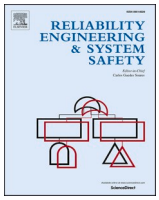
Reliability Engineering and System Safety 250 (2024) 110284

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress



Multi-stream domain adversarial prototype network for integrated smart roller TBM main bearing fault diagnosis across various low rotating speeds

Xingchen Fu, Keming Jiao, Jianfeng Tao *, Chengliang Liu

State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, PR China

ARTICLE INFO

Keywords:
Tunnel boring machine main bearing
Smart roller
Cross-domain fault diagnosis
Domain adversarial learning
Few-shot

ABSTRACT

The well-being of tunnel boring machines (TBM) heavily relies on the health status of main bearings, significantly impacting both safety and operational efficiency. Investigating the operation monitoring and diagnosis algorithms of TBM main bearings in practical scenarios with limited fault data and real-time variations in low-speed heavy-load operating conditions poses a research challenge. To solve these issues, a novel domain-adversarial prototype network diagnostic algorithm with a multi-stream fusion feature encoder (MDAPN) based on smart roller monitoring data is proposed. In contrast to existing external vibration monitoring schemes for bearings, monitoring data from smart rollers can mitigate complex external interference and greatly shorten the transmission path of fault sources. The algorithm encompasses a multi-stream fusion feature encoder, domain discriminator, and prototype network classifier. Specifically, the multi-stream fusion feature encoder adopts the two-stream convolutional network to deeply extract the axial-radial fused features of rollers, incorporating shallow statistical feature information. Domain-invariant features are generated based on domain adversarial learning strategies, and the prototype network classifier reduces dependence on target domain samples. An integrated smart roller main bearing fault simulation testbed was built, conducting 6 sets of cross-domain experiments. The proposed method reached an average accuracy of 98.41 % under 5-shot, surpassing the baseline method by 6.57 % at least. This validates that the MDAPN based on smart roller state exhibits excellent fault recognition performance for the main bearing under varying operational conditions with scarce available samples.

1. Introduction

As one of the most advanced tunnel construction equipment, tunnel boring machines (TBM) are widely employed in major engineering projects such as road network development and water diversion tunnels, owing to their advantages of cleanliness, efficiency, and safety [1]. As the core transmission component within the TBM, the main bearing plays a pivotal role in supporting the cutterhead rotation for rock excavation and bearing the load from associated components during the tunneling process. Owing to permanent operation in complex and harsh subterranean conditions [2], only 5 % of the actual service life of TBM main bearings has reached their designed lifespan actually [3]. In addition, the replacement of main bearings is costly, difficult, and seriously affects the project schedules [4], so the operational performance and efficiency of the entire drive system are directly impacted by the health condition of main bearings [5]. Researching the TBM main bearing fault diagnosis is crucial for reducing downtime, minimizing maintenance costs, and enhancing excavation efficiency.

Conventional bearing sensing schemes typically involve placing sensor units on the static bearing housing or seat ring, correlating the sensor output signals with structural conditions [6]. The sensing signals from such wired external monitoring means are greatly influenced by factors such as path loss and strong background noise [7]. However, the TBM main bearing is large and operates in harsh conditions, making it apparent that external vibration monitoring means are difficult to detect fault source information effectively. Despite extensive research on state monitoring of low-speed heavy-duty slewing bearings [8,9], they are difficult to apply to industrial sites. Research indicates [10] that smart bearings integrated with wireless embedded sensors can achieve self-diagnostic functions, and directly measure state parameters such as vibration, load, temperature, and speed in the core working area of bearings, while shielding against numerous external interference signals effectively. At the same time, self-powered technology also provides energy security for the practical application of such bearings. For

* Corresponding author.
E-mail address: jftao@sjtu.edu.cn (J. Tao).

https://doi.org/10.1016/j.ress.2024.110284
Received 28 December 2023; Received in revised form 28 May 2024; Accepted 11 June 2024
Available online 15 June 2024

example, wireless power transmission technology based on inductive coupling and electromagnetic radiation [11], as well as energy harvesting technology based on piezoelectric [12], electromagnetic [13], and triboelectricity [14] effects to collect energy from motion. Smart bearings developed by well-known bearing companies such as SKF, NTN, and Schaeffler find extensive applications across various industries, including wind power, railways, and automobiles [15,16]. To perceive the state information of the working area of large bearings and maximize the utilization of the limited sensing range of sensors, SKF first proposed the magic roller system [17], which monitors the large bearing state by collecting roller load data. In recent years SKF has applied the roller sensing system to collect load, roller speed, and temperature data of main bearings of wind turbines, for monitoring the contact stress, thermal effects, and roller deflection of main bearings [18]. In addition, well-known bearing manufacturing companies such as ThyssenKrupp [19] and JTEKT [20] have also developed sensing rollers to measure the load of large bearings. It is evident that roller monitoring signals are capable of effectively capturing state characteristics of the large bearing and can be used as a reference for monitoring the TBM main bearing state, providing the data foundation for fault diagnosis.

For low-speed heavy-duty large bearings, the analysis of fatigue life typically involves the examination of load distribution and friction torque [21], while vibration analysis is employed to identify fault modes [22]. In comparison to high-speed bearings, TBM main bearings are affected by low-speed and heavy-duty working conditions, and their vibration signals are affected by the long period between defect-related pulses, containing less fault-related information during the same period [23]. In addition, although roller state monitoring signals have a great shielding effect on a large number of external interference signals, they introduce roller rotation information, greatly increasing the difficulty of extracting fault features. Caesarendra [9] systematically summarized common vibration feature extraction methods and their applications in low-speed slewing bearing health management, including time-domain statistical characteristics [24]: pulse factor, kurtosis, etc.; frequency-domain features [25]: energy spectrum, envelope spectrum, etc.; time-frequency features [26]: wavelet decomposition, intrinsic time-scale decomposition, etc.; and nonlinear features [27] like the maximum Lyapunov exponent. Intelligent fault diagnosis techniques have advanced in the last several years due to the rise of deep learning. Various deep learning methods, including residual networks [28], autoencoders [29], convolutional neural networks (CNN) [30], long short-term memory [31], and Transformers [32], have been proposed. These approaches exploit multi-layer network architectures to end-to-end mine effective fault information, thereby mitigating the dependence on expert knowledge. Tang et al. [33] introduced fault identification methods in low-speed bearings amidst significant background noise, employing robust principal component analysis and multi-kernel deep neural networks. Fu et al. [34] introduced a two-stream CNN designed to diagnose low-speed heavy-load main bearing faults in TBM operating under strong noise conditions. Although the aforementioned research has shown how deep learning can be when applied to diagnosing low-speed bearing faults, there still exists a gap in research on roller state-based large-scale bearing diagnosis algorithms. Further analysis of the correspondence between roller space state characteristics and deep learning network structure is needed. Research into methods for separating roller motion information is also required to enhance the representation capability of bearing states. In practical engineering applications, the real-time variations in operating conditions, such as the speed and load of TBM main bearings, result in a noticeable decrease in diagnostic accuracy when applying the same trained model to different conditions. Moreover, data volume under various working scenarios is hard to guarantee.

Changes in working conditions cause domain shifts during the training and testing phases. Existing solutions to cross-domain fault diagnosis issues mainly include two solutions: domain generalization and domain adaptation [35]. Domain generalization achieves cross-domain tasks by learning diagnostic knowledge from multi-source domains and applying it to unseen target domains [36]. A typical example is using the federated domain generalization method to build multi-client, scalable joint diagnosis models and achieve data decentralization [37,38]. The fundamental idea of domain adaption techniques is to transfer the source domain knowledge to the target domain, achieving cross-domain adaptation of the model to different domains [39]. In comparison, domain generalization methods require a large amount of data and complex model calculations. Domain adaptation is more suitable for variable working conditions scenarios where available samples for TBM main bearings are scarce. Domain adaptation techniques broadly include two categories. The first one involves methods based on domain distance metrics, which employ loss functions such as Kullback-Leibler divergence, maximum mean difference (MMD), etc., to measure the differences between samples from two different domains in the feature space, with the goal of minimizing these differences to achieve domain adaptation tasks. The second one involves methods based on domain adversarial learning, which generate features with domain invariant characteristics by forming adversarial relationships between feature encoders and domain discriminators. This ensures that features extracted for diagnostic tasks in different domains exhibit good classification performance. Zhao et al. [40] proposed a deep multi-scale adversarial network to realize domain alignment. Its shared feature generator contains two novel multi-scale modules and attention mechanisms. Wu et al. [41] introduced adversarial learning strategies and the MMD loss function separately into the feature and prediction label spaces, achieving domain adaptation across varying speeds and loads. The above research proves that compared with simply relying on distance metrics, domain adversarial learning can more accurately capture shared features of variable operating conditions, improving the robustness and versatility of cross-domain models. In recent years, a meta-learning method commonly used for few-shot classification tasks, has gained attention due to its ability to quickly and accurately adapt to unseen tasks with few available samples [42]. Many scholars have applied metric-based meta-learning paradigms in cross-domain fault diagnosis [43], such as prototype networks and relation networks. Compared with common domain adaptation methods, meta-learning has significant advantages in addressing the issue of scarcity samples under variable operating conditions. Feng et al. [44] introduced a meta-learning method that leverages domain adversarial similarity, improving generalization performances of the algorithm for unseen tasks. Liu et al. [45] introduced a metadata-based residual network combined with a weight allocation method for bearing scenarios under class-imbalanced variable operating conditions. From the survey of the above literature, it is evident that domain adversarial learning-based deep learning methods achieved significant advancements in cross-condition high-speed small bearing fault detection. However, they tend to focus solely on deep fault information while ignoring shallow fault information, which may lose crucial fault information when facing various low-speed working conditions of large main bearings. Additionally, the rock-breaking impact during tunneling operations and momentary changes in cutterhead speed result in frequent variations in operating conditions, and the time for stable data acquisition under new working conditions is limited. Therefore, the diagnostic model is required to achieve rapid adaptation under the condition of scarce samples in the target domain.

This study developed a smart roller monitoring system that integrates vibration sensing and wireless communication functions, aiming to obtain the status information of main bearings at the TBM working site. On this basis, a novel multi-stream fusion feature encoder-based domain adversarial prototype network diagnostic algorithm (MDAPN) is proposed, which is suitable for solving the domain shift issue that arises during different working condition data training of the TBM main bearing with integrated smart rollers. The proposed MDAPN method comprises a multi-stream fusion feature encoder, a domain discriminator, and a prototype network classifier. Firstly, the multi-

stream fusion feature encoder is constructed based on roller state signals. It separately extracts radial vibration features, axial vibration features, and time-domain statistical information from three-axis vibration signals of the rollers. These features are then fused at the fully-connected layer to boost the network generalization performance. Secondly, an adversarial relationship is formed between the feature extractor and domain discriminator. Through maximizing the domain discriminator loss value and minimizing the classifier loss value, the feature encoder can generate features that are invariant across domains, eliminating domain shift impact on the classifier. Finally, a prototype network is employed as the classifier, following the meta-learning paradigm to recognize faults with extremely limited target domain data. The efficiency of MDAPN is demonstrated through numerous experiments under various operating conditions of main bearings. The primary breakthroughs and contributions of this article can be summarized as follows:

(1) This study developed a smart roller monitoring system that integrates vibration sensing and wireless communication to perceive the state information of the TBM main bearing. By utilizing a two-stream CNN to merge deep features of roller axial-radial direction and introducing shallow time-domain statistical features of three-axis vibration signals, the domain-generalization features extracted by the feature encoder are applicable to diagnostic issues associated with low-speed, heavy-load large bearings across different working conditions. Additionally, the effectiveness of the multi-stream feature encoder and prototype network classifier was further validated through ablation experiments.

(2) Utilizing domain adversarial learning and meta-learning methods, the proposed MDAPN model forms an adversarial relationship between its feature encoder and domain discriminator. This addresses the domain shift issue affecting classification accuracy after changes in operating conditions. The model incorporates a prototype network classifier to accomplish cross-domain fault recognition tasks of TBM main bearings in real industrial scenarios, especially under class imbalance and few shot conditions. A series of few-shot cross-domain experiments with variable speeds were carried out to verify the superior performance of the proposed method.

The subsequent sections of this article are structured as follows: Section 2 provides an introduction to the smart roller system and experimental testbed. Section 3 elucidates basic principles such as adversarial domain adaptation and prototype networks. Section 4 provides a detailed overview of the principles and implementation process of the MDAPN method. Section 5 conducts variable speed experiments and analyzes the results. Section 6 provides a conclusion.

2. Roller in-situ monitoring for TBM main bearing

Fig. 1(a) shows the typical TBM structure diagram, where main bearings play a role in transmitting power from the driving motor and supporting the rotation of the cutterhead for rock excavation. The commonly used low-speed heavy-load large main bearings are depicted in Fig. 1(b). This triple-row cylindrical roller combination structure can withstand high torque and complex loads from associated components. In specific working scenarios such as tunneling (TBM main bearings), pitch and yaw (wind turbine bearings), the rated speed operating range of these bearings is usually 0.5 to 15 rpm. However, when the main bearing is only used as a transmission system or in non-operational testing scenarios, the operating speed can reach 40 rpm [46,9]. Due to the large structure, the compact and relatively enclosured internal space of main bearings, external sensing schemes face difficulties in monitoring its structural damages. Fig. 2 illustrates the structure of the smart roller developed in this study, integrating functions such as state perception, wireless communication, and power management. The designed smart roller structure exhibits superior contact stress distribution compared to solid rollers, effectively preventing bending fatigue fracture of the rolling elements under low-speed heavy-load conditions. Fig. 3 presents the state monitoring system of the smart roller, comprising a miniature low-power controller unit, a triaxial accelerometer (sampling rate: 3200 Hz, transmission bandwidth: 153.6 kbps), a gyroscope (sampling frequency: 480 Hz, transmission bandwidth: 153.6 kbps), a WiFi module, and a energy management unit. This monitoring circuit sends roller state data containing internal fault source information of the main bearing to the router through a semi-enclosed end cover, and transmits it to the upper computer.

To verify if the roller state-based intelligent algorithm is feasible for cross-speed fault diagnosis in the TBM main bearing. This study developed a large-scale main bearing testbed integrated with smart rollers as shown in Fig. 4, which is used to monitor the roller state signals of different fault modes of large bearings under low speed, heavy load, and various operating conditions. The drive system of the testbed consists of a motor, reducer, and coupling. The loading system consists of a CP-180 hydraulic loading hand pump and an FPY-5 jack installed on top of the auxiliary bearing. The spindle is driven by the motor through the coupling, and the speed is controlled by the reducer. It withstands the load applied by the auxiliary bearing, allowing the test bearing to simulate low speeds ranging from 5 to 50 rpm and withstand radial loads of up to 100 MPa. The test bearing employs eight rollers with a diameter R_r of 50 mm, and the inner ring raceway diameter R_i and outer ring raceway diameter R_o are 122 and 222 mm, respectively. The design life of the testbed is determined by the auxiliary bearing, model GB/T-276 6224. After calculation, its design life is 3723.875×10^6 r. The number of smart rollers determines the monitoring efficiency of the bearing condition. When the number is n, assuming that the state of the bearing remains unchanged during the process of the rollers traversing $1/n$ revolutions along the outer race in pure rolling. In other words, the maximum distance at which the smart roller can detect bearing component damage pulses must be within the range of $1/n$ revolutions. After testing, it was found that in this experiment, when the number of smart rollers is 1, it is capable of monitoring the entire raceway of the test bearing of this testbed to detect any occurrence of damage.

Ultimately, the monitoring data of the smart roller in the test bearing are collected to analyze the status types of large bearings. In this

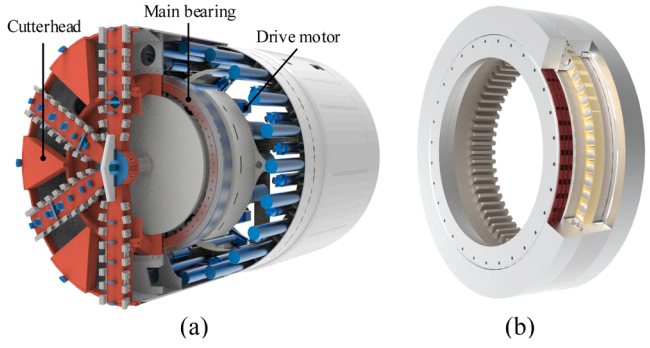


Fig. 1. Structure diagram: (a) TBM, (b) The triple-row cylindrical roller bearing.

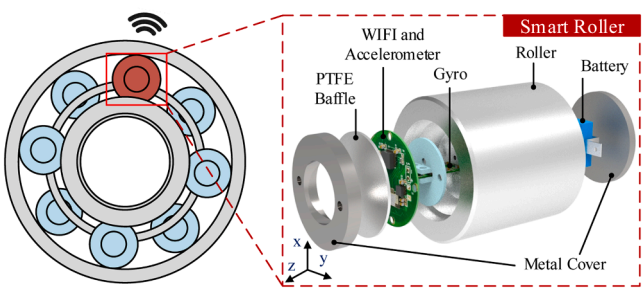


Fig. 2. Schematic diagram of smart roller structure.

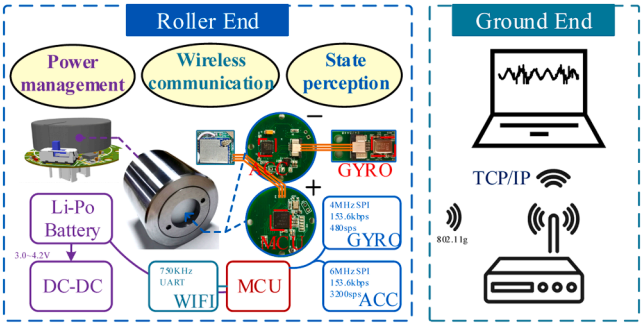


Fig. 3. Smart roller condition monitoring system.

experiment, crack faults were simulated by processing defects with 2 mm width and 1 mm depth on the bearing. Fault locations are set on the inner ring, outer ring, and rollers, respectively. It should be noted that roller faults are not set on smart rollers, and their occurrence location and quantity do not change the frequency characteristics of the faults. The hydraulic loading device is set to a 20 MPa load, and the three-axis vibration signals of rollers corresponding to normal bearings (NB), inner ring faults (IF), outer ring faults (OF), and roller faults (RF) are collected at speeds of 30 rpm, 40 rpm, and 50 rpm. The time-domain waveform of the roller vibration data for four bearing damage types at 30 rpm are depicted in Fig. 5.

3. Preliminaries

3.1. Domain adaptation

During the excavation process, alterations in geological conditions across various sections result in fluctuations in the operational parameters of the TBM main bearing. Typically, under different rotation speed conditions, the data distribution of corresponding domains differs, giving rise to domain shift phenomena. This implies that the deep learning model trained exclusively on data from a specific working condition may not be suitable for application in other conditions. Domain adaptation seeks to acquire features that are consistent across various operating conditions, enabling the trained classifier to be robust to domain shifts [47]. In order to facilitate method description, we define domain $D = \{(X, Y)\}$ as a collection of sample datasets X and corresponding labels Y from the same operating condition. $D_S = \left\{\left(x_i^S, y_i^S\right)\right\}_{i=1}^{N_s}$ represent N_s labeled source domain samples, and $D_T = \left\{\left(x_i^T\right)\right\}_{i=1}^{N_t}$ represent N_t unlabeled target domain samples. When a change occurs in operating conditions, the distributions of corresponding domains differ, i.e., $F_{dis}(D_S) \neq F_{dis}(D_T)$, where $F_{dis}(\cdot)$ represents the distribution function. Domain adaptation, as depicted in Fig. 6, is the process of predicting sample labels for target domain by extracting domain-invariant features, assuming that the label sets match for the source and target domains. Among them, $F_{invar} \in (FE(D_S) \cap FE(D_T))$, where $FE(\cdot)$ represents the feature encoder.

3.2. Domain-adversarial learning

As the common cross-domain adaptive method [48], the network architecture of domain adversarial method includes a feature encoder FE, a class discriminator M_c, and a domain discriminator M_d. The key to forming an adversarial relationship lies in: on the one hand, the feature encoder FE strives to generate features that are as unaffected by domain shift as possible to deceive the domain discriminator M_d; on the other hand, M_d endeavors to distinguish features based on their originating domain as accurately as possible, resisting deception by the feature encoder FE. The goal of training is roughly stated as: while ensuring the minimum loss term L_c of the class discriminator M_c, also minimize the loss term L_d of the domain discriminator M_d, which is inherently contradictory to the requirement of feature extraction. Therefore, in the process of backpropagation of training parameters, taking the partial derivative of L_c updates the feature encoder parameters θ_e and class discriminator parameters θ_c in the direction of gradient descent, while taking the partial derivative of L_d achieves gradient reversal by multiplying "−1" and updates the domain discriminator parameters θ_d in the direction of gradient ascent. Based on the above adversarial ideas, define the total loss of network L_{total} as shown in the following Eq. (1):

$$L_{total}(\theta_e, \theta_c, \theta_d) = \frac{1}{n_s} \sum_{x_i \in D_s} L_c(M_c(FE(x_i)), y_i) - \frac{1}{n_s + n_t} \sum_{x_i \in D_s \cup D_T} L_d(M_d(FE(x_i)), d_i) \tag{1}$$

where θ_e, θ_c, θ_d are parameters corresponding to FE, M_c, and M_d. Designate n_s as the batch sample quantity from the source domain and n_t from the target domain. The variable d_i represent the domain label.

As depicted in Fig. 7, the concept of adversarial learning guides the update of parameters in L_{total}. The feature encoder parameter θ_e and the class discriminator θ_c are optimized to minimize the loss function L_c obtained by the class discriminator M_c on the source domain data, namely minimizing L_{total}, thereby improving the distinguishability of features; Meanwhile, update the domain discriminator parameter θ_d separately. This update minimizes the domain discriminator loss function L_d on different domain data, i.e. maximizing L_{total}, thereby improving the domain discrimination ability to distinguish feature sources. In adversarial learning, this enhances the transferability of features extracted by the feature encoder. The updated parameter values are denoted as $\hat{\theta}_e$, $\hat{\theta}_c$, and $\hat{\theta}_d$. The extraction of features that are invariant across domains is ultimately achieved by means of the above

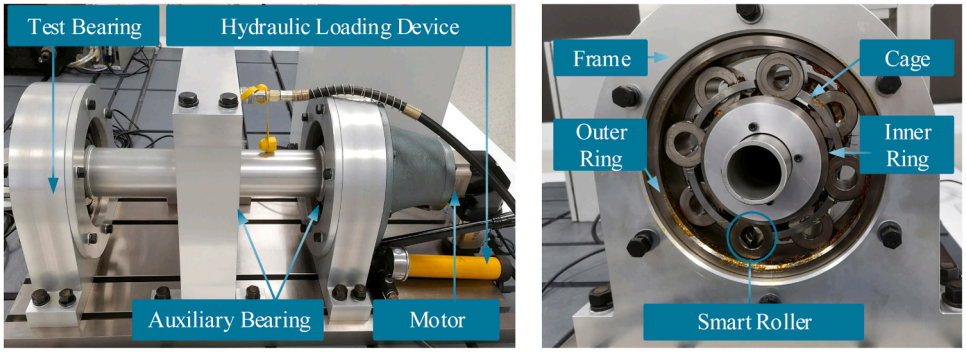


Fig. 4. Large-scale main bearing testbed with integrated smart rollers: (a) Overall perspective, (b) Test bearing perspective.

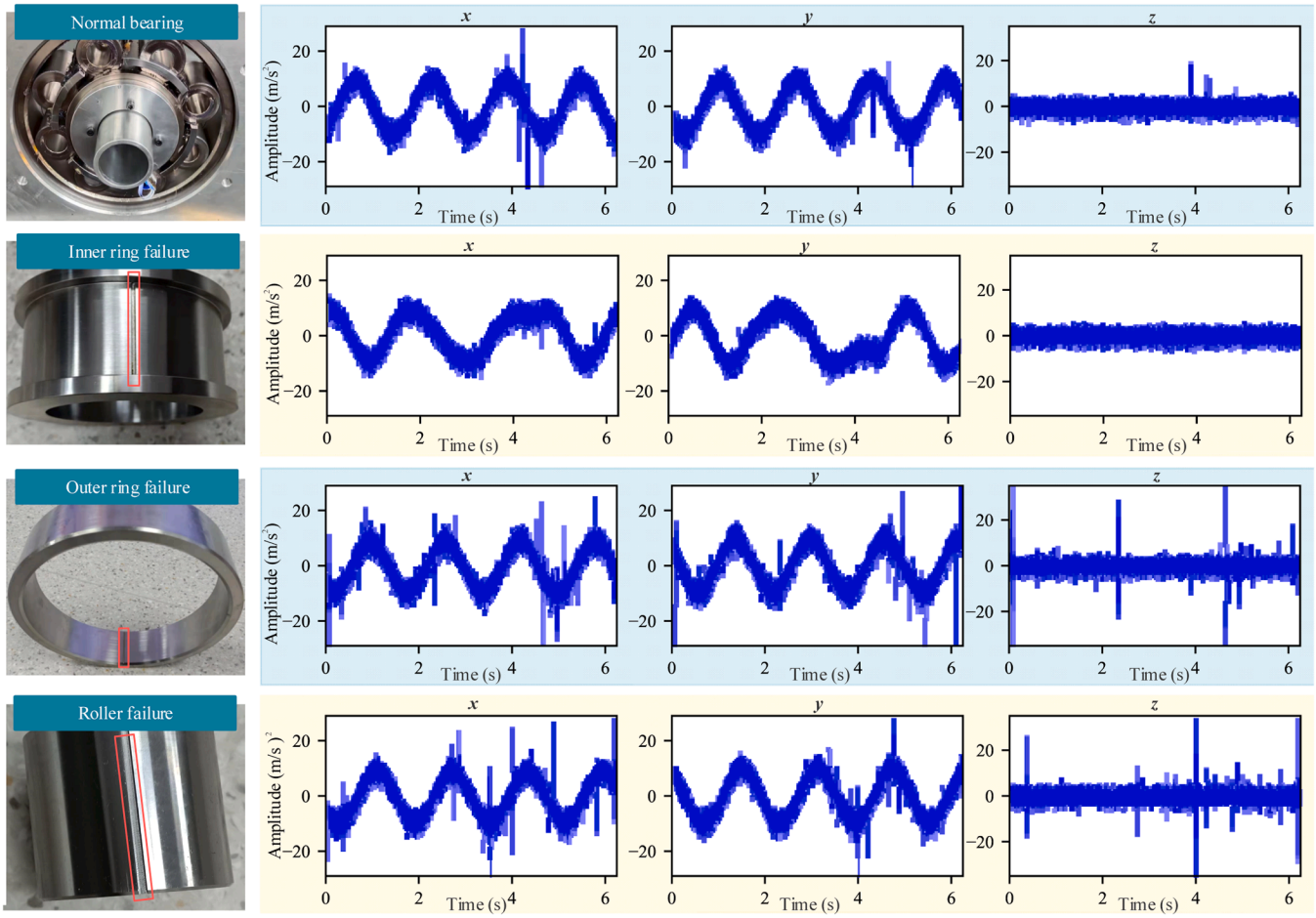


Fig. 5. The roller three-axis vibration signals corresponding to NB, IF, OF, and RF state at a speed of 30 rpm.

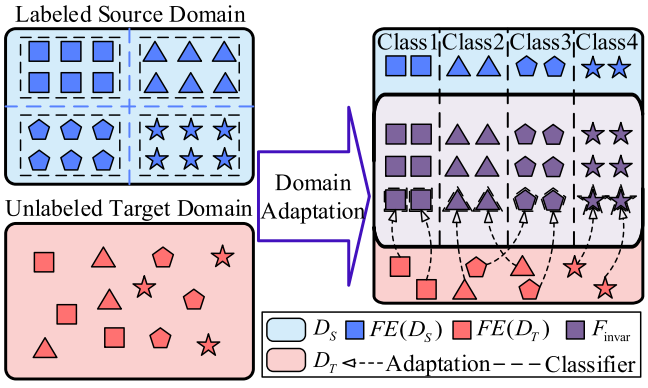


Fig. 6. The schematic of the domain adaptation process.

calculation process, allowing for the generalization of bearing states across different conditions. The mathematical expression of this process is shown in Eq. (2):

$$\begin{aligned}(\widehat{\theta}_e, \widehat{\theta}_c) &= \arg\min_{\theta_e,\theta_c} L_{total}(\theta_e, \widehat{\theta}_d, \theta_c) \\ \widehat{\theta}_d &= \arg\max_{\theta_d} L_{total}(\widehat{\theta}_e, \theta_d, \widehat{\theta}_c)\end{aligned} \tag{2}$$

3.3. Prototype network

Meta-learning is a cross-task learning approach with the ability to generalize to new tasks, often applied in transfer learning [49] and few-shot learning [50] tasks. Meta-learning divides the dataset into Z (support set) and Q (query set), where $Z = \left\{\left(x_i^Z, y_i^Z\right)\right\}_{i=1}^{N_Z}$ and $Q = \left\{\left(x_i^Q, y_i^Q\right)\right\}_{i=1}^{N_Q}$. x_i^Z and x_i^Q represent the i-th sample in set Z and Q, respectively. The sample label y_i has the same range of values, with N sample classes and each class comprising K samples, and $N_Z = N_Q = N \times K$. Therefore, meta-learning classification problems are commonly abbreviated as "N-way K-shot".

Prototype networks [51] are portrayed in Fig. 8 as excellent metric-based meta-learning approaches. With its outstanding few-shot generalization characteristics, the prototype network functions as the classifier in this study, enhancing the domain adaptive classification performance under class-imbalanced and few-shot samples. The implementation of prototype networks involves two main stages: prototype generation and distance measurement. The principle behind prototype generation is to compute the mean of feature points of each class in the metric space from the support set. Eq. (3) illustrates the calculation process.

$$p_n = \frac{1}{K}\sum_{i=1}^{K} FE\left(x_i^Z\right) \tag{3}$$

where p_n represents the prototype generated by the n-th class sample. The distance measurement process is achieved by calculating the Euclidean distance $d\left(FE\left(x_i^Q\right), p_n\right)$ between the prototype p_n and the feature points $FE\left(x_i^Q\right)$ of the query set sample. This distance is then input into the softmax function to predict the category labels for the query set samples. Its mathematical expression is shown in Eq. (4):

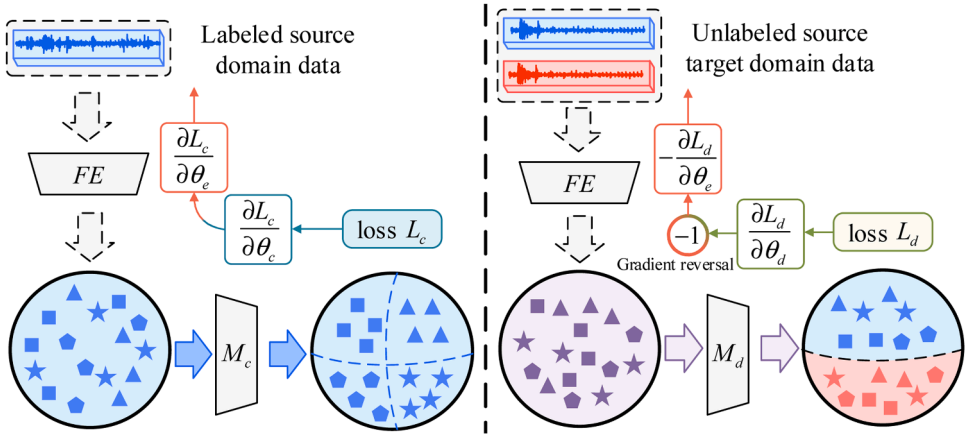


Fig.7. The schematic of the domain-adversarial learning process.

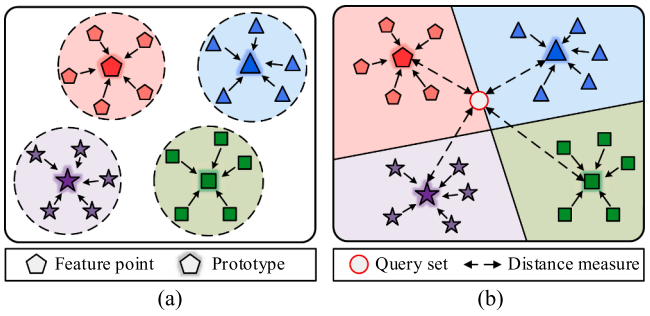


Fig.8. Prototype network: (a) Prototype generation, (b) Distance measure.

$$P(y=n|x\in Q)=\frac{\exp\left[-d\left(FE\left(x_i^Q\right),p_n\right)\right]}{\sum_{j=1}^{N}\exp\left[-d\left(FE\left(x_i^Q\right),p_j\right)\right]} \tag{4}$$

The training strategy of the prototype network involves training on the support set during each iteration and conducting pre-testing on the query set. The objective is to consistently decrease the cross-entropy loss L_q between forecasted and actual labels in query set, as shown in Eq. (5), to achieved the model generalized learning for new fault categories.

$$L_q=-\frac{1}{n_q}\sum_{(x,y)\in Q}\log P(y=n|x\in Q) \tag{5}$$

4. Proposed method

The proposed domain-adaptive fault diagnosis method is outlined in Fig. 9. This method is made up of four primary components: preprocessing, a multi-stream fusion feature encoder, a domain discriminator, and a prototype network class discriminator, which respectively play the roles of obtaining features from data, determining the origin of features, and identifying fault types. In the training phase, training involves the inclusion of labeled samples from D_S and unlabeled samples from D_T. Among them, the domain discriminator receives unlabeled source and target domain samples, whereas the class discriminator is trained exclusively on labeled source domain data. The encoder and M_d discriminator are key components forming an adversarial relationship, where their mutual competition during training facilitates domain adaptation from the D_S to the D_T.

4.1. Detrending: roller motion trend

Monitoring the roller state signal during rotation will inevitably introduce motion information, greatly affecting the recognition of the bearing state. Common signal trend removal algorithms include the

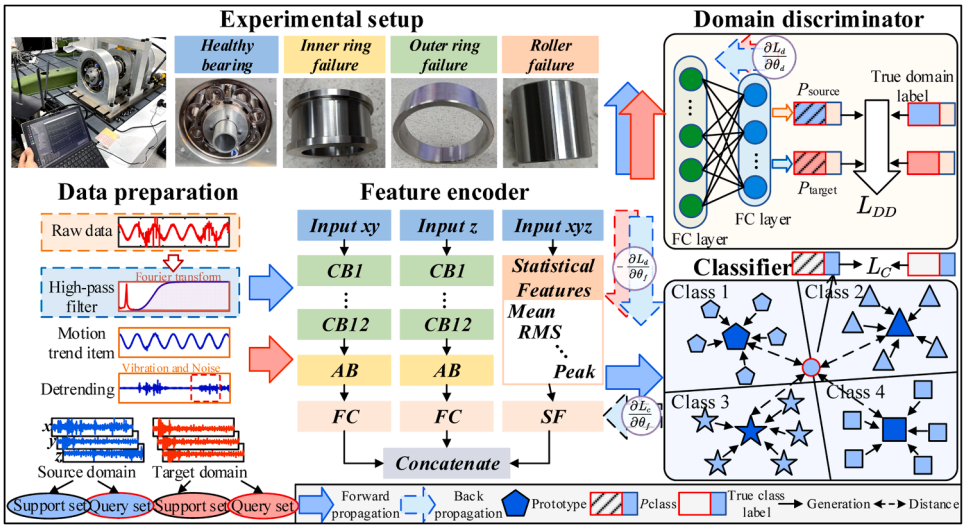


Fig. 9. The architecture of cross-domain few-shot large main bearing diagnosis algorithm based on MDAPN.

finite/infinite impulse response filter method [52], exponential weighted moving average method, smoothing prior method, variational mode decomposition [53], etc. Due to the characteristics of high amplitude and low frequency of the roller motion trend signal, designing a high-pass filter can efficiently achieve the separation of low-speed roller motion signals under various speed conditions, as shown in Fig. 10. The specific implementation process includes: (1) Performing Fourier transform on the signals, detecting the position of maximum amplitude in the frequency domain as the design basis for the high-pass filter; (2) Designing an infinite impulse response high-pass filter based on Butterworth, with the following parameters: the order is 5, passband cutoff frequency f_P=3 Hz, stopband cutoff frequency f_S=7 Hz, actual passband ripple R_P=3 dB, minimum stopband attenuation R_S=30 dB. The amplitude frequency characteristic curve of the filter is shown in Fig. 10; (3) Performing high-pass filtering on the signal to separate the motion trend term.

4.2. Architecture of MDAPN

(1) Feature encoder: The proposed method designs a multi-stream fusion feature encoder, which utilizes a two-stream CNN to merge roller axial-radial features. The model also incorporates time-domain statistical features from triaxial vibration signals, thereby improving its generalization capabilities. The architecture of the encoder is depicted in Fig. 11, and its information stream processing can be primarily separated into three sections: the first section involves feeding the detrending x- and y-axis roller radial vibration signals into a multi-channel CNN with attention mechanisms; the second part processes the z-axis roller axial vibration signals using a neural network with a similar structure; the third part extracts time-domain statistical features from the three-axis roller signals and incorporates them into the fully connected layer. Finally, three information streams converge within the fully-connected layer. Among them, the network architecture for the first two parts includes 12 identical convolutional modules (The number of convolution kernels remains unchanged after increasing to 128), one attention module, and one flatten layer, respectively. The convolutional module consists of convolutional layers f_{conv}, batch normalization layers f_{BN}, activation functions f_{ReLU}, and pooling layers f_{pool} connected in sequence. The expression for the convolution processing process is shown in Eq. (6):

$$f_{conv}(x) = \left[\sum_{i=1}^{n_c} w(j) * x(i) + b(j)\right]_{j=1}^{n_k} \tag{6}$$

where w represents weights and b denotes biases of convolutional kernels, n_c and n_k are the channel quantity and convolutional kernel quantity, respectively. $*$ denotes the convolutional operation. The expression for the batch normalization process is given by Eq. (7):

$$f_{BN}(x) = \frac{x - \mathrm{E}[x]}{\sqrt{\mathrm{Var}[x] + \varepsilon}} * \gamma + \beta \tag{7}$$

where $\mathrm{E}[x]$ and $\mathrm{Var}[x]$ represent the mean and variance of samples, γ and β are the parameters to be trained, and the hyperparameter ε defaults to a value of 1×10^{-5}. f_{ReLU} employs Rectified Linear Unit (ReLU) to perform nonlinear processing on the output of convolutional layers, which is expressed as $f_{ReLU} = \max\{0, x\}$. Max pooling is applied in the pooling layer to downsample the data. The attention module adopts the "Squeeze-and-Excitation" channel attention mechanism [54]. The flatten layer flattens the local features obtained by convolutional kernels. Among them, the specific implementation steps of the attention module consist of squeeze, excitation, and re-calibration. Assuming the input feature map is $f_{map} \in \mathbb{R}^{K \times W}$, K and W represent the number of convolutional kernels in the previous layer and the width of the feature map, respectively. The squeeze step reduces the feature map dimensionality by global average pooling to obtain the importance weights for each convolutional kernel. The expression for the squeezed tensor $q_k \in \mathbb{R}^{K \times 1}$ is given by Eq. (8). The excitation step involves learning the importance weights $e_k \in \mathbb{R}^{K \times 1}$ through a two-layer fully-connected network, as shown in Eq. (9).

$$s_k = F_{sq}\left(f_{map}^k\right) = \frac{1}{W}\sum_{i=1}^{W} f_{map}^k(i) \tag{8}$$

$$e_k = F_{ex}(s_k) = f_{sig}\left(W_2 f_{\mathrm{ReLU}}(W_1 s_k)\right) \tag{9}$$

where f_{sig} is the sigmoid activation function, $W_1 \in \mathbb{R}^{\frac{K}{r} \times K}$ and $W_2 \in \mathbb{R}^{K \times \frac{K}{r}}$ is the weight matrix, and r is the channel reduction ratio. Finally, in the re-calibration process, the importance weight is multiplied element-wise with the original feature map to obtain a weighted feature map.

The input of the third information stream includes the detrending radial signals and the axial signal of smart rollers. The three-axis signal extracts 8 time-domain statistical features as listed in Table 1, including mean, Rectified mean, RMS, peak, peak-to-peak, shape factor, clearance factor, and margin factor. Ultimately, three sets of information stream features are fused at the fully connected layer as inputs to the M_d and M_c discriminator. Model parameters of the MDAPN feature encoder are shown in Table 2.

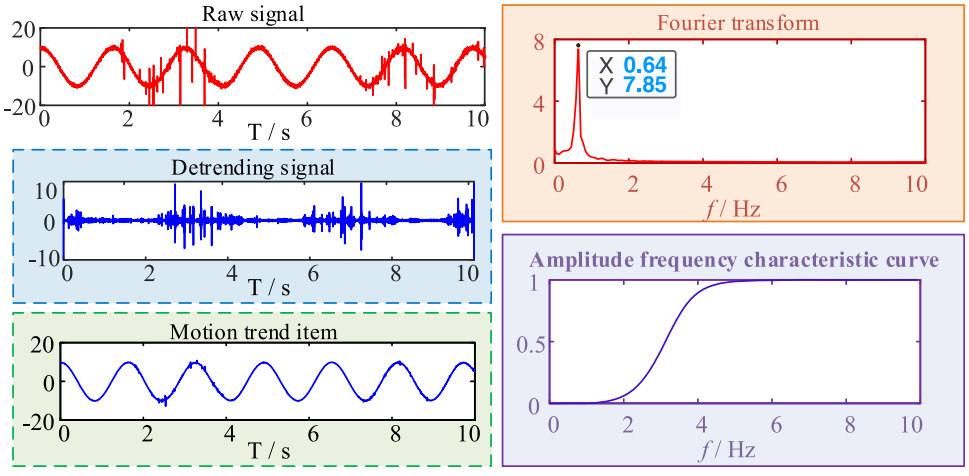


Fig.10. Schematic diagram of the filtering process of the roller motion trend item.

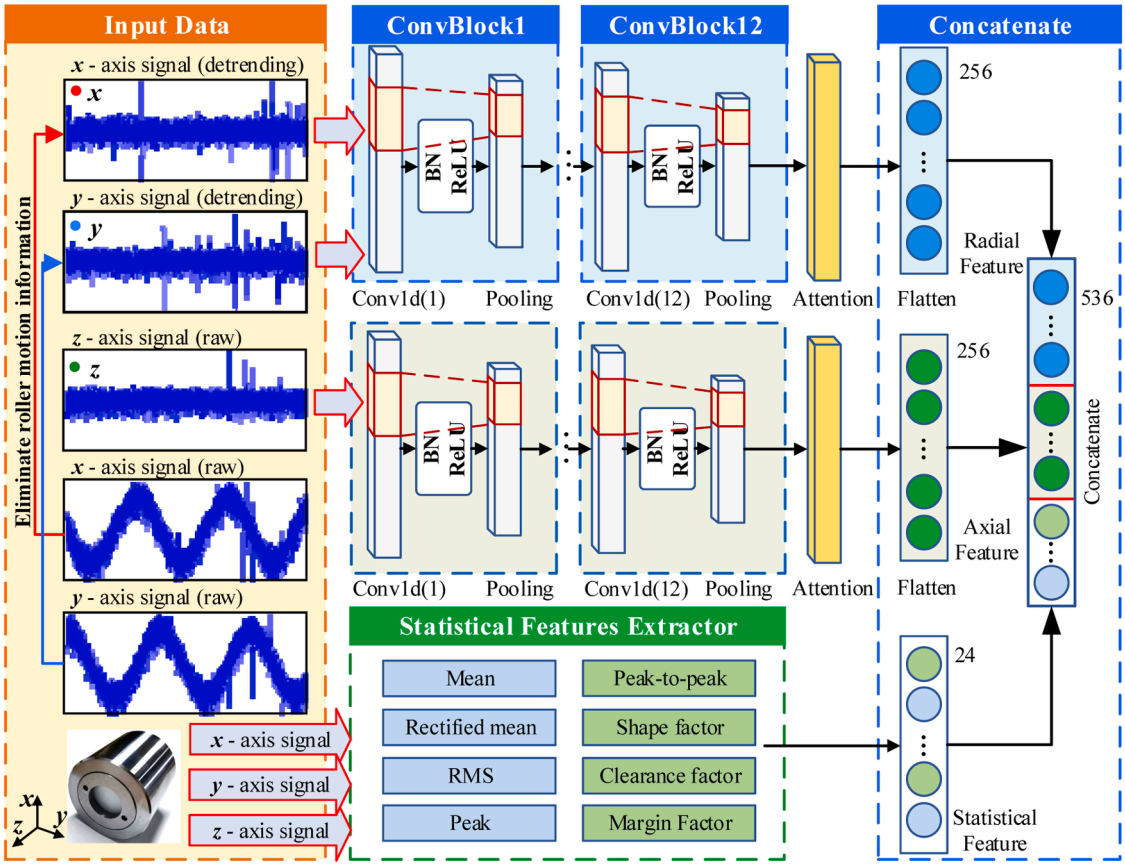


Fig.11. The network architecture of the multi-stream fusion feature encoder.

Table 1
Mathematical expressions for time-domain statistical features.

Name	Expression	Name	Expression
Mean	$\bar{x} = \frac{1}{L}\Sigma_{i=1}^{L} x_i$	Peak-to-peak	$x_{pp} = \max(x_i) - \min(x_i)$
Rectified mean	$\bar{x}_{abs} = \frac{1}{L}\Sigma_{i=1}\lvert x_i\rvert$	Shape factor	$x_{sf} = \frac{x_{rms}}{\bar{x}_{abs}}$
RMS	$x_{rms} = \sqrt{\frac{1}{L}\Sigma_{i=1}^{L} x_i^2}$	Clearance factor	$x_{cf} = \frac{x_{pp}}{x_{rms}}$
Peak	$x_{pk} = max(\lvert x_i\rvert)$	Margin factor	$x_{mf} = \frac{x_{pp}}{\left[\frac{1}{L}\Sigma_{i=1}^{L}\sqrt{\lvert x_i\rvert}\right]^2}$

(2) Domain discriminator: A simple multi-layer perceptron and a softmax layer comprise the domain discriminator. Among them, the multi-layer perceptron consists of two linear modules and activation functions. The output is computed by the following Eq. (10):

$$z_i = M_d(FE(x_i)) = w_2 f_{ReLU}(w_1 FE(x_i) + b_1) + b_2 \tag{10}$$

The multi-layer perceptron output is processed through a softmax layer to calculate the domain probabilities, as shown in the following Eq. (11):

$$P(y=i) = \text{softmax}(z_i) = \frac{e^{z_i}}{e^{z_0} + e^{z_1}} \tag{11}$$

where i denotes the judgment of M_d, with a integer of 0 or 1, indicating D_S or D_T, respectively.

Table 2
Parameter description of MDAPN feature encoder model.

Part	Description	Details	Output size
Radial feature	Input_radial	Shape = (2 × 8192)	2 × 8192
	Conv01_block	Kernel_num = 16, Size = 3, Step = 1, BatchNorm1d, Activation = ReLU, Maxpool1d (size = 2)	16×4096
	…	…	…
	Conv12_block	Kernel_num = 128, Size = 3, Step = 1, BatchNorm1d, Activation = ReLU, Maxpool1d (size = 2)	128 × 2
	Attention	Avgpool1d(1), Linear,Activation = ReLU, Linear, Activation = Sigmoid	128 × 2
	Flatten	/	256
Axial feature	Input_axial	Shape = (1 × 8192)	1 × 8192
	Conv01_block	Kernel_num = 16, Size = 3, Step = 1, BatchNorm1d, Activation = ReLU, Maxpool1d (size = 2)	16×4096
	…	…	…
	Conv12_block	Kernel_num = 128, Size = 3, Step = 1, BatchNorm1d, Activation = ReLU, Maxpool1d (size = 2)	128 × 2
	Attention	Avgpool1d(1), Linear,Activation = ReLU, Linear, Activation = Sigmoid	128 × 2
	Flatten	/	256
Statistical feature	Input_sf	Shape = (3 × 8192)	3 × 8192
	Feature extractor	Mean, Rectified mean, RMS, Peak, Peak-to-peak, Shape factor, Clearance factor, Margin factor	8; 8; 8
	Flatten	/	24
	Concatenate	Radial feature, Axial feature, Statistical feature	536

(3) Class discriminator: M_c employs a prototype network, calculating the prototype for each category according to Eq. (3), and the probability that each sample in the query set belongs to each category is calculated using Eq. (4).

4.3. Learning strategy and optimization

The learning strategy for the MDAPN method is described in Section 3.2, in accordance with the domain adversarial learning theory. The input of the class discriminator is labeled source domain samples. The loss function L_c for M_c during the training process is shown in Eq. (12). The input for M_d is unlabeled source and target domain samples. The loss function L_d for the domain discriminator is shown in Eq. (13).

$$L_c\left(X^S, Y^S; \theta_e, \theta_c\right) = -\frac{1}{N \times n_q} \sum_{i=1}^{N \times n_q} \log P\left(y = y_i^S | x_i \in Q\right) \tag{12}$$

$$L_d\left(X^S, X^T; \theta_e, \theta_d\right) = -\frac{1}{n_s} \sum_{i=1}^{n_s} \log P\left(y = 0 | x_i \in X^S\right) - \frac{1}{n_t} \sum_{i=1}^{n_t} \log P\left(y = 1 | x_i \in X^T\right) \tag{13}$$

Algorithm 1 displays the training process of MDAPN. The loop is executed as long as the loss function continues to decrease. Training is stopped when consecutive increases in the loss function L_c and L_d are detected, at which point the obtained parameters are considered the optimal parameters, following the "early stopping" method. The model utilizes the Adam optimizer [55] to update the parameters. This optimizer incorporates momentum and adaptive learning rates, offering the advantages of fast convergence and high computational efficiency. The parameters set in this paper are as follows: first-order and second-order moment estimation exponential decay rates are $\beta_1 = 0.9$ and $\beta_2 = 0.99$, numerical stability constant $\epsilon = 10^{-6}$, and initial learning rate $\eta = 0.001$.

5. Experimental verification and discussion

In this section, the efficiency of the MDAPN algorithm is validated by extensive fault diagnosis experiments. The process of dataset partitioning and the experimental details for conducting comparative experiments. This section further explores the impact of different sample lengths and class imbalance circumstances on MDAPN in Section 5.3 and elucidates the advantages of the multi-stream fusion feature encoder and prototype network classifier. The proposed MDAPN model is implemented using the PyTorch deep learning framework, with batch size and epochs set to 20. The model updates parameters using the Adam optimizer. The computer configuration is NVIDIA GeForce RTX 3060, CPU i5–12600KF, and 16 GB of memory.

5.1. Descriptions of datasets

Section 2 introduces details of conducting low-speed heavy-load large-bearing fault simulation experiments. This section will further partition the smart roller state signals collected from the experiments into datasets. A sliding window of length 8192 is selected to perform overlapping sampling on the three-axis vibration data of rollers under working conditions of 30 rpm, 40 rpm, and 50 rpm, resulting in individual samples of size 3×8192. For each working condition, 200 samples are chosen at random for both the training and testing sets, ensuring that there is no data leakage between the training and testing sets. This study sets 6 variable-speed diagnosis tasks outlined in Table 3, to validate the cross-domain diagnostic capability of the proposed algorithm through subsequent comparative experiments.

5.2. Experimental details and result analysis

Although the smart roller monitoring system can avoid complex external interference signals and retain fault information close to the fault source. However, the low operational speed of the TBM main bearing results in limited effective information in individual samples of roller state data. Additionally, changes in working conditions cause a shift in sample distribution, and samples are scarce for a single working condition. This section conducts experiments on the six cross-domain diagnostic tasks established in Table 3 and verifies the superiority of the MDAPN algorithm through comparisons with existing approaches. Since the classifier adopts a prototype network structure, the model training process follows the 4-way 5-shot meta-learning paradigm, with 200 training samples per class and only 5 available samples per class in the D_T. The sample quantity in query set and support set is kept consistent. In terms of experiments, the comparative experiment applied five classic methods as shown in Table 4. This includes the traditional machine learning method: support vector machines (SVM) [56]; the

Table 3
Details of cross-domain fault task for main bearings.

Tasks	Path	Source domain	Target domain	Fault types	Load/ MPa
T_{34}	30 rpm → 40 rpm	30 rpm/ n_{train}=200	40 rpm/ n_{target}=5	NB/IF/ OF/RF	20
T_{35}	30 rpm → 50 rpm	30 rpm/ n_{train}=200	50 rpm/ n_{target}=5	NB/IF/ OF/RF	20
T_{43}	40 rpm → 30 rpm	40 rpm/ n_{train}=200	30 rpm/ n_{target}=5	NB/IF/ OF/RF	20
T_{45}	40 rpm → 50 rpm	40 rpm/ n_{train}=200	50 rpm/ n_{target}=5	NB/IF/ OF/RF	20
T_{53}	50 rpm → 30 rpm	50 rpm/ n_{train}=200	30 rpm/ n_{target}=5	NB/IF/ OF/RF	20
T_{54}	50 rpm → 40 rpm	50 rpm/ n_{train}=200	40 rpm/ n_{target}=5	NB/IF/ OF/RF	20

Algorithm 1
The training procedure of MDAPN.

Input: The labeled source domain data $D_S = \left\{\left(x_i^S, y_i^S\right)\right\}_{i=1}^{N_s}$ and the unlabeled target domain data $D_T = \left\{\left(x_i^T\right)\right\}_{i=1}^{N_t}$.
Output: The feature extractor ***FE*** able to extract domain invariant features.
1: Initialize the feature encoder parameter θ_e, the class discriminator parameter θ_c, and the domain discriminator parameter θ_d.
2: **WHILE** the loss function L_c and L_d is continuously reducing **DO**
3: Randomly select the support set $\left\{\left(x_i^Z, y_i^Z\right)\right\}_{i=1}^{N_z}$, i.e. Z from D_S.
4: Randomly select the query set $\left\{\left(x_i^Q, y_i^Q\right)\right\}_{i=1}^{N_q}$, i.e. Q from D_S.
5: Calculate the prototype p_n by Eq. (3).
6: Calculate the $L_c\left(X^S, Y^S; \theta_e, \theta_c\right)$ on the query set Q by Eq. (12).
7: Calculate the $L_d\left(X^S, X^T; \theta_e, \theta_d\right)$ on X^S and Y^S by Eq. (13).
8: Calculate the $L_{total}\left(X^S, Y^S, X^T; \theta_e, \theta_d\right)$ by Eq. (1).
9: Optimize the parameter θ_c, θ_e.
10: Optimize the parameter θ_d, θ_e.
11: **END WHILE**

feature-based transfer learning approach, transfer component analysis (TCA) [57], facilitates domain adaptation through the reduction of distance in edge probability distribution between D_S and D_T; Deep domain confusion (DDC) approach [58], implemented using CNN, adds an adaptation layer and an additional domain confusion loss function to achieve the goal of obtaining domain-invariant features; Deep model based domain adaptation for fault diagnosis (DAFD) [59], a domain adaptive method also rooted in CNN architecture, minimizes the MMD between D_S and D_T to lessen domain disparities; Domain adversarial neural networks (DANN) [60], an adversarial mechanism-based approach, generates features to confuse the domain discriminators. To truly reflect the performance of the comparison method, DDC, DAFD, DANN, and the proposed method adopt the same feature extractor. Since the input data for the experiments are the three-dimensional vibration signals of rollers, manually extracted features from each dimension of samples are concatenated for input into SVM and TCA models. For other deep learning models, the CNN channel quantity is changed to adapt to the multi-dimensional data.

Table 5 depicts the cross-domain diagnostic outcomes of the MDAPN algorithm and other comparative algorithms. The diagnostic accuracy of MDAPN for six different cross-domain tasks is 97.42 %, 97.38 %, 99.05 %, 98.60 %, 98.81 %, and 99.16 %, respectively. The average recognition accuracy of the MDAPN algorithm is compared to five comparative algorithms, with improvements of 35.13 %, 33.41 %, 25.42 %, 20.68 %, and 6.57 %, respectively. It is evident that the MDAPN accuracy outperforms existing approaches in different tasks. The average accuracy of DAFD is increased by 14.45 % and 12.73 % compared to SVM and TCA, respectively, indicating that deep learning has better feature expression ability than machine learning and traditional transfer learning algorithms. The average accuracy of DANN has increased by 18.85 % and 14.11 % compared to DDC and DAFD, respectively. This indicates that the introduction of adversarial mechanisms significantly reduces the impact of domain shift on model accuracy, and extracted domain-invariant features have stronger generalization ability. Compared to DANN, the average accuracy of MDAPN is improved by 6.57 %, indicating that the improved feature encoder and prototype network classifier enhance the ability to extract domain-generalized features from roller state data with a limited target domain sample quantity. The specific roles of each module are discussed in Section 5.3 and 5.4.

Further analysis of the experimental outcomes of the MDAPN model across various tasks reveals that the diagnostic accuracy for tasks T_{34}, T_{35}, and T_{45} is lower than that for other tasks. This is attributed to the longer periods between defect-related pulses in fixed-length samples under low-speed operating conditions, making it more challenging to extract fault information for identifying fault types under relatively higher-speed conditions. Consequently, the T_{35} task is more challenging compared to other tasks. When the speed is transferred from 30 to 50 rpm, the accuracy of SVM, TCA, DDC, DAFD, DANN, and the proposed methods are 67.52 %, 69.25 %, 72.62 %, 93.99 %, 88.66 %, and 97.39 %, respectively. To further illustrate the specific correct classification and misclassification information of each method under this task, a confusion matrix, as depicted in Fig. 12, was generated. It is evident from the graphic that DDC has obvious phenomena of missed alarms and false alarms. Although SVM, TCA, DAFD, and DANN perform well in recognizing certain states, they all suffer from severe misclassification issues. In contrast, the proposed method accurately identifies various fault types. Combining the T-Stochastic Neighbor Embedding (T-SNE) visualization in Fig. 13, it makes sense intuitively that the intra-class clustering effect and inter-class discriminability of the MDAPN model are significantly higher than other algorithms. The better the clustering effect of feature points with the same color, and the more obvious the differentiation between feature points with different shapes, the stronger the fault recognition ability of the features. Besides, better clustering of feature points with the same shape signifies smaller differences in feature distribution between D_S and D_T, resulting in less domain shift impact.

Table 4
Key parameters and network structure of comparative methods.

Methods	Feature extractor parameters	Domain adaption key modules
SVM [56]	Three-axis statistical fusion feature	Gaussian kernel function mapping
TCA [57]	Three-axis statistical fusion feature	Reproducing kernel Hilbert Space mapping→MMD
DDC [58]	Three-channel CNN (12 Conv blocks)	Adaptation layer→MMD
DAFD [59]	Three-channel CNN (12 Conv blocks)	MMD→Domain adaptation
DANN [60]	Three-channel CNN (12 Conv blocks)	Adversarial domain adaptation

Table 5
The fault recognition accuracy of different methods under cross-speed conditions.

Methods	T_{34}	T_{35}	T_{43}	T_{45}	T_{53}	T_{54}	avg
SVM	63.53 ±0.14	67.52 ±0.20	57.16 ±0.08	69.04 ±0.14	50.25 ±0.12	72.17 ±0.07	63.28 ±7.72
TCA	62.69 ±0.17	69.25 ±0.13	54.68 ±0.06	63.08 ±0.14	69.82 ±0.06	70.47 ±0.08	65.00 ±5.74
DDC	71.37 ±3.22	72.62 ±2.42	71.14 ±4.15	79.35 ±2.51	72.67 ±2.12	70.81 ±2.50	72.99 ±3.88
DAFD	56.90 ±2.61	93.99 ±1.78	74.84 ±2.93	82.97 ±2.16	76.10 ±2.85	81.54 ±2.42	77.73 ±11.70
DANN	87.69 ±1.36	88.66 ±1.56	91.31 ±1.27	91.76 ±0.99	96.37 ±1.02	95.24 ±1.24	91.84 ±3.41
Proposed	97.42 ±1.19	97.39 ±0.34	99.05 ±0.68	98.60 ±0.72	98.81 ±0.56	99.16 ±0.20	98.41 ±0.95

The comprehensive analysis above leads to the conclusion that the MDAPN model possesses the capability to accomplish cross-speed diagnosis tasks under extremely limited target domain data. In practical applications, it can accurately identify the TBM main bearing status under various low-speed heavy-load conditions, avoiding the occurrence of major safety accidents.

5.3. Discussion

5.3.1. Selection of sample length under low-speed operating conditions

Taking into account that the smart roller of the TBM main bearing gathers vibration signals during low-speed operations is affected by the long period between defect-related pulses, especially for cases in which the target domain sample quantity is scarce, the diagnostic model needs to choose a reasonable sample length. Considering that the integral power of 2 can maximize the utilization of computational resources, sample lengths (1024, 2048, 4096, 8192, and 16384) were selected to observe experimental results. The experimental setup remained consistent with Section 5.3, and the accuracy of MDAPA under different cross-speed tasks is illustrated in Fig. 14. During the increase in sample length from 1024 to 4096, there was a slight improvement in model accuracy. However, noticeable differences in model accuracy for various tasks indicated that the sample interval could not cover the complete bearing fault information, resulting in a lack of sufficient fault information in the model. When the sample length reached 8192, the accuracy of the model under different tasks improves significantly and there is no significant difference. After further increasing the sample length, the model accuracy is essentially unaffected by the sample length, as the sample interval now fully covers the fault impulse signals of each component. Therefore, choosing 8192 as the sample length is the most appropriate.

5.3.2. Performance analysis of multi-stream fusion feature encoder

The network architecture of the multi-stream fusion feature encoder is depicted in Fig. 11. To explore the advantage of the two-stream CNN in processing roller state data and the importance of introducing

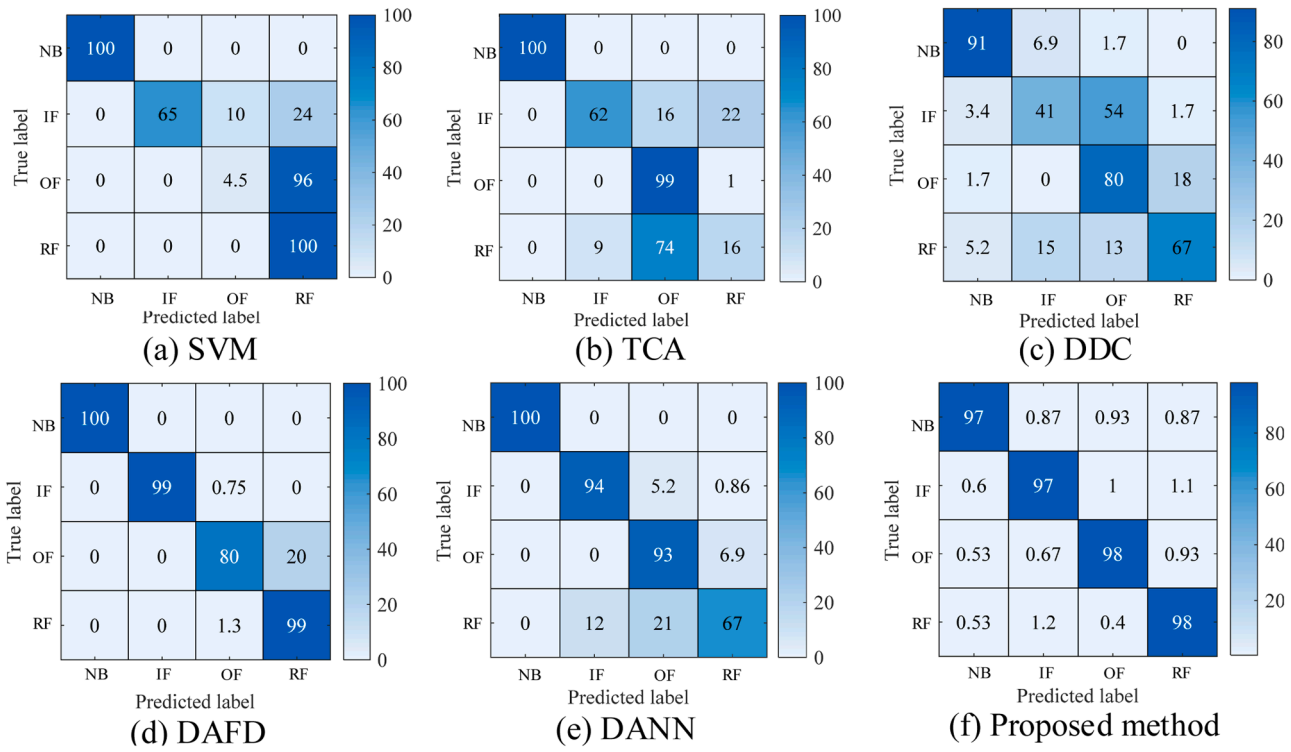


Fig.12. Confusion matrix of experimental results from various algorithms under T_{35} task.

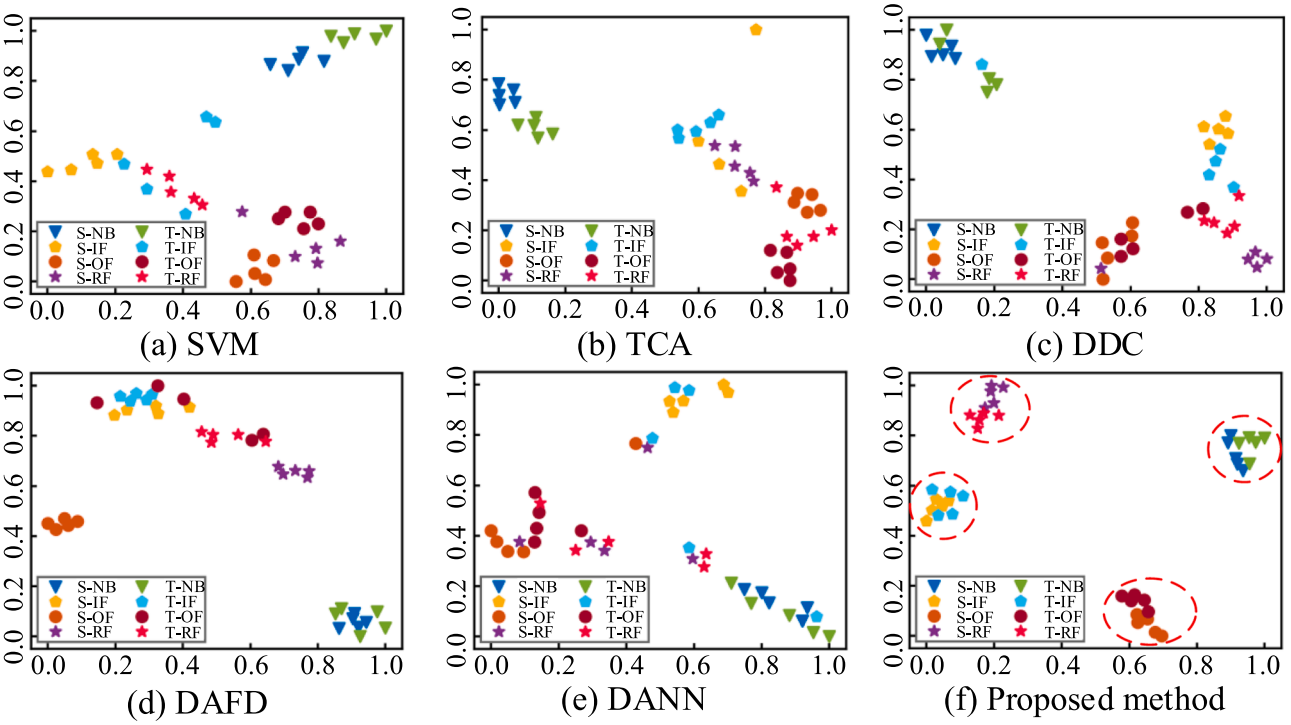


Fig.13. T-SNE dimensionality reduction visualization results using different methods under T_{35} task.

statistical feature information. In this section, the feature encoder of the MDAPN model is replaced with a three-channel CNN with statistical features and a two-stream CNN, resulting in the comparative methods DCNN-SF-DAPN and TSCNN-DAPN. The network structures of the feature encoders for the three methods are depicted in Fig. 15. Cross-domain diagnostic experiments were carried out to compare results of three methods in various speed conditions, as shown in Table 6. Compared to the MDAPN method, the accuracy of the DCNN-SF-DAPN method decreased by 5.90 %, 4.63 %, 5.99 %, 4.30 %, 2.68 %, and 0.07 % in different tasks, indicating that using a two-stream CNN to analyze roller data can obtain more complete state characteristics. However, the multi-channel CNN structure assumes that the axial and radial features of rollers have the same modal characteristics, leading to a decrease in recognition accuracy. Meanwhile, the accuracy of the

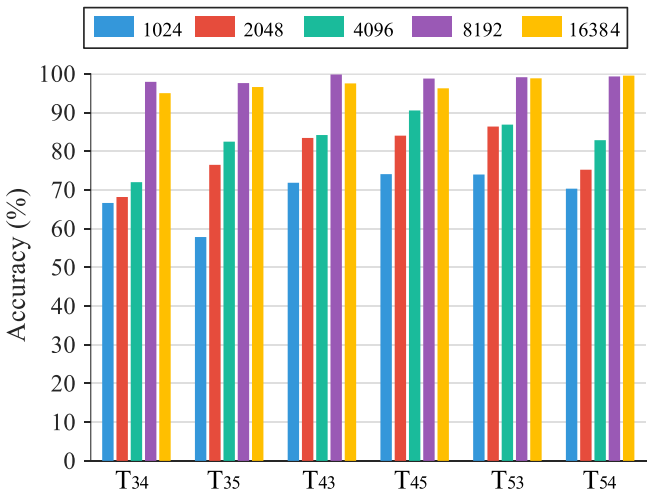


Fig.14. Experimental results of MDAPN under different cross-domain tasks with different sample lengths.

TSCNN-DAPN method when operating conditions transition from low to high speeds ($T_{34}/T_{35}/T_{45}$) is significantly higher than when the operating conditions transition from high to low speeds ($T_{43}/T_{53}/T_{54}$). In comparison, the accuracy of the proposed method was improved by −0.96 %, −0.40 %, 9.72 %, 7.10 %, 16.54 %, and 18.05 % in different tasks, indicating that the introduced statistical feature can significantly enhance the few-shot cross-domain diagnostic performance during the transition from high-speed to low-speed conditions. This improvement is attributed to the fact that when the target domain samples are scarce, shallow fault feature information compensates for the crucial fault information lost during the learning process of deep network structures.

To intuitively explain the influence of different feature encoder structures, visualization of convolution kernels in the feature encoding process was conducted, as depicted in Fig. 16. By comparing the feature maps in Fig. 16(b) (c) and Fig. 16(a), it can be observed that two-stream CNN-based feature encoders have a pronounced energy concentration effect on the extracted features, with clear positions of energy components representing the state and a significant reduction in discrete energy. Comparing Fig. 16(a) (c) and Fig. 16(b) reveals that the feature energy components, extracted by the feature encoder with time-domain statistical features, are densely distributed, while the energy component distribution in Fig. 16(b) is relatively scattered, possibly retaining ineffective energy components that interfere with model judgment.

5.3.3. Performance analysis of prototype network classifier

To validate the significant role of prototype networks in enhancing the few-shot learning capability of domain adaptation models, in this section, we introduced a prototype network classifier into the baseline DANN method, resulting in the variant method DCNN-DAPN. Additionally, we replaced the prototype network classifier in the TSCNN-DAPN method with a multi-perceptron, yielding the variant method TSCNN-DANN. We conducted ablation experiments around the prototype network classifier, comparing four variant methods and the proposed method, analyzing their performance on six cross-domain diagnostic tasks under different numbers of training samples. The experimental results are shown in Table 7. The number of training samples for each class was set to 500, 200, 100, 50, and 30, respectively, and the average and standard deviation of the accuracy of the six cross-domain tasks for each method were recorded. Compared to DANN and TSCNN-DANN without prototype network classifiers, DCNN-DAPN and TSCNN-DAPN showed the similar diagnostic accuracy when $n_{\text{train}} = 500, 200, 100$. However, under conditions of scarce samples ($n_{\text{train}} = 50, 30$), the accuracy of DCNN-DAPN increased by 5.75 % and 18.06 %, respectively, and the accuracy of TSCNN-DAPN increased by 8.85 % and 19.18 %, respectively. This indicates that the application of prototype network classifiers in domain adaptation models exhibits good few-shot learning performance under variable conditions with scarce available samples.

In this experiment, it is noteworthy that the baseline domain adaptation method, DANN, did not show a significant improvement in average diagnostic accuracy before and after using two-stream CNN. However, as indicated in Section 5.3.2, MDAPN showed a significant improvement in accuracy compared to DCNN-SF-DAPN. Analysis combined with Fig. 16 shows that the reason is that two-stream CNN has strong capabilities in extracting energy components but it is also more susceptible to the influence of ineffective components. This leads to excellent performance of two-stream CNN-based methods in individual cross-domain tasks, but the average diagnostic accuracy is not high. Introducing time-domain statistical feature information removes the

Table 6
Ablation experimental results of three methods under different variable speed working conditions.

Methods	T_{34}	T_{35}	T_{43}	T_{45}	T_{53}	T_{54}	avg
DCNN-SF-DAPN	92.02	92.97	93.83	94.47	96.43	99.25	94.83
TSCNN-DAPN	98.28	98.00	90.10	91.67	82.57	81.27	90.32
MDAPN	97.92	97.60	99.82	98.77	99.11	99.32	98.76

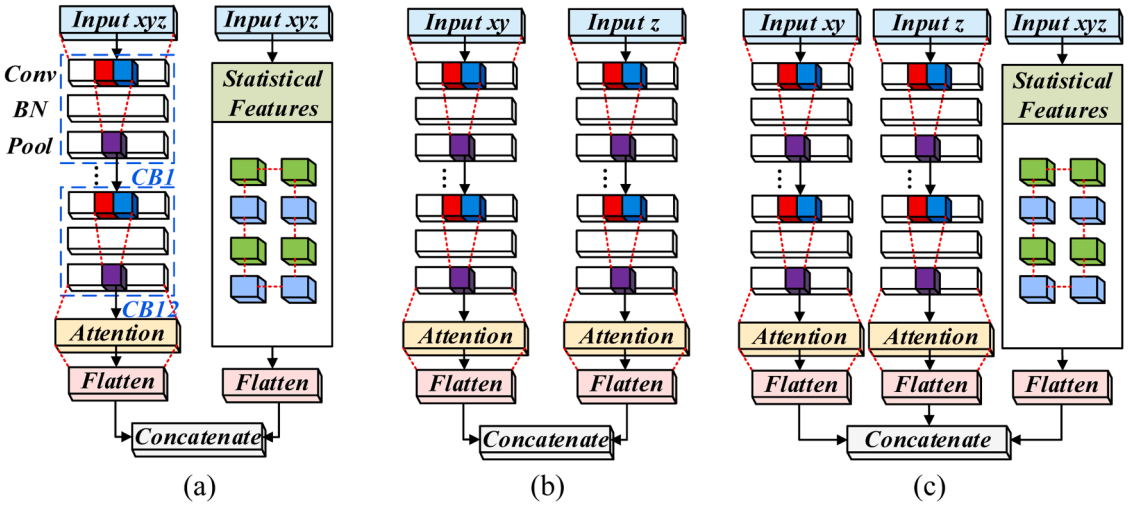


Fig.15. Feature encoder network structure: (a) DCNN-SF-DAPN, (b) TSCNN-DAPN, (c) MDAPN.

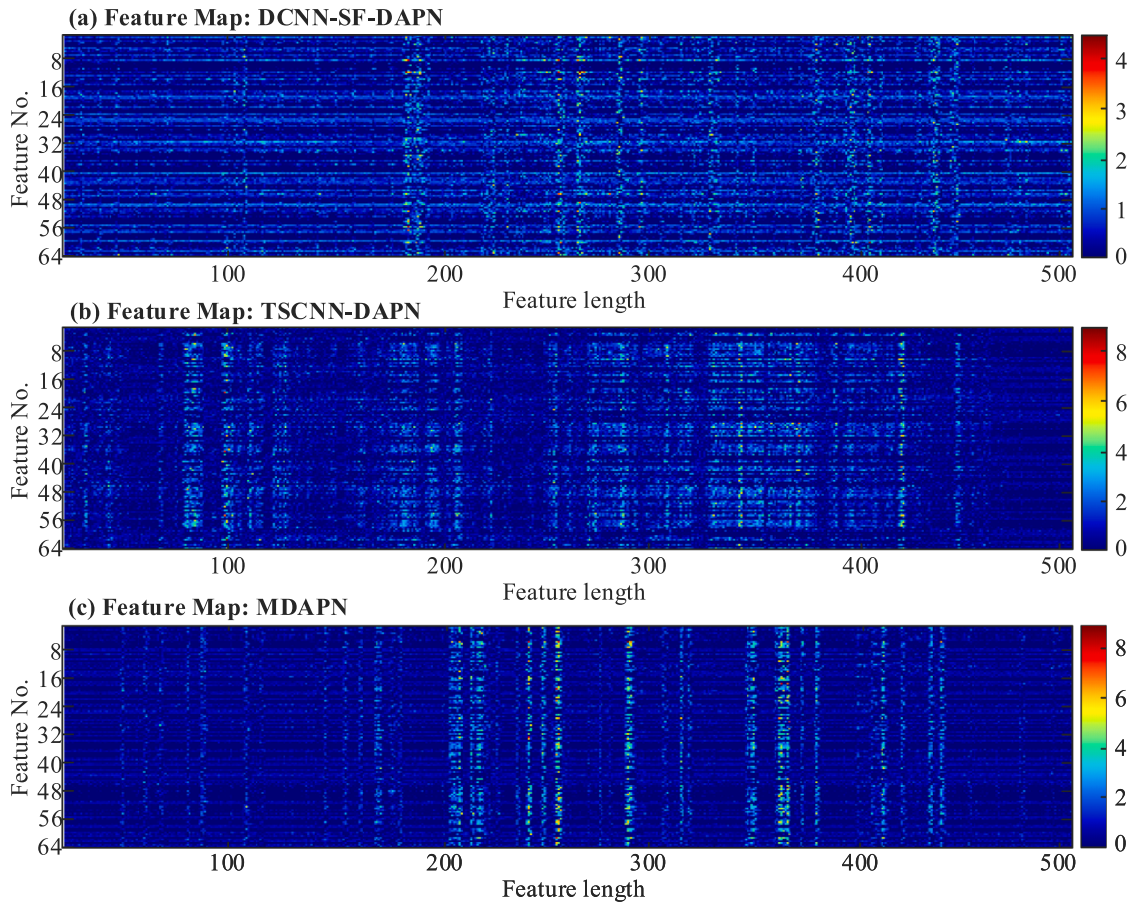


Fig.16. Visual feature maps of convolutional kernels in the feature encoding process for three methods.

Table 7
Ablation experimental results of prototype network classifier under different training sample sizes.

Methods	n_{train}				
	500	200	100	50	30
DANN	94.29 ±4.36	92.10 ±4.06	87.17 ±3.58	74.35 ±2.17	59.76 ±3.74
DCNN-DAPN	91.41 ±6.16	88.25 ±6.55	85.09 ±7.51	80.10 ±6.49	77.82 ±6.59
TSCNN-DANN	92.66 ±6.51	90.13 ±7.34	86.43 ±7.17	74.02 ±5.78	61.66 ±4.58
TSCNN-DAPN	92.65 ±5.95	90.32 ±7.30	87.20 ±6.97	82.87 ±6.41	80.84 ±6.63
MDAPN	98.72 ±0.75	98.76 ±0.85	96.65 ±0.69	92.40 ±1.22	90.74 ±1.04

interference of ineffective energy components, greatly leveraging the advantages of two-stream CNN, thus resulting in a significant improvement in average diagnostic accuracy.

5.3.4. *The influence of class-imbalance working conditions*

The multi-stream fusion feature encoder-based domain adversarial prototype network diagnostic algorithm is proposed, achieving cross-condition fault recognition when target domain samples are sparse. In practical application scenarios, normal data obtained from monitoring devices is rich while fault data is scarce. Therefore, this section further discusses the diagnostic ability of the model under class-imbalance conditions. The number of normal training samples is kept constant at 200, with class-imbalance rates α set at 100 %, 50 %, 20 %, 10 %, and 5 %, respectively. That is, 200, 100, 40, 20, and 10 fault samples are selected for each type of fault as training data. The target domain still maintains 5 available samples per class, and the model is trained according to the 4-way 5-shot paradigm. The experimental findings are visually presented in Fig. 17. The MDAPN average diagnostic accuracy under different class-imbalanced rates is 98.76 %, 95.52 %, 90.49 %, 86.87 %, and 82.23 %, respectively. This indicates that even with only 10 training samples per class in the source domain where normal samples are abundant and faulty samples are scarce, the MDAPN method can still achieve 82.23 % average accuracy. Verified that the MDAPN model still exhibits the capability to identify the types of TBM main bearing

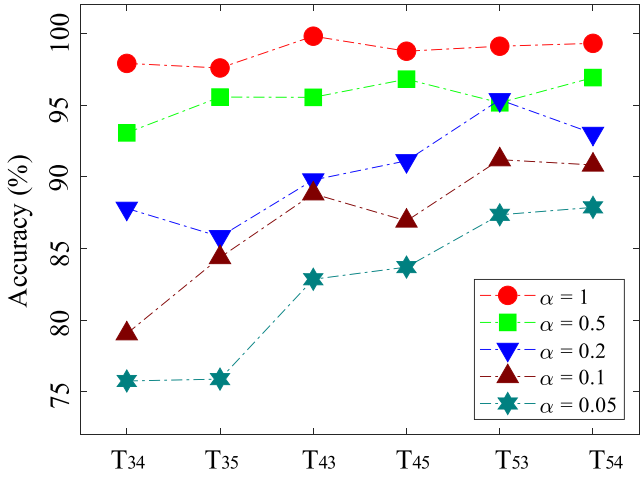


Fig.17. Experimental results of MDAPN under different class-imbalance rates in cross-domain tasks.

faults under extreme class-imbalance conditions.

6. Conclusion

In this work, we conducted research on the TBM main bearing fault diagnosis utilizing data from smart roller monitoring, effectively addressing the challenge of monitoring internal damage behavior in main bearings under complex noise and strong impact conditions at industrial sites. Considering that the TBM main bearing actually operates in conditions characterized by low speed and heavy load, scarce fault samples, and variable working conditions, this study proposes a novel multi-stream fusion feature encoder-based domain adversarial prototype network diagnostic algorithm. It achieves cross-domain fault diagnosis by leveraging roller states with few samples in the target domain. The MDAPN model utilizes a two-stream CNN to deeply mine the axial-radial fusion features of the three-axis monitoring signals of rollers and combines shallow statistical feature information to accurately characterize the main bearing state. Among them, interference signals generated by roller rotation are filtered using modal decomposition. To eliminate the domain shift effects caused by variations in main bearing speeds, an adversarial learning strategy is applied to generate domain-invariant features. In addition, a prototype network classifier is used to achieve fault diagnosis under 5-shot conditions. The results of six cross-domain diagnostic experiments conducted at speeds of 30–40–50 rpm showed that the proposed MDAPN method achieved an average diagnostic accuracy of 98.41 %, significantly outperforming domain-adaptation approaches such as SVM, TCA, DDC, DAFD, and DANN. The efficiency of the MDAPN method is further verified by confusion matrices and T-SNE dimensionality reduction visualization. Additionally, the discussion section explains the selection of sample lengths under low-speed conditions and evaluates the performance of the multi-stream fusion feature encoder and prototype network classifier. Finally, class-imbalance experiments demonstrate that the model maintains an accuracy of 82.23 % when there are only 10 samples per fault class in the source domain, indicating its applicability under class-imbalanced conditions.

In future research, our focus will be on developing dynamic models for atypical faults in TBM main bearings and exploring the domain generalization technology in open-set scenarios. In addition, how to implement the algorithms to sequential online data in industrial settings is also an issue worth researching.

CRediT authorship contribution statement

Xingchen Fu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Data curation, Conceptualization. **Keming Jiao:** Software, Methodology, Data curation, Conceptualization. **Jianfeng Tao:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition. **Chengliang Liu:** Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Key R&D Program of China [Grant No. 2020YFB2007202].

References

[1] Liu Q, Huang X, Gong Q, et al. Application and development of hard rock TBM and its prospect in China[J]. Tunnell Underground Space Techn 2016;57:33–46.
[2] Rostami J. Performance prediction of hard rock tunnel boring machines (TBMs) in difficult ground[J]. Tunnell Underground Space Techn 2016;57:173–82.
[3] Zhang H. Analysis on the causes of main bearing damage of TB880E tunnel boring machine [J]. Tunnel Construct 2014;34(11):1092–7 (in Chinese).
[4] Sun B, Yang S. An improved 3D finite difference model for simulation of double shield TBM tunnelling in heavily jointed rock masses: the DXL tunnel case[J]. Rock Mech Rock Eng 2019;52(7):2481–8.
[5] Huo J, Hou N, Sun W, et al. Analyses of dynamic characteristics and structure optimization of tunnel boring machine cutter system with multi-joint surface [J]. Nonlinear Dyn 2016;87(1):237–54.
[6] Li X, Zhong X, Shao H, et al. Multi-sensor gearbox fault diagnosis by using feature-fusion covariance matrix and multi-Riemannian kernel ridge regression[J]. Reliab Eng Syst Saf 2021;216:108018.
[7] Zhang Y, Wang W, Wu X, et al. A comprehensive review on self-powered smart bearings [J]. Renew Sustain Energy Rev 2023;183:113446.
[8] Liu C, Wang F. A review of current condition monitoring and fault diagnosis methods for low-speed and heavy-load slewing bearings[C]. In: 2017 9th International conference on modelling, identification and control (ICMIC). IEEE; 2017. p. 104–9.
[9] Caesarendra W, Tjahjowidodo T. A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing [J]. Machines 2017;5(4):21.
[10] Bashir I, Wang L, Harvey T, et al. Integrated smart bearings for next generation aero-engines Part 1: development of a sensor suite for automatic bearing health monitoring. In: Proceedings of the 1st World Congress on Condition Monitoring (WCCM); 2017. p. 13–6.
[11] Zhang Z, Pang H, Georgiadis A, et al. Wireless power transfer—An overview [J]. IEEE Transact Industrial Electr 2019;66(2):1044–58.
[12] Fu H, Yeatman EM. A methodology for low-speed broadband rotational energy harvesting using piezoelectric transduction and frequency up-conversion [J]. Energy 2017;125:152–61.
[13] Wang Z, Wang W, Gu F, et al. On-rotor electromagnetic energy harvester for powering a wireless condition monitoring system on bogie frames [J]. Energy Convers Manag 2021;243:114413.
[14] Han Q, Ding Z, Qin Z, et al. A triboelectric rolling ball bearing with self-powering and self-sensing capabilities [J]. Nano Energy 2020;67:104277.
[15] De Bruijn E, Gerard M, Corno M, et al. On the performance increase of wheel deceleration control through force sensing [C]. In: 2010 IEEE International conference on control applications; 2010.
[16] Zhang Y, Cao J. Development of self-powered smart bearing for health condition monitoring [C]. In: 2018 IEEE/ASME International conference on advanced intelligent mechatronics (AIM); 2018.
[17] Bankestrom J. O. Load sensing bearing: U.S. Patent 5,503,030[P]. 1996-4-2.
[18] Den Haak N S W., Van Amerongen G. Roller with integrated load detection: U.S. Patent 10,767,703[P]. 2020-9-8.
[19] Elfert G., Lueneburg B., Rollmann J., et al. Rolling element for use in a rolling-element bearing: U.S. Patent 11,226,004[P]. 2022–1-18.
[20] Matsuda S., Teramoto T. Load detecting device for roller bearing and roller bearing apparatus: U.S. Patent 8,672,553[P]. 2014–3-18.
[21] Ding P, Wang H, Bao W, et al. HYGP-MSAM based model for slewing bearing residual useful life prediction [J]. Measurement 2019;141:162–75.
[22] Jin X, Chen Y, Wang L, et al. Failure prediction, monitoring and diagnosis methods for slewing bearings of large-scale wind turbine: a Review [J]. Measurement 2021;172:108855.
[23] Fu X, Tao J, Jiao K, et al. A novel semi-supervised prototype network with two-stream wavelet scattering convolutional encoder for TBM main bearing few-shot fault diagnosis [J]. Knowl Based Syst 2024;286:111408.
[24] Yu X, Zhao Z, Zhang X, et al. Statistical identification guided open-set domain adaptation in fault diagnosis [J]. Reliab Eng Syst Saf 2023;232:109047.
[25] Yazdanianasr M, Mauricio A, Gryllias K. Evaluation of the improved envelope spectrum via feature optimization-gram (IESFOgram) for bearing diagnostics under low rotating speeds[C]. In: 2021 7th International conference on condition monitoring of machinery in non-stationary operations (CMMNO). IEEE; 2021. p. 319–24.
[26] Bai R, Meng Z, Xu Q, et al. Fractional Fourier and time domain recurrence plot fusion combining convolutional neural network for bearing fault diagnosis under variable working conditions[J]. Reliab Eng Syst Saf 2023;232:109076.
[27] Caesarendra W, Kosasih B, Tieu AK, et al. Application of the largest Lyapunov exponent algorithm for feature extraction in low speed slew bearing condition monitoring[J]. Mech Syst Signal Process 2015;50:116–38.
[28] Qin C, Shi G, Tao J, et al. RCLSTMNet: a Residual-convolutional-LSTM neural network for forecasting cutterhead torque in shield machine[J]. Internat J Control, Automat Syst 2024:1–17.
[29] Deng C, Deng Z, Miao J. Semi-supervised ensemble fault diagnosis method based on adversarial decoupled auto-encoder with extremely limited labels [J]. Reliab Eng Syst Saf 2024;242:109740.
[30] Jin YR, Qin CJ, Zhang ZN, et al. A multi-scale convolutional neural network for bearing compound fault diagnosis under various noise conditions[J]. Sci China Technol Sci 2022;65(11):2551–63.
[31] Meng H, Geng M, Han T. Long short-term memory network with Bayesian optimization for health prognostics of lithium-ion batteries based on partial incremental capacity analysis[J]. Reliab Eng Syst Saf 2023;236:109288.

[32] Qin CJ, Wu RH, Huang GQ, et al. A novel LSTM-autoencoder and enhanced transformer-based detection method for shield machine cutterhead clogging[J]. Sci China Technol Sci 2023;66(2):512–27.
[33] Tang H, Liao Z, Chen P, et al. A Robust Deep Learning Network for Low-Speed Machinery Fault Diagnosis Based on Multikernel and RPCA [J]. IEEE/ASME Transact Mechatr 2022;27(3):1522–32.
[34] Fu X, Tao J, Qin C, et al. A Roller State-Based Fault Diagnosis Method for Tunnel Boring Machine Main Bearing Using Two-Stream CNN With Multichannel Detrending Inputs [J]. IEEE Trans Instrum Meas 2022;71:1–12.
[35] Zhao C, Zio E, Shen W. Domain generalization for cross-domain fault diagnosis: an application-oriented perspective and a benchmark study [J]. Reliab Eng Syst Saf 2024;245:109964.
[36] Yang B, Lei Y, Li X, et al. Targeted transfer learning through distribution barycenter medium for intelligent fault diagnosis of machines with data decentralization [J]. Expert Syst Appl 2024;244:122997.
[37] Russell M, Wang P. Maximizing model generalization for machine condition monitoring with self-supervised learning and federated learning [J]. J Manuf Syst 2023;71:274–85.
[38] Manjunatha KA, Agarwal V, Palas H. Federated-transfer learning for scalable condition-based monitoring of nuclear power plant components[C]. In: Proceedings of 16th conference of probabilistic safety assessment and management; 2022.
[39] Lin X, Jiang Q, Shen Y, et al. Multi-scale pooled convolutional domain adaptation network for intelligent diagnosis of rolling bearing under variable conditions [J]. IEEE Sens J 2023:1. –1.
[40] Zhao B, Zhang X, Zhan Z, et al. Deep multi-scale adversarial network with attention: a novel domain adaptation method for intelligent fault diagnosis [J]. J Manuf Syst 2021;59:565–76.
[41] Wu Y, Zhao R, Ma H, et al. Adversarial domain adaptation convolutional neural network for intelligent recognition of bearing faults [J]. Measurement 2022;195:111150.
[42] Feng Y, Chen J, Xie J, et al. Meta-learning as a promising approach for few-shot cross-domain fault diagnosis: algorithms, applications, and prospects [J]. Knowl Based Syst 2022;235:107646.
[43] Luo J, Shao H, Lin J, et al. Meta-learning with elastic prototypical network for fault transfer diagnosis of bearings under unstable speeds [J]. Reliab Eng Syst Saf 2024;245:110001.
[44] Feng Y, Chen J, Yang Z, et al. Similarity-based meta-learning network with adversarial domain adaptation for cross-domain fault identification [J]. Knowl Based Syst 2021;217:106829.
[45] Liu X, Chen J, Zhang K, et al. Cross-domain intelligent bearing fault diagnosis under class imbalanced samples via transfer residual network augmented with explicit weight self-assignment strategy based on meta data [J]. Knowl Based Syst 2022;251:109272.
[46] Kandukuri ST, Klausen A, Karimi HR, et al. A review of diagnostics and prognostics of low-speed machinery towards wind turbine farm-level health management [J]. Renew Sustain Energy Rev 2016;53:697–708.
[47] Ding P, Zhao X, Shao H, et al. Machinery cross domain degradation prognostics considering compound domain shifts[J]. Reliab Eng Syst Saf 2023;239:109490.
[48] Mao W, Liu Y, Ding L, et al. A new structured domain adversarial neural network for transfer fault diagnosis of rolling bearings under different working conditions [J]. IEEE Trans Instrum Meas 2020;70:1–13.
[49] Xia PC, Huang YX, Wang YX, et al. Augmentation-based discriminative meta-learning for cross-machine few-shot fault diagnosis[J]. Sci China Technol Sci 2023;66(6):1698–716.
[50] Ding P, Jia M, Zhao X. Meta deep learning based rotating machinery health prognostics toward few-shot prognostics[J]. Appl Soft Comput 2021;104:107211.
[51] Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning[J]. Adv Neural Inf Process Syst 2017:30.
[52] Agrawal N, Kumar A, Bajaj V, et al. Design of digital IIR filter: a research survey[J]. Appl Acoust 2021;172:107669.
[53] Chen S, Dong X, Peng Z, et al. Nonlinear chirp mode decomposition: a variational method[J]. IEEE Transact Signal Process 2017;65(22):6024–37.
[54] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 7132–41.
[55] Kingma DP, Adam Ba J. A Method for Stochastic Optimization [J]. In: International conference on learning representations; 2014.
[56] Chorowski J, Wang J, Zurada JM. Review and performance comparison of SVM- and ELM-based classifiers[J]. Neurocomputing 2014;128:507–16.
[57] Pan SJ, Tsang IW, Kwok JT, et al. Domain adaptation via transfer component analysis[J]. IEEE Trans Neural Netw 2010;22(2):199–210.
[58] Tzeng E., Hoffman J., Zhang N., et al. Deep domain confusion: maximizing for domain invariance[J]. arXiv preprint arXiv:1412.3474, 2014.
[59] Lu W, Liang B, Cheng Y, et al. Deep model based domain adaptation for fault diagnosis[J]. IEEE Transact Industrial Electr 2016;64(3):2296–305.
[60] Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks[J]. J Mach Learn Res 2016;17(59):1–35.