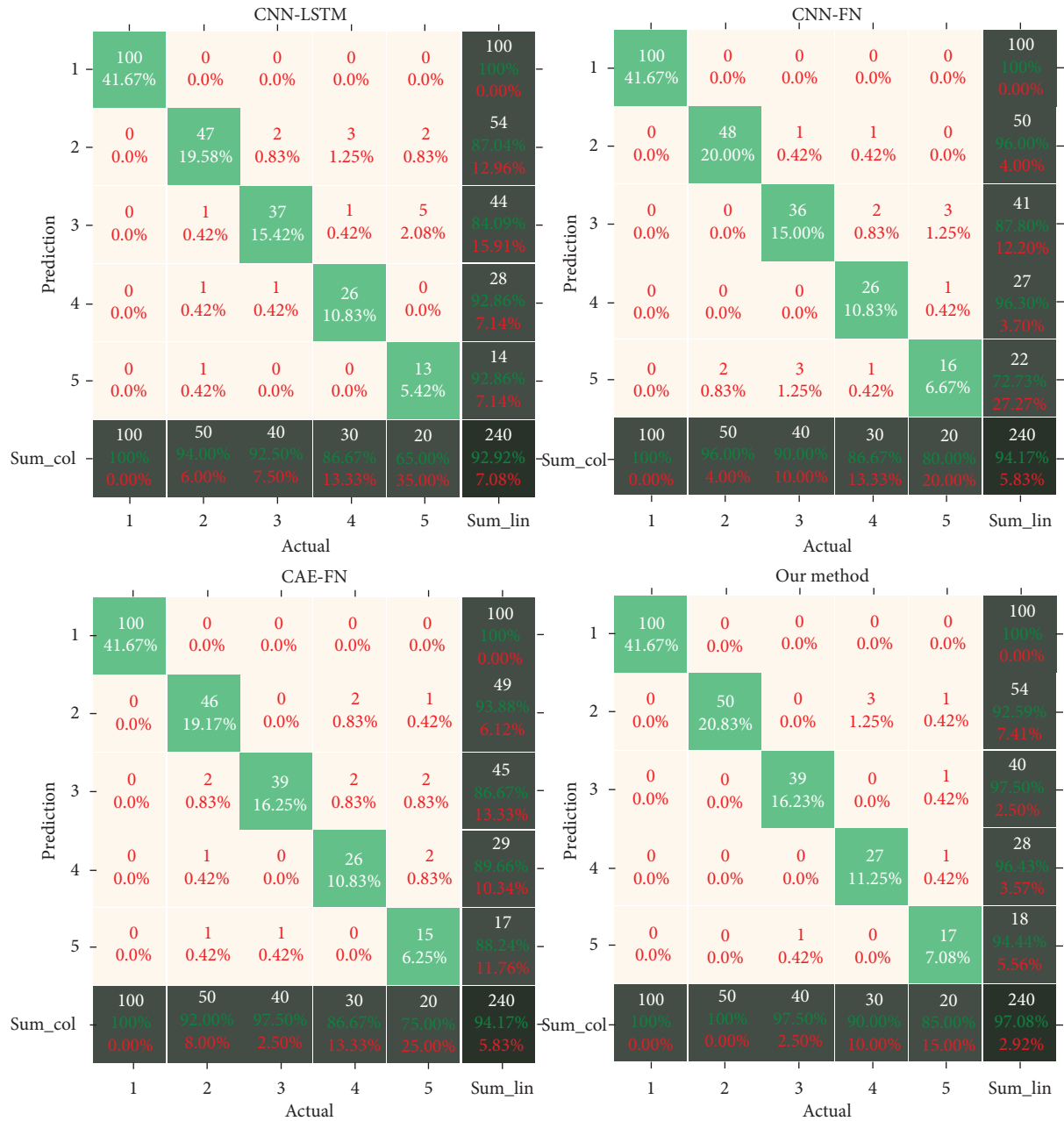


FIGURE 7: The confusion matrices.

can effectively suppress the influences of unbalancing. The similarity loss also has a relatively large impact on our network, due to that acquiring the invariance between the two modalities using this loss function before fusion can reduce the computation load while obtaining more comprehensive fault features. The recon loss has a relatively small impact on the network because the task loss can learn the trivial representations without this loss function. The impact of the triple loss is the smallest, since both the fusion and the similarity loss can achieve the clustering of this loss function, thereby weakening its contribution. Furthermore, it can be observed from the ablation experiment results of the

network structure that each module (pretrain CAE, MLP, LSTM, and CBAM) has a positive impact on the network. This indicates that the design of proposed network is reasonable and effective.

4.3.4. Generalization Performance Analysis. To validate the generalization capability of our model, in addition to the bearing data from the University of Paderborn, this paper used bearing-data from Xi'an Jiaotong University (XJTU-SY) and Case Western Reserve University (CWRU) for testing.

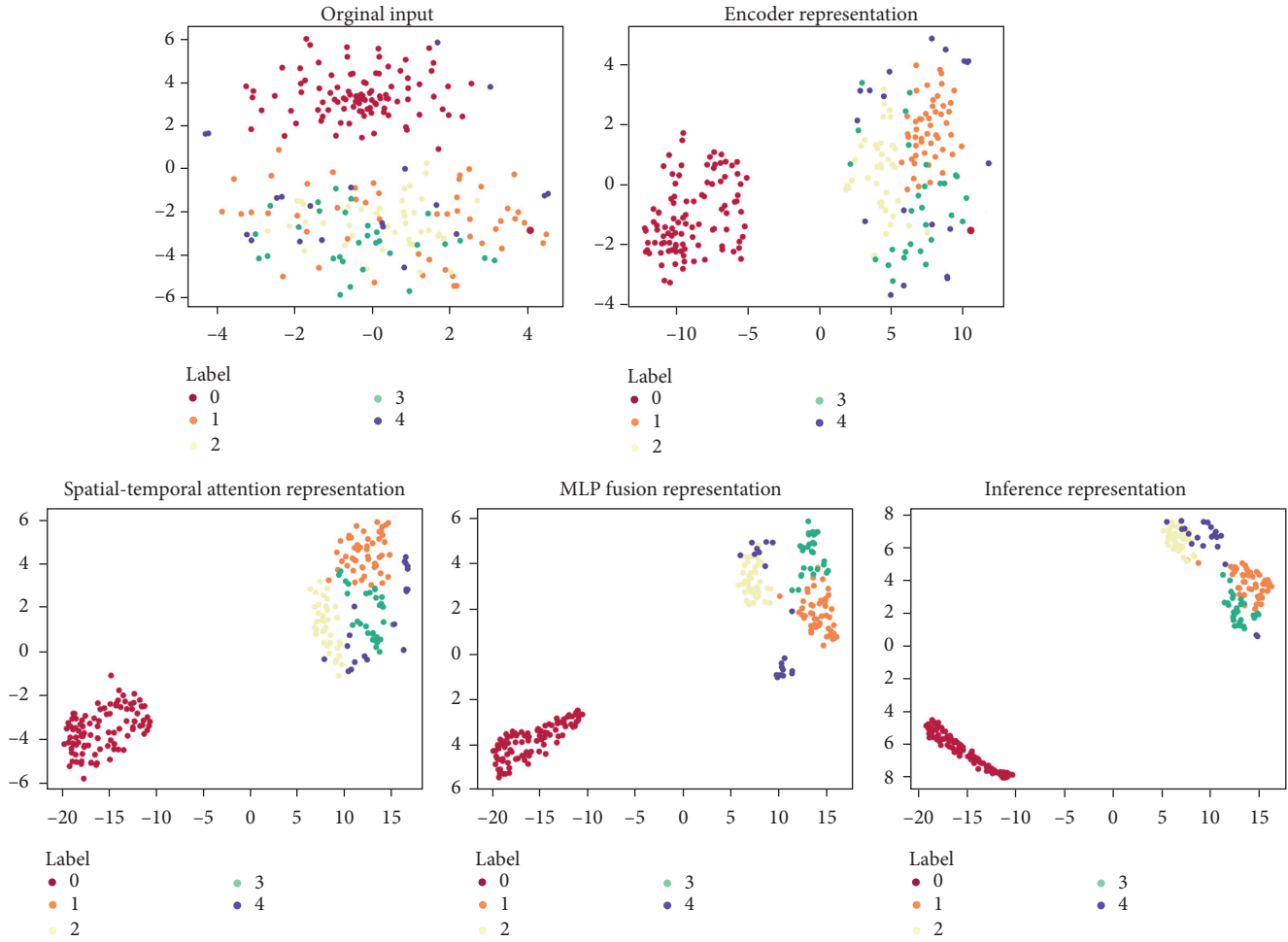


FIGURE 8: T-SNE visualization.

TABLE 6: Experiment of ablation studies.

Experiment type	Ablation factors	Accuracy (%)
Loss ablation	Without focal loss	93.33
	Without recon loss	95.42
	Without similarity loss	94.58
	Without triplet loss	96.67
	Without CBAM	95.83
Structure ablation	Without LSTM	94.58
	Without MLP	94.17
	Without pretrain CAE	95
—	None	97.08

The data from the University of Paderborn consist of vibration signals from two sensors in the target domain dataset of the experiment section. The data from XJTU-SY include vibration signals of five fault categories for the outer ring, inner ring, outer ring + inner ring, bearing cage, and inner ring + ball + cage. The data from CWRU contain vibration signals from five categories of ball fault (fault

diameter = 0.007 inches), ball-1 fault (0.014 inches), inner ring fault (0.007 inches), outer ring fault (0.007 inches), and normal condition. The test results are illustrated in Figure 9. It can be observed that our proposed model can still achieve good performance on the different datasets under a variety of unbalanced conditions, which has proved that our model has sound generalizing capability.

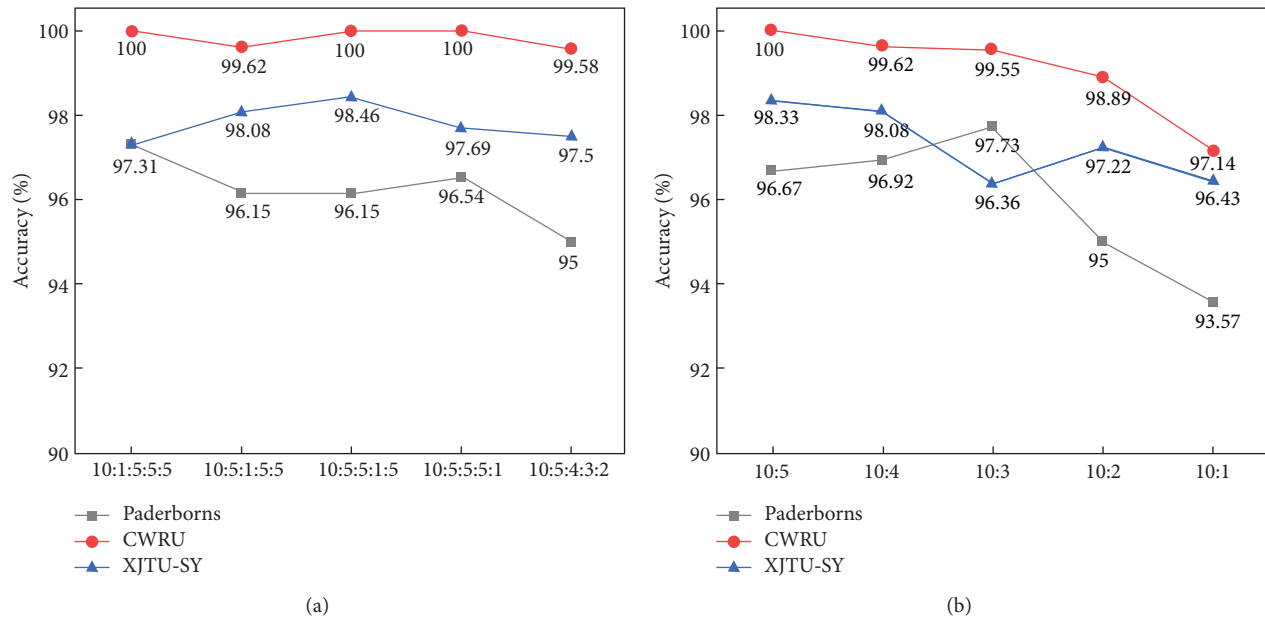


FIGURE 9: Generalization experiment of ITSA-FN.

5. Conclusions

Focusing on the issue of unbalanced bearing fault samples, we found that in addition to resampling methods and algorithm-level strategies, transfer learning and feature fusion methods can effectively improve the network's reasoning performance under unbalanced conditions. Moreover, we also discover that obtaining modal invariance through similarity measurement constraints can improve the efficiency of feature fusion, thereby further improving network performance. Therefore, this paper innovatively combines the algorithm-level strategy (focal loss), transfer learning, and the feature fusion method under similarity constraints, and presents an ISTA-FN model for vibration and current signals. The model consists of two sections of invariant spatial-temporal representation and constrained fusion representation. The invariant temporal-spatial attention representation section includes a pretrained CAE model (trained by balanced samples), CBAM, and LSTM. The fusion representation section includes an MLP. First, we extract independent features and modal-invariant features (interactive features) of the vibration and current signals using the invariant temporal-spatial attention representation section. Then, we use MLP to fuse and infer the features of the two modalities and guide the network's learning through a new loss function. Finally, this paper validates the effectiveness of our model through experiments and a visualization analysis. However, our approach has not touched on the situation of an extremely small sample population which we plan to investigate in future research using a zero-sample learning method.

Data Availability

The Paderborn bearing dataset was used in our study, which is publicly available via the following link: [https://mb.uni-](https://mb.uni-paderborn.de/kat/forschung/datacenter/bearing-datacenter/)

[paderborn.de/kat/forschung/datacenter/bearing-datacenter/](https://mb.uni-paderborn.de/kat/forschung/datacenter/bearing-datacenter/). The CWRU bearing dataset was used in our study, which is publicly available via the following link: <https://engineering.case.edu/bearingdatacenter>. The XJTU-SY bearing dataset was used in our study, which is publicly available via the following link: <https://biaowang.tech/xjtu-sy-bearing-datasets/>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Jianhua Liu and Haonan Yang conceived and designed the experiments; Haonan Yang proposed the method; Jing He performed the experiments; Zhenwen Sheng analyzed the data; and Shou Chen prepared the original draft.

Acknowledgments

This work was supported by the National Key R&D Program of China (2021YFF0501101), National Natural Science Foundation of China (61733004 and 62173137), Hunan Provincial Natural Science Foundation of China (2021JJ30217 and 2021JJ50001), and Project of Hunan Provincial Department of Education (19A137).

References

- [1] Y. Wang, L. Zeng, L. Wang, Y. Shao, Y. Zhang, and X. Ding, "An efficient incremental learning of bearing fault imbalanced data set via filter StyleGAN," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.
- [2] R. P. Monteiro, M. Cerrada, D. R. Cabrera, R. V. Sánchez, and C. J. A. Bastos-Filho, "Using a support vector machine based decision stage to improve the fault diagnosis on gearboxes,"

- Computational Intelligence and Neuroscience*, vol. 2019, pp. 1–13, Article ID 1383752, 2019.
- [3] W. Deng, J. Xu, H. Zhao, and Y. Song, “A novel gate resource allocation method using improved PSO-based QEA,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, pp. 1–9, 2020.
 - [4] W. Deng, J. Xu, X. Gao, and H. Zhao, “An enhanced MSIQDE algorithm with novel multiple strategies for global optimization problems,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, pp. 1–10, 2020.
 - [5] X. Cheng, D. Wang, C. Xu, and J. Li, “Sensor fault diagnosis method based on α -grey wolf optimization-support vector machine,” *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–15, 2021.
 - [6] S. Mao, M. Ye, and X. Li, “Rapid vehicle logo region detection based on information theory,” *Computers & Electrical Engineering*, vol. 39, no. 03, pp. 863–872, 2013.
 - [7] J. Liu, M. Zhang, H. Wang et al., “Sensor fault detection and diagnosis method for AHU using 1-D CNN and clustering analysis,” *Computational Intelligence and Neuroscience*, vol. 2019, pp. 1–20, 2019.
 - [8] W. Deng, S. Shang, X. Cai, H. Zhao, Y. Song, and J. Xu, “An improved differential evolution algorithm and its application in optimization problem,” *Soft Computing*, vol. 25, no. 7, pp. 5277–5298, 2021.
 - [9] C. Wang, C. Xin, and Z. Xu, “A novel deep metric learning model for imbalanced fault diagnosis and toward open-set classification,” *Knowledge-Based Systems*, vol. 220, Article ID 106925, 2021.
 - [10] Z. Wang, L. Zheng, W. Du et al., “A novel method for intelligent fault diagnosis of bearing based on capsule neural network,” *Complexity*, vol. 2019, pp. 1–17, Article ID 6943234, 2019.
 - [11] X. Li, W. Zhang, N.-X. Xu, and Q. Ding, “Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places,” *IEEE Transactions on Industrial Electronics*, vol. 67, no. 8, pp. 6785–6794, 2020.
 - [12] B. Zhao, X. Zhang, H. Li, and Z. Yang, “Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions,” *Knowledge-Based Systems*, vol. 199, Article ID 105971, 2020.
 - [13] C. Wang, H. Li, K. Zhang, S. Hu, and B. Sun, “Intelligent fault diagnosis of planetary gearbox based on adaptive normalized CNN under complex variable working conditions and data imbalance,” *Measurement*, vol. 180, Article ID 109565, 2021.
 - [14] T. Zhang, J. Chen, F. Li et al., “Intelligent fault diagnosis of machines with small & imbalanced data: a state-of-the-art review and possible extensions,” *ISA Transactions*, vol. 119, 2021.
 - [15] A. Duan, L. Guo, H. Gao, X. Wu, and X. Dong, “Deep focus parallel convolutional neural network for imbalanced classification of machinery fault diagnostics,” *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 11, pp. 8680–8689, 2020.
 - [16] R. Batuwita and V. Palade, “Efficient resampling methods for training support vector machines with imbalanced datasets,” in *Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, Barcelona, Spain, July 2010.
 - [17] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, no. 1, pp. 321–357, 2002.
 - [18] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, “Safe-level-SMOTE: safe-level-synthetic minority over-sampling TEchnique for handling the class imbalanced problem,” *Advances in Knowledge Discovery and Data Mining*, in *Proceedings of the 13th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD)*, pp. 475–482, Bangkok, Thailand, April 2009.
 - [19] H. Han, W.-Y. Wang, and B.-H. Mao, “Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning,” in *Proceedings of the 2005 international conference on Advances in Intelligent Computing*, pp. 878–887, Berlin, Germany, August 2005.
 - [20] D. Cieslak, N. Chawla, and A. Striegel, “Combating imbalance in network intrusion datasets,” in *Proceedings of the International Conference on Granular Computing*, pp. 732–737, Atlanta, GA, USA, May 2006.
 - [21] H. He, Y. Bai, E. Garcia, and L. Shuetao, “ADASYN: adaptive synthetic sampling approach for imbalanced learning,” in *Proceedings of the International Joint Conference on Neural Networks*, pp. 1322–1328, Hong Kong, China, June 2008.
 - [22] G. Douzas, F. Bacao, and F. Last, “Improving imbalanced learning through a heuristic oversampling method based on K-means and SMOTE,” *Information Sciences*, vol. 465, pp. 1–20, 2018.
 - [23] J. Wei, H. Huang, L. Yao, Y. Hu, Q. Fan, and D. Huang, “New imbalanced fault diagnosis framework based on Cluster-MWMOTE and MFO-optimized LS-SVM using limited and complex bearing data,” *Engineering Applications of Artificial Intelligence*, vol. 96, Article ID 103966, 2020.
 - [24] J. Wei, H. Huang, L. Yao, Y. Hu, Q. Fan, and D. Huang, “New imbalanced bearing fault diagnosis method based on Sample-characteristic Oversampling Technique (SCOTE) and multi-class LS-SVM,” *Applied Soft Computing*, vol. 101, Article ID 107043, 2021.
 - [25] Z. Wu, W. Lin, B. Fu, J. Guo, Y. Ji, and M. Pecht, “A local adaptive minority selection and oversampling method for class-imbalanced fault diagnostics in industrial systems,” *IEEE Transactions on Reliability*, vol. 69, no. 4, pp. 1195–1206, 2020.
 - [26] Z. Li, T. Zheng, Y. Wang, Z. Cao, Z. Guo, and H. Fu, “A novel method for imbalanced fault diagnosis of rotating machinery based on generative adversarial networks,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–17, 2021.
 - [27] G. Yang, Y. Zhong, L. Yang, H. Tao, J. Li, and R. Du, “fault diagnosis of harmonic drive with imbalanced data using generative adversarial network,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.
 - [28] Y.-r. Wang, G.-d. Sun, and Q. Jin, “Imbalanced sample fault diagnosis of rotating machinery using conditional variational auto-encoder generative adversarial network,” *Applied Soft Computing*, vol. 92, Article ID 106333, 2020.
 - [29] Z. Zhi-Hua Zhou and X Xu-Ying Liu, “Training cost-sensitive neural networks with methods addressing the class imbalance problem,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 63–77, 2006.
 - [30] J.-S. Lee, “AUC4.5: auc-based C4.5 decision tree algorithm for imbalanced data classification,” *IEEE Access*, vol. 7, Article ID 106034, 2019.
 - [31] W. Zong, G.-B. Huang, and Y. Chen, “Weighted extreme learning machine for imbalance learning,” *Neurocomputing*, vol. 101, pp. 229–242, 2013.
 - [32] P. Peng, W. Zhang, Y. Zhang, Y. Xu, H. Wang, and H. Zhang, “Cost sensitive active learning using bidirectional gated

- recurrent neural networks for imbalanced fault diagnosis,” *Neurocomputing*, vol. 407, pp. 232–245, 2020.
- [33] C. Zhang, K. C. Tan, H. Li, and G. S. Hong, “A cost-sensitive deep belief network for imbalanced classification,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 1, pp. 109–122, 2019.
- [34] J. Wang, P. Zhao, and S. C. H. Hoi, “Cost-sensitive online classification,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2425–2438, 2014.
- [35] P. Zhao, Y. Zhang, M. Wu, S. C. H. Hoi, M. Tan, and J. Huang, “Adaptive cost-sensitive online classification,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 2, pp. 214–228, 2019.
- [36] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal loss for dense object detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.
- [37] Q. He, Y. Pang, G. Jiang, and P. Xie, “A spatio-temporal multiscale neural network approach for wind turbine fault diagnosis with imbalanced SCADA data,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, pp. 6875–6884, 2021.
- [38] K. Bu, Y. He, X. Jing, and J. Han, “Adversarial transfer learning for deep learning based automatic modulation classification,” *IEEE Signal Processing Letters*, vol. 27, pp. 880–884, 2020.
- [39] Z. Ding and Y. Fu, “Deep transfer low-rank coding for cross-domain learning,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 6, pp. 1768–1779, 2019.
- [40] M. Fu, J. Liu, H. Zhang, and S. Lu, “Multisensor fusion for magnetic flux leakage defect characterization under information incompleteness,” *IEEE Transactions on Industrial Electronics*, vol. 68, no. 5, pp. 4382–4392, 2021.
- [41] C. Xing, H. Zhao, S. Shang et al., “An improved quantum-inspired cooperative Co-evolution algorithm with multi-strategy and its application,” *Expert Systems with Applications*, vol. 171, Article ID 114629, 2021.
- [42] H. Dong, K. Song, Y. He, J. Xu, Y. Yan, and Q. Meng, “PGANet: pyramid feature fusion and global context attention network for automated surface defect detection,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 12, pp. 7448–7458, 2020.
- [43] J. Yu and X. Zhou, “One-dimensional residual convolutional autoencoder based feature learning for gearbox fault diagnosis,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, pp. 6347–6358, 2020.
- [44] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: convolutional block Attention module,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19, Munich, Germany, September 2018.
- [45] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, “Spatial-temporal recurrent neural network for emotion recognition,” *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 839–847, 2019.
- [46] S. Mao, H. Wu, and M. Lu, “Multiple 3D marker localization and tracking system in image-guided radiotherapy,” *International Journal of Robotics and Automation*, vol. 32, no. 05, pp. 517–523, 2017.
- [47] W. Deng, S. Shang, X. Cai et al., “Quantum differential evolution with cooperative coevolution framework and hybrid mutation strategy for large scale optimization,” *Knowledge-Based Systems*, vol. 224, Article ID 107080, 2021.
- [48] Z. Zhao, Z. Bao, Z. Zhang et al., “Automatic assessment of depression from speech via a hierarchical attention transfer network and attention autoencoders,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 423–434, 2020.
- [49] J. Ma, H. Liu, C. Peng, and T. Qiu, “Unauthorized broadcasting identification: a deep lstm recurrent learning approach,” *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 5981–5983, 2020.
- [50] E. Principi, D. Rossetti, S. Squartini, and F. Piazza, “Unsupervised electric motor fault detection by using deep autoencoders,” *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 2, pp. 441–451, 2019.
- [51] W. Zelling, T. Grubinger, E. Lughofer, T. Natschläger, and S. Saminger-Platz, “Central moment discrepancy (CMD) for domain-invariant representation learning,” in *Proceedings of the 5th International Conference on Learning Representations (ICLR-2017)*, Toulon, France, April 2017.
- [52] F. Qu, J. Liu, X. Liu, and L. Jiang, “A multi-fault detection method with improved triplet loss based on hard sample mining,” *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 127–137, 2021.
- [53] L. Christian, K. J. Kuria, Z. Detmar, and W. Sextro, “Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification,” in *Proceedings of the European Conference of the Prognostics and Health Management Society*, pp. 5–8, Bilbao, Spain, July 2016.