CROSS-DOMAIN BEARING FAULT DIAGNOSIS BASED ON A TRIPLE ATTENTION MULTI-SCALE SPATIOTEMPORAL CONVOLUTIONAL NETWORK

1st Haojie Deng
School of Software
Taiyuan University of Technology
Taiyuan, China
2022511305@link.tyut.edu.cn

2nd Yanrong Hao
School of Software
Taiyuan University of Technology
Taiyuan, China
haoyanrong@tyut.edu.cn

3rd Xin Wen
School of Software
Taiyuan University of Technology
Taiyuan, China
xwen@tyut.edu.cn

4th Jing Bian
School of Software
Taiyuan University of Technology
Taiyuan, China
bianjing210@126.com

5th Jin Li
School of Software
Taiyuan University of Technology
Taiyuan, China
lijin0106@link.tyut.edu.cn

6th Rui Cao*
School of Software
Taiyuan University of Technology
Taiyuan, China
caorui@tyut.edu.cn

***Abstract*—Critical to mechanical systems, rolling bearings face severely degraded fault diagnosis performance under complex industrial conditions due to domain distribution shifts. To address the above issues, we propose a domain-adaptive deep transfer learning model for fault diagnosis, which incorporates a Triple-Attention Mechanism-Based Multi-Scale Spatiotemporal Network (TAM-MST) to enhance feature representation and facilitate effective knowledge transfer. Concurrently, we design a multi-layer maximum mean difference to improve distribution similarity between source and target domains, thereby strengthening the model's transfer learning capabilities. Our approach overcomes the low diagnostic accuracy of traditional models under varying operating conditions, endowing the model with adaptability to unknown operating scenarios. Finally, to validate the model's transfer learning capability, we compared it with several classical algorithms on the CWRU and PU datasets. Experimental results show diagnostic accuracies of 99.86% and 98.88% on the two datasets, respectively, providing a reliable research direction for bearing fault diagnosis in industrial equipment.**

***Index Terms*—Cross-condition fault diagnosis, Domain adaptation (DA), Attention mechanism,Industrial Anomaly Detection**

I. INTRODUCTION

In the context of the deep integration of industrial IoT and smart manufacturing, massive amounts of monitoring data have brought revolutionary opportunities to industrial equipment health monitoring technology [1]. As a core component of rotating machinery, the operating status of rolling bearings directly affects equipment reliability [2]. Statistics show that over 40% of mechanical failures are caused by bearing failures [3]. Simultaneously, bearing failures directly impact the performance of rotating machinery. Improper handling of these failures poses significant threats to personal safety and has major implications for social and economic development [4]. Accurate identification of bearing health is crucial for ensuring safe working conditions and efficient production. However, traditional methods struggle to meet the demands of modern industrial equipment, which is becoming increasingly complex and integrated [5].

Deep learning has demonstrated significant advantages in bearing fault diagnosis due to its end-to-end feature extraction capabilities [6], with wide applications in this domain. Representative techniques include convolutional neural networks (CNN) [7], sparse autoencoders (SAE) [8], and deep belief networks (DBN) [9]. Lee et al. [10] further developed a recurrent neural network incorporating attention mechanisms to enhance fault feature extraction accuracy. However, a critical limitation persists: vibration data collected under non-stationary operating conditions exhibit substantial distribution shifts [11]. Consequently, models trained on single-condition data fail to generalize effectively to new operational scenarios [12] , leading to substantial deterioration in generalization performance [13], [14] .

Experiments on two bearing fault datasets demonstrate outstanding performance and strong generalization under complex conditions, overcoming accuracy degradation of existing models to advance practical fault diagnosis and provide a reliable solution for industrial health monitoring.Tao et al. [15]enhanced cross-domain adaptability by aligning time-frequency

* is corresponding author. This work was supported by the National Natural Science Foundation of China (62206196) , the the Natural Science Foundation of Shanxi (202403011222002,202403021222057)and the Department of Science and Technology of Shanxi Province (202203021212216).

features using MMD. Subsequently, Yu et al. [16]addressed the single-kernel limitation of traditional MMD by integrating a multi-scale framework with MK-MMD. Complementing this kernel optimization work, the research focus was shifted to sample-level weighting by Qin et al. [17], who proposed Source-Weighted Maximum Mean Discrepancy (SWMMD). Building on the foundation of multi-kernel domains, the Multi-kernel Weighted Joint Domain Adaptive Network (MKWJ-DAN) was developed by Li et al. [18]. Beyond MMD-based methods, Zhao et al. [19] proposed a distinct self-supervised transfer learning framework (MTSTLF) that extracts robust domain-invariant features via self-supervision. Meanwhile, Fang et al. [20] used Local MMD (LMMD) to align global and sub-domain distributions, capturing fine-grained fault pattern differences. To highlight critical fault features, Jia et al. [21] further combined attention networks with Joint MMD (JMMD).

However, most existing domain-adaptive fault diagnosis methods employ simple feature extractors that limit the capture of deep and complex features. Their alignment strategies only focus on high-level features in the final layer, ignoring low-level fine-grained distribution alignment, which easily causes local optima, unstable training, and wrong diagnosis. To solve these problems, this paper proposes a novel transfer learning method called Triple-Attention based Multi-Scale Spatiotemporal feature learning framework (TAM-MST) for rolling bearing fault diagnosis. It can effectively learn deep fault features and achieve precise feature distribution alignment between source and target domains.

The main contributions of this paper are as follows:

- A triple hybrid attention spatiotemporal feature fusion module is proposed, which learns the useful information features of bearing faults and suppresses the useless ones via multi-scale attention, channel attention and temporal attention.
- To address cross-condition issues, we combine gradient descent with multi-layer MMD minimization to reduce feature distribution differences across domains. This approach overcomes the limitation of single-layer MMD, which focuses solely on a single level and struggles to fully adapt to the hierarchical variations in deep network features, thereby enhancing overall transfer learning capabilities.
- Through experimental analysis, we examined the impact of different attention modules on fault diagnosis performance. Tests conducted across various datasets and operating conditions demonstrated that this transfer diagnosis method achieves higher accuracy than alternative approaches.

II. RELATED WORKS

The working condition of rotating machinery in the actual industrial scene is complex, which leads to that the source and target domain does not follow the same distribution, and this difference becomes more and more obvious with the change of speed or load. In order to solve this problem, fault diagnosis methods based on DA have sprung up. Unlike supervised methods that require stringent labeling of data, these methods can greatly alleviate the tedious data labeling process [22].

DA is mainly categorized into distance-based and adversarial-based methods. The former maps cross-domain features into a shared space and quantifies their distribution divergence. For example, Qin et al. [23]proposed an adaptive middle class distribution alignment method, which realized domain confusion through dynamic middle class and introduced AdaSoftmax at the terminal to improve the separability of the model.An et al. [24]extracted periodic and local features through gaussian mixture transform layer, and used domain alignment term to map features to shared space. Yang et al. [25]extracted cross-domain data features by constructing a shared residual network, and collected target domain samples by using a clustering module, so as to realize the correction of joint distribution offset by associating source domain labels. Qian et al. [26]proposed a Relational Transfer Domain Generalized Network (RTDGN), which used two-stage DA to enhance the domain identification ability of the network, and designed the inverse entropy loss to improve the accuracy of the classifier. Yu et al. [27] realized local–global feature extraction by combining time–frequency features and global statistical matrix, and adopted Multi-Kernel MMD (MK-MMD) to improve the learning ability of the network.

The adversarial-based DA learns from the Generative Adversarial Network (GAN) idea, and plays a game between feature extractor and domain discriminator, thus reducing the negative impact brought by domain shift. For example, Gao et al. [28]utilized adversarial TL to obtain generalized feature, and proposed a joint distributed DA to further promote the learning of domain invariant feature. Chen et al. [29]used edge adversarial module to calculate the adversarial loss, and constructed an internal adversarial module to perform feature extraction, which enhanced the confusion performance of network. Li et al. [30]proposed a data-combined labeled domain expansion method, and improved the feature extraction ability of the model. Deng et al. [31] used progressive learning strategy to disperse unknown samples to improve the separability between classes. Fan et al. [32]proposed a deep mixed domain generalization network, which applied data expansion to feature space and introduced the adversarial perturbation for the computation of different distributions.

While existing research has advanced fault diagnosis, most methods only focus on global distribution consistency and ignore intra-class fine-grained differences, domain-specific features, and network generalization. Accordingly, multilayer pre-adaptation has attracted increasing attention. As shown in Fig. 1, global pre-adaptation aligns overall domain distributions, whereas multilayer pre-adaptation better captures fine-grained discrepancies and achieves more accurate class-level feature alignment between domains.

III. PROPOSED METHOD

As illustrated in Fig. 2, the workflow of this study comprises three main phases: data preprocessing, algorithm implementa-

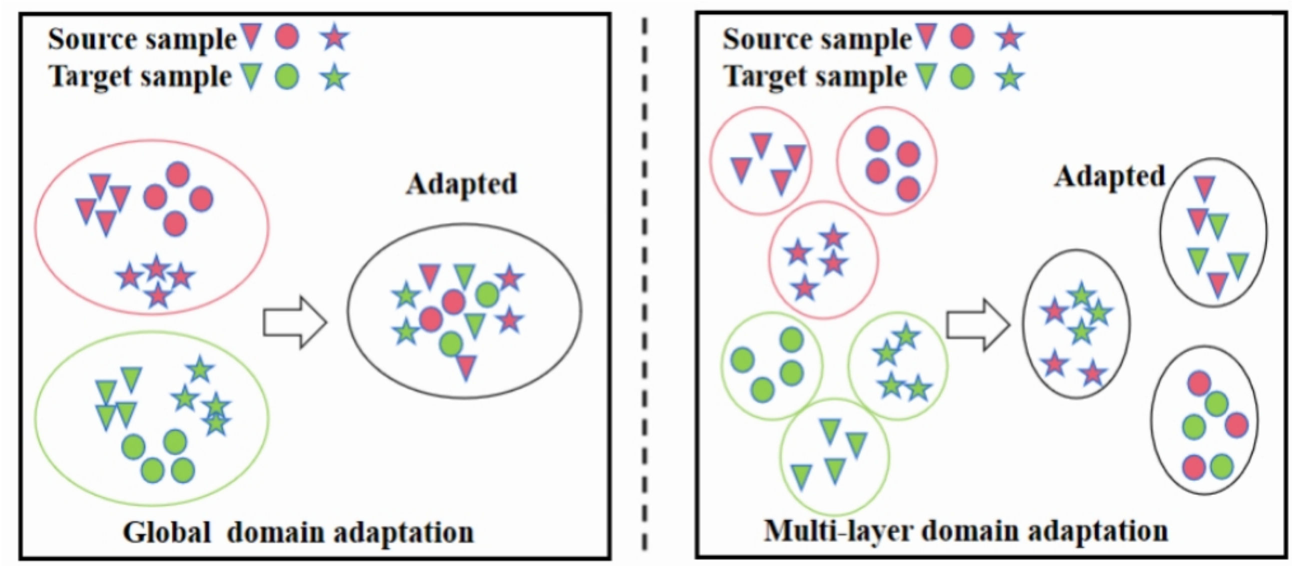

Fig. 1: Description of different DA methods.

tion, and result analysis. We propose a TAM-MST framework comprising three core components:a feature extraction module, a triple-attention mechanism that combines spatial, channel, and temporal attention for multidimensional feature extraction, and a hierarchical domain adaptation strategy designed to handle distribution shifts. These components work synergistically to achieve effective cross-domain fault diagnosis.

The triple attention mechanism enables precise multi-dimensional feature extraction for non-stationary vibration signals, overcoming traditional attention's cross-domain bearing diagnosis limitations: multi-scale attention uses 3/5/7 kernels to extract full-frequency features (avoiding weak fault omissions), channel attention adjusts weights to highlight valid features and suppress noise (yielding high-purity inputs), and temporal attention leverages bidirectional LSTM to capture long-term dependencies (breaking convolution's local constraints). This progressive pipeline (multi-scale coverage $\rightarrow$ channel denoising $\rightarrow$ temporal modeling) generates cross-domain invariant features—multi-dimensional, pure, temporally robust—to drive optimal diagnostic performance.

A. Data Preprocessing

To preserve the integrity of raw signals, data preprocessing employs a segmentation strategy using fixed-length non-overlapping sliding windows. Raw vibration signals are partitioned into consecutive segments of 1,024 data points, where the window length equals the sliding step size.

B. Triple Attention Module

The segmented data is a one-dimensional time series, and the input 1D vibration signal$X \in \mathbb{R}^{B\times L}$undergoes parallel multi-scale feature extraction through three convolutional kernels $\{k_1 = 3, k_2 = 5, k_3 = 7\}$.

$$F_i = \text{ReLU}\left(\text{Conv1D}\left(X; C, k_i, p = \left\lfloor \frac{k_i}{2} \right\rfloor\right)\right) \tag{1}$$

The feature maps $\mathbf{F} \in \mathbb{R}^{B\times C\times L}$ ($C = 16$) are generated.To avoid dimensionality explosion from naive multi-scale feature concatenation, a dynamic attention mechanism enables weighted fusion. Adaptive global average pooling extracts channel-wise statistics, which are processed through two FC layers to generate attention weights. This dynamically recalibrates feature contributions, enhancing sensitivity to subtle faults. where $\mathbf{W}_3$ and $\mathbf{W}_4$ are learnable parameters, σ denotes the ReLU activation function, and $\|\cdot\|$ represents the concatenation operation.

$$\mathbf{z}_i = \frac{1}{L}\sum_{t=1}^{L} \mathbf{F}_i[:, :, t], \quad \mathbf{z}_i \in \mathbb{R}^{B\times C} \tag{2}$$

$$\alpha_i = \frac{\exp\left(\mathbf{W}_2\sigma\left(\mathbf{W}_1\left[\mathbf{z}_1\|\mathbf{z}_2\|\mathbf{z}_3\right]\right)\right)}{\sum_{j=1}^{3}\exp\left(\mathbf{W}_2\sigma\left(\mathbf{W}_1\left[\mathbf{z}_1\|\mathbf{z}_2\|\mathbf{z}_3\right]\right)\right)} \tag{3}$$

$$\mathbf{F}_{\text{ms}} = \sum_{i=1}^{3} \alpha_i \cdot \mathbf{F}_i, \quad \mathbf{F}_{\text{ms}} \in \mathbb{R}^{B\times C\times L} \tag{4}$$

By dynamically regulating channel contributions, fault-related features are enhanced. Global average pooling is applied temporally to extract channel-wise statisticsz_{ch}, which are nonlinearly transformed into attention weightsβ. This allows the model to autonomously emphasize informative channels and suppress noisy ones. Finally, the weights are multiplied channel-wise with multi-scale features, enabling dual attention refinement and providing optimized representations for downstream modules.The formula is as

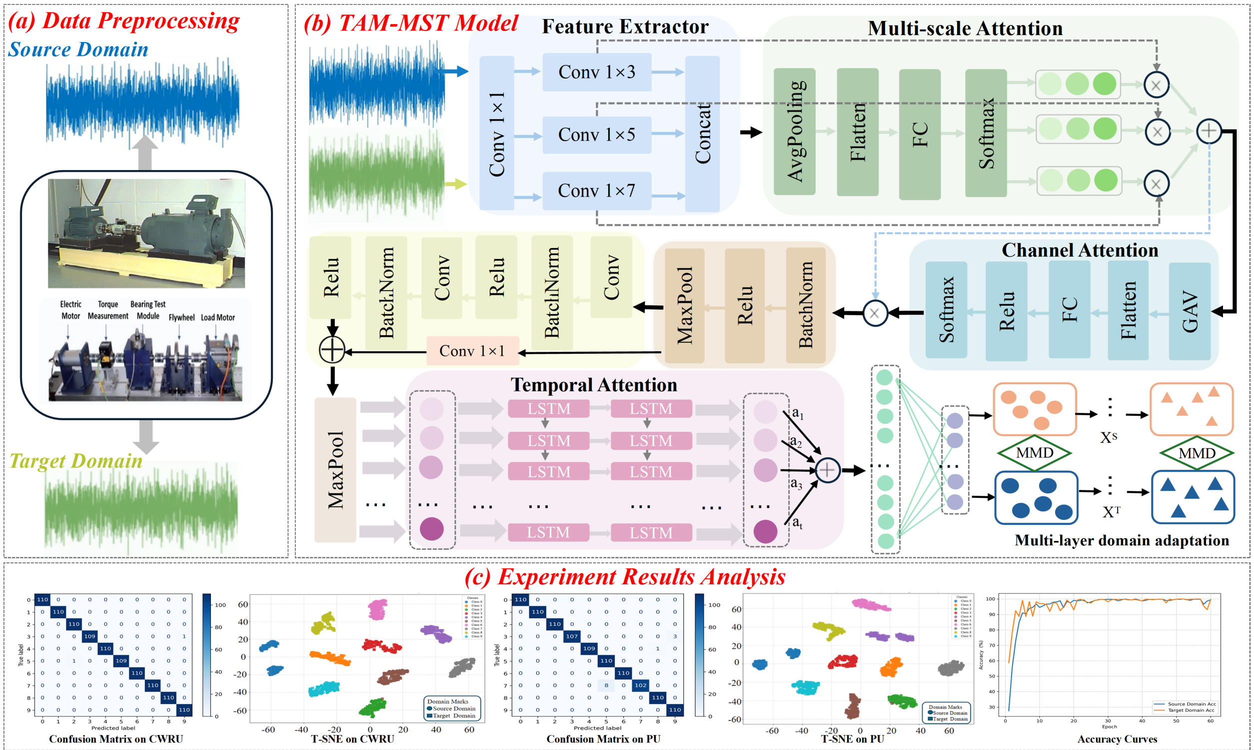

Fig. 2: The overall architecture of the proposed TAM-MST framework.

follows:

$$\mathbf{F}_{\text{mc}} = \mathbf{F}_{\text{ms}} \otimes \left(\text{Sigmoid}\left(\mathbf{W}_4 \cdot \text{ReLU}\left(\mathbf{W}_3 \cdot \mathbf{z}_{\text{ch}}\right)\right) \odot \mathbf{1}_L^T\right) \tag{5}$$

where $\mathbf{W}_3$ and $\mathbf{W}_4$ are learnable channel-wise weight matrices, $\mathbf{z}_{\text{ch}}$ denotes the result of global average pooling over the temporal dimension of $\mathbf{F}_{\text{ms}}$, $\otimes$ represents element-wise multiplication, $\odot$ denotes broadcasting multiplication to match the spatial-temporal dimension, and $\mathbf{1}_L^T$ is an all-ones vector of length $L$for dimension alignment. After being fused via channel attention, $\mathbf{F}_{\text{mc}}$ is fed into a residual convolutional block, which employs a dual-branch architecture consisting of a main path and shortcut connection for spatial feature extraction, effectively mitigating the gradient vanishing problem in deep networks. Following this, max-pooling operations (kernel size=4, stride=4) are applied to compress the feature sequence length from 1024 to 64.

To capture long-term temporal dependencies, the spatial features are reshaped into a 2D tensor and input into a bidirectional LSTM network. The LSTM outputs hidden states $\mathbf{H} \in \mathbb{R}^{B \times L' \times 2D}$ (where $D = 512$ denotes the hidden dimension) at each timestep, along with the final timestep hidden state $\mathbf{h}_n$ and cell state $\mathbf{c}_n$.

Attention scores are generated by measuring the similarity between the final hidden state and outputs at each timestep. Finally, weighted temporal features and spatial features are concatenated to form spatiotemporal features, with the corresponding formulations provided as follows:

$$\text{attscores} = \text{Softmax}\left(\frac{\mathbf{h}_n \mathbf{H}^T}{\sqrt{d}}\right) \tag{6}$$

$$\mathbf{F}_{\text{attn}} = \sum_{t=1}^{L'} \text{attnscores}[:, t] \odot \mathbf{H}[t, :, :] \tag{7}$$

C. Domain adaptation and classification

To align features across domains, we introduce a dynamic attenuation mechanism for joint loss construction. The discrepancy between source and target domains is measured with Maximum Mean Discrepancy (MMD). MMD is computed at multiple feature levels and adaptively weighted via attenuation coefficients to form the final domain loss. The MMD loss is calculated as:

$$\begin{aligned}\mathcal{L}_{\text{MMD}}^{(l)} = \frac{1}{n^2} \sum_{i,j=1}^{n} k(x_i^s, x_j^s) + \frac{1}{n^2} \sum_{i,j=1}^{n} k(x_i^t, x_j^t) \\ - \frac{2}{n^2} \sum_{i,j=1}^{n} k(x_i^s, x_j^t)\end{aligned} \tag{8}$$

where x_i^s and x_i^t denote the sample features of the source domain and the target domain, $k(\cdot, \cdot)$ represents the Gaussian kernel function.$l = 1$ corresponds to $\mathbf{F}_{\text{ms}}$ features, $l = 2$ corresponds to $\mathbf{F}_{\text{attn}}$ features, and $l = 3$ corresponds to $\mathbf{F}_{\text{final}}$ features. The total domain dynamic attenuation loss is given by:

$$\mathcal{L}_{\text{dom}} = \sum_{l=1}^{3} \left(\lambda_l \cdot \left(1 - \frac{t}{T}\right)\right) \cdot \mathcal{L}_{\text{MMD}}^{(l)} \tag{9}$$

The final extracted feature vector is projected by a fully connected layer into a 10-dimensional category space corresponding to the 10 bearing fault types. The classification loss employs the cross-entropy loss function, and the total model loss is the weighted sum of the classification loss and the domain loss.The formula is as follows:

$$\mathcal{L}_{\text{total}} = -\frac{1}{B}\sum_{i=1}^{B}\sum_{c=1}^{10}\mathbb{I}(y_i = c)\log\frac{\exp\left(y_{\text{ds}}^{(i,c)}\right)}{\sum_{k=1}^{10}\exp\left(y_{\text{ds}}^{(i,k)}\right)} + \mathcal{L}_{\text{dom}} \tag{10}$$

IV. EXPERIMENTS AND RESULTS

A. Datasets and Experimental Setup

(1)Datasets

Our model is evaluated on two benchmark datasets: the CWRU dataset [33] and the PU dataset [34]. Collected from the drive end, fan end, and base at sampling rates of 12 kHz and 48 kHz, the CWRU dataset is divided into four working conditions (A, B, C, D). It contains normal conditions as well as inner race, outer race, and rolling element faults, with three severity levels (7, 14, and 21 mils), and detailed information is provided in TableI. The PU dataset includes 64 kHz vibration signals from both healthy and faulty bearings (with artificial and natural damage), and is categorized into four working conditions (E, F, H, G). It covers normal conditions, inner race faults, outer race faults, and multiple fault conditions, and specific details are shown in TableII. Both datasets consist of ten classes: one for normal conditions and nine for different fault conditions.

TABLE I: Four Working Conditions of CWRU Dataset

Dataset setting	Motor load (hp)	Speed (rpm)	Sampling rate (kHz)	Fault type
A	0	1797	12	Normal, IR, OR, Ball
B	1	1772	12	Normal, IR, OR, Ball
C	2	1750	12	Normal, IR, OR, Ball
D	3	1730	12	Normal, IR, OR, Ball

TABLE II: Four Working Conditions of PU Dataset

Dataset setting	Load torque (Nm)	Speed (rpm)	Radial force (N)	Fault type
E	0.7	1500	1000	Normal, IR, OR, MF
F	0.7	900	1000	Normal, IR, OR, MF
G	0.1	1500	1000	Normal, IR, OR, MF
H	0.7	1500	400	Normal, IR, OR, MF

(2) Cross - domain Fault Diagnosis Tasks

The CWRU dataset encompasses six cross-domain fault diagnosis tasks(A→B, A→C, A→D, B→C, B→D, C→D). Here, labels before and after the arrow "→" denote source and target domain data, respectively. The PU dataset exhibits more pronounced inter-domain distribution shifts, for which we have correspondingly established cross-domain fault diagnosis tasks (E→F, E→H, E→G, F→H, F→G, H→G). All experiments were repeated five times to mitigate randomness effects.

(3) Parameter setup

The experiments were conducted on a workstation equipped with an Intel(R) Core(TM) i5-10500 CPU and an NVIDIA GeForce RTX 4090 GPU, utilizing the PyTorch 2.1.3 deep learning framework. The model was trained using the Adam optimizer with an initial learning rate of 0.001, a weight decay factor of 0.0009, a batch size of 32, and 100 epochs.

(4) Comparison Experiment

To validate the effectiveness of the proposed method, we conducted comparative experiments with several recent approaches, including MTSTLF [19] , LMMD [20] , JMMD [21] ,FreqMGCN [35], MTDA-IRP [36], and IND [37]. The comparative models above encompass not only typical JMDD and LMMD approaches but also pre-adaptive models incorporating attention mechanisms or networks integrating physical knowledge to enrich the model structure. Additionally, we compared graph-based diagnostic methods. In summary, we conducted comprehensive and meticulous testing of these methods' diagnostic performance under various operating conditions.

B. Experiment Results

We conducted six cross-domain comparison experiments on the CWRU and PU datasets. The results are shown in TableIII and TableIV. Our results achieve optimal accuracy on the CWRU and PU datasets. On the most challenging task A→D in the CWRU dataset, where other methods reached a maximum accuracy of 99.12%, our approach improved this by 0.88%, attaining 100% diagnostic accuracy. The PU dataset has more complex and subtle hidden fault information. Particularly in H→G scenarios with non-adjacent operating conditions, our method achieves 99.36% accuracy, surpassing the highest-performing alternative methods by 1.12%.

We performed comprehensive multi-dimensional verification on the CWRU and PU datasets to assess experimental reliability and robustness. A confusion matrix over ten categories (one normal, nine fault types) revealed cross-domain confusion points and validated our feature extraction mechanism. Additionally, t-SNE visualization showed that source and target domain features form compact, distinguishable clusters after domain adaptation.

Furthermore, we plotted a domain adaptation accuracy curve to dynamically track the alignment process between the source and target domains during model training. These multi-faceted verification results are presented in Fig. 2(c). Collectively, the experimental results on the two datasets fully validate the superior performance of the TAM-MST framework in cross-domain bearing fault diagnosis.

TABLE III: Comparison results on CWRU dataset (Accuracy %)

Method	A → B	A → C	A → D	B → C	B → D	C → D	AVG
JMMD	93.25	91.87	65.87	74.18	88.26	83.94	82.89
MTSTLF	93.11	91.30	90.89	91.17	91.20	89.83	91.25
LMMD	99.62	85.83	95.43	94.54	94.55	96.20	94.36
FreqMGCN	96.53	93.62	94.59	97.05	93.97	98.56	95.72
MTDA-IRP	99.79	99.36	98.27	97.13	97.58	98.45	98.43
IND	99.27	99.45	99.12	99.36	**99.64**	99.14	99.33
TAM-MST	**100.00**	**99.81**	**100.00**	**100.00**	99.54	**99.81**	**99.86**

TABLE IV: Comparison results on PU dataset (Accuracy %)

Method	E → F	E → H	E → G	F → H	F → G	H → G	AVG
JMMD	81.26	76.34	65.04	63.29	83.17	61.95	71.84
MTSTLF	95.74	91.58	94.87	80.26	68.94	92.83	87.37
LMMD	94.31	93.89	91.00	76.95	89.73	88.62	91.75
FreqMGCN	94.75	93.22	89.16	95.73	92.88	94.90	93.41
MTDA-IRP	92.14	92.20	96.34	98.67	98.78	97.45	95.93
IND	96.09	99.22	98.44	98.16	98.44	98.24	98.09
TAM-MST	**98.27**	**99.27**	**98.90**	**98.81**	**98.72**	**99.36**	**98.88**

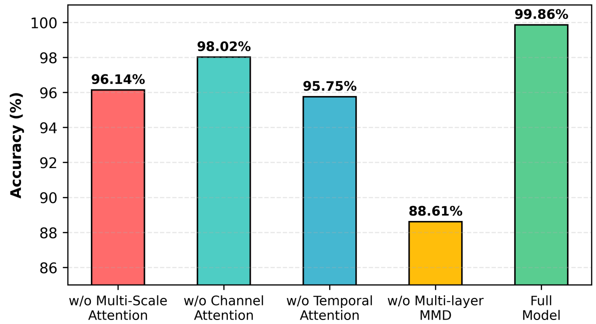

Fig. 3: Ablation study classification results on CWRU dataset.

C. Ablation Experiments

To validate the effectiveness of each module, we conduct ablation studies on four key components: multi-scale attention, channel attention, temporal attention, and multi-layer MMD. The corresponding ablation results are illustrated in Fig. 3.

The first three ablation experiments, in which the multi-scale attention module, channel attention module and temporal attention module are removed separately, verify the core functions of each component. The multi-scale attention mechanism is capable of capturing multi-band fault features and significantly enhancing the ability to extract diverse features. Removing this module causes the diagnostic accuracy to drop from 99.86% to 96.14%. The channel attention module serves to filter out noisy and redundant feature channels, while the temporal attention component is designed to model long-term temporal dependencies embedded in vibration signals.

The fourth experiment removes the multi-layer MMD module, and the accuracy plummets from 99.86% to 88.61%, with a decline of more than 10%. Multi-layer MMD is critical for cross-domain feature alignment. The triple attention mechanism exhibits synergy, as the full model outperforms all dual-attention variants. The complete model's optimal performance confirms the superiority of collaborative integration and the necessity of domain adaptation.

D. Hyperparameter Selection Analysis

Our approach combines classification loss with multi-layer MMD to achieve feature distribution alignment. The selection of the multi-layer MMD loss weighting coefficient α has a decisive impact on model performance. When the weighting coefficient α is set to 0.1, 0.05, and 0.2 respectively, the traditional fixed-weight strategy typically yields relatively low classification accuracy, ranging from 95.59% to 97.44%(see Fig.4, left). To address the limitations of fixed weights, this study innovatively proposes a dynamic weight adjustment strategy, which realizes linear weight decay throughout the training process through the formulation $\alpha = \lambda_l \times (1 - t/T)$. This strategy is observed to significantly outperform fixed-weight methods across all three parameter configurations, with the dynamic weight scheme employing $\lambda_l = 0.1$ achieving optimal performance at 99.86%(see Fig.4, right).

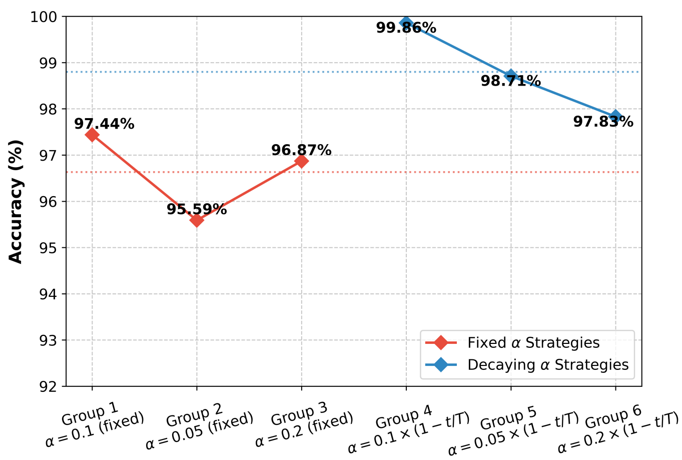

Fig. 4: The influence of MMD weights on model accuracy

V. CONCLUSION

To address the challenge of highly variable operating conditions in real-world fault diagnosis, we propose a novel deep transfer learning framework (TAM-MST). This framework combines a multi-scale spatio-temporal feature learning architecture with a triple attention mechanism. Multi-layer maximum mean discrepancy is introduced to effectively align feature distributions between source and target domains, thereby enhancing domain adaptation capabilities. Extensive experiments on two public bearing fault datasets demonstrate outstanding performance and strong generalization under complex conditions. The proposed design effectively overcomes accuracy degradation issues of existing models, advancing practical bearing fault diagnosis and providing a reliable solution for industrial equipment health monitoring and maintenance.

References

[1] J. He, Z. Cheng, and B. Guo, "Anomaly detection in telemetry data using a jointly optimal one-class support vector machine with dictionary learning," *Reliability Engineering & System Safety*, vol. 242, p. 109717, 2024.

[2] K. Huang, S. Tao, D. Wu, C. Yang, and W. Gui, "Robust condition identification against label noise in industrial processes based on trusted connection dictionary learning," *Reliability Engineering & System Safety*, vol. 247, p. 110133, 2024.

[3] E. Zio, "Prognostics and health management (phm): Where are we and where do we (need to) go in theory and practice," *Reliability Engineering & System Safety*, vol. 218, p. 108119, 2022.

[4] Y. Xiao, D. Liu, L. Cui, and H. Wang, "Heterogeneous graph representation-driven multiplex aggregation graph neural network for remaining useful life prediction of bearings," *Mechanical Systems and Signal Processing*, vol. 220, p. 111679, 2024.

[5] L. Cui, Z. Jiang, D. Liu, and D. Zhen, "A novel weighted sparse classification framework with extended discriminative dictionary for data-driven bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 222, p. 111777, 2025.

[6] Y. Ding, M. Jia, J. Zhuang, Y. Cao, X. Zhao, and C.-G. Lee, "Deep imbalanced domain adaptation for transfer learning fault diagnosis of bearings under multiple working conditions," *Reliability Engineering & System Safety*, vol. 230, p. 108890, 2023.

[7] T. Xie, X. Huang, and S.-K. Choi, "Intelligent mechanical fault diagnosis using multisensor fusion and convolution neural network," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 5, pp. 3213–3223, 2021.

[8] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, 2016.

[9] S. Xing, Y. Lei, S. Wang, and F. Jia, "Distribution-invariant deep belief network for intelligent fault diagnosis of machines under new working conditions," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 3, pp. 2617–2625, 2020.

[10] H. Lee, H. Jeong, G. Koo, J. Ban, and S. W. Kim, "Attention recurrent neural network-based severity estimation method for interturn short-circuit fault in permanent magnet synchronous machines," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 4, pp. 3445–3453, 2020.

[11] L. Cui, Q. Shen, Y. Xiao, D. Liu, and H. Wang, "Sparse graph structure fusion convolutional network for machinery remaining useful life prediction," *Reliability Engineering & System Safety*, vol. 254, p. 110592, 2025.

[12] L. Cui, Y. Xiao, D. Liu, and H. Han, "Digital twin-driven graph domain adaptation neural network for remaining useful life prediction of rolling bearing," *Reliability Engineering & System Safety*, vol. 245, p. 109991, 2024.

[13] S. Zheng, C. Wang, E. Zio, and J. Liu, "Fault detection in complex mechatronic systems by a hierarchical graph convolution attention network based on causal paths," *Reliability Engineering & System Safety*, vol. 243, p. 109872, 2024.

[14] S. Li, J. Ji, Y. Xu, X. Sun, K. Feng, B. Sun, Y. Wang, F. Gu, K. Zhang, and Q. Ni, "Ifd-mdcn: Multibranch denoising convolutional networks with improved flow direction strategy for intelligent fault diagnosis of rolling bearings under noisy conditions," *Reliability Engineering & System Safety*, vol. 237, p. 109387, 2023.

[15] H. Tao, J. Qiu, Y. Chen, V. Stojanovic, and L. Cheng, "Unsupervised cross-domain rolling bearing fault diagnosis based on time-frequency information fusion," *Journal of the Franklin Institute*, vol. 360, no. 2, pp. 1454–1477, 2023.

[16] S. Yu, L. Song, S. Pang, M. Wang, X. He, and P. Xie, "M-net: a novel unsupervised domain adaptation framework based on multi-kernel maximum mean discrepancy for fault diagnosis of rotating machinery," *Complex & Intelligent Systems*, vol. 10, no. 3, pp. 3259–3272, 2024.

[17] A. Qin, Q. Hu, Q. Zhang, and H. Mao, "A partial domain adaptation broad learning system for machinery fault diagnosis," *Measurement*, vol. 243, p. 116437, 2025.

[18] X. Li, H. Chen, S. Li, D. Wei, X. Zou, L. Si, and H. Shao, "Multi-kernel weighted joint domain adaptation network for cross-condition fault diagnosis of rolling bearings," *Reliability Engineering & System Safety*, vol. 261, p. 111109, 2025.

[19] L. Zhao, Y. He, D. Dai, X. Wang, H. Bai, and W. Huang, "A novel multi-task self-supervised transfer learning framework for cross-machine rolling bearing fault diagnosis," *Electronics*, vol. 13, no. 23, p. 4622, 2024.

[20] L. Fang, Y. Liu, X. Li, and J. Chang, "Intelligent fault diagnosis of rolling bearing based on deep transfer learning," in *2024 6th International Conference on Natural Language Processing (ICNLP)*. IEEE, 2024, pp. 753–757.

[21] S. Jia, Y. Deng, J. Lv, S. Du, and Z. Xie, "Joint distribution adaptation with diverse feature aggregation: A new transfer learning framework for bearing diagnosis across different machines," *Measurement*, vol. 187, p. 110332, 2022.

[22] Y. Chen, C. Yang, Y. Zhang, and Y. Li, "Deep conditional adaptation networks and label correlation transfer for unsupervised domain adaptation," *Pattern Recognition*, vol. 98, p. 107072, 2020.

[23] Q. Qian, J. Luo, and Y. Qin, "Adaptive intermediate class-wise distribution alignment: A universal domain adaptation and generalization method for machine fault diagnosis," *IEEE transactions on neural networks and learning systems*, vol. 36, no. 3, pp. 4296–4310, 2024.

[24] Y. An, K. Zhang, Y. Chai, Z. Zhu, and Q. Liu, "Gaussian mixture variational-based transformer domain adaptation fault diagnosis method and its application in bearing fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 1, pp. 615–625, 2023.

[25] B. Yang, Y. Lei, X. Li, and C. Roberts, "Deep targeted transfer learning along designable adaptation trajectory for fault diagnosis across different machines," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 9, pp. 9463–9473, 2022.

[26] Q. Qian, J. Zhou, and Y. Qin, "Relationship transfer domain generalization network for rotating machinery fault diagnosis under different working conditions," *IEEE transactions on industrial informatics*, vol. 19, no. 9, pp. 9898–9908, 2023.

[27] X. Yu, Y. Wang, Z. Liang, H. Shao, K. Yu, and W. Yu, "An adaptive domain adaptation method for rolling bearings' fault diagnosis fusing deep convolution and self-attention networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–14, 2023.

[28] T. Gao, J. Yang, and Q. Tang, "A multi-source domain information fusion network for rotating machinery fault diagnosis under variable operating conditions," *Information Fusion*, vol. 106, p. 102278, 2024.

[29] X. Chen, H. Shao, Y. Xiao, S. Yan, B. Cai, and B. Liu, "Collaborative fault diagnosis of rotating machinery via dual adversarial guided unsupervised multi-domain adaptation network," *Mechanical Systems and Signal Processing*, vol. 198, p. 110427, 2023.

[30] Q. Li, L. Chen, L. Kong, D. Wang, M. Xia, and C. Shen, "Cross-domain augmentation diagnosis: An adversarial domain-augmented generalization method for fault diagnosis under unseen working conditions," *Reliability Engineering & System Safety*, vol. 234, p. 109171, 2023.

[31] Y. Deng, J. Lv, D. Huang, and S. Du, "Combining the theoretical bound and deep adversarial network for machinery open-set diagnosis transfer," *Neurocomputing*, vol. 548, p. 126391, 2023.

[32] Z. Fan, Q. Xu, C. Jiang, and S. X. Ding, "Deep mixed domain generalization network for intelligent fault diagnosis under unseen conditions," *IEEE Transactions on Industrial Electronics*, vol. 71, no. 1, pp. 965–974, 2023.

[33] K. Loparo, "Case western reserve university (cwru) bearing data center," 2012.

[34] C. Lessmeier, J. K. Kimotho, D. Zimmer, 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 *PHM society European conference*, vol. 3, no. 1, 2016.

[35] J. Wang, S. Abdullah, C. Gao, A. Arifin, and S. S. K. Sing, "Freqmgcn-net: An iot-integrated multi-parallel graph convolutional network with frequency attention for motor bearing fault diagnosis," *Alexandria Engineering Journal*, vol. 128, pp. 175–185, 2025.

[36] G. Bai, T. Xing, W. Sun, H. Chi, Z. Zhong, Q. Sun, and L. Sun, "Unsupervised multiple-target domain adaptation for bearing fault diagnosis," *Engineering Applications of Artificial Intelligence*, vol. 154, p. 111063, 2025.

[37] H. He, C. Zhuang, and H. Xiong, "A novel incremental method for bearing fault diagnosis that continuously incorporates unknown fault types," *Mechanical Systems and Signal Processing*, vol. 216, p. 111524, 2024.