machines

MDPI

Article

Adaptive Deeping Siamese Residual Network: A Novel Model for Few-Shot Bearing Fault Diagnosis

Yonghua Jiang [1,2], Maoli Lu [3], Zhilin Dong [3,*], Zhichao Jiang [4], Weidong Jiao [3,*], Chao Tang [1], Jianfeng Sun [3] and Zhongyi Xuan [1]

[1] Xingzhi College, Zhejiang Normal University, Lanxi 321100, China; yonghua_j82@zjnu.edu.cn (Y.J.); ethan95_tang@zjnu.edu.cn (C.T.); zyx@zjnu.cn (Z.X.)
[2] Lanxi Magnesium Materials Research Institute, Lanxi 321100, China
[3] Key Laboratory of Intelligent Operation and Maintenance Technology and Equipment for Urban Rail Transit of Zhejiang Province, Zhejiang Normal University, Jinhua 321004, China; lml_study@zjnu.edu.cn (M.L.); sunjf1991@zjnu.edu.cn (J.S.)
[4] Key Laboratory of Advanced Manufacturing Technology, Beijing University of Technology, Beijing 100124, China; zhichaojiang@emails.bjut.edu.cn
* Correspondence: d18133679022@zjnu.edu.cn (Z.D.); jiaowd1970@zjnu.cn (W.J.)

Abstract: The diagnostic performance of deep learning models is heavily reliant on large volumes of labeled training data. However, in practical applications, bearing fault samples are relatively scarce, and the availability of samples for effective model training is even more limited, leading to the suboptimal performance of traditional deep learning methods in bearing fault diagnosis. To address the issue of poor performance in few-shot bearing fault diagnosis, a novel Adaptive Deep Siamese Residual Network (ADSRN) is proposed in this study. Frequency-domain information is extracted using the Fourier Transform, and training samples are randomly paired according to the matching criteria defined by the Siamese network to augment the dataset. A novel Dynamic Time Warping (DTW) technique is applied to non-linearly adjust the sequence information, allowing for the precise calculation of the optimal match between two sequences by detecting subtle differences. Additionally, inspired by the concept of dynamic soft-hard threshold matching in unsupervised learning, an innovative strategy for dynamically adjusting the adaptive threshold has been developed to enhance the generalization capability of the proposed ADSRN. Multiple few-shot fault diagnosis experiments were conducted on two bearing datasets and compared with several state-of-the-art methods. Through rigorous experimental evaluations, the effectiveness and superiority of the proposed ADSRN, as well as the advantages of DTW, were validated.

Keywords: Adaptive Deep Siamese Residual Network; Dynamic Time Warping; few-shot; bearing fault diagnosis

check for updates

Academic Editor: Hui Ma

Received: 13 January 2025
Revised: 8 February 2025
Accepted: 26 February 2025
Published: 27 February 2025

Citation: Jiang, Y.; Lu, M.; Dong, Z.; Jiang, Z.; Jiao, W.; Tang, C.; Sun, J.; Xuan, Z. Adaptive Deeping Siamese Residual Network: A Novel Model for Few-Shot Bearing Fault Diagnosis. *Machines* **2025**, *13*, 193. https://doi.org/10.3390/machines13030193

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Electric motor systems, which are extensively utilized in mechanical devices, find widespread application in various industrial applications [1,2]. The performance of motor systems is significantly influenced by rolling bearings, which play a crucial role. The health condition of these bearings is directly linked to the stability and safety of the entire mechanical system. Rolling bearings are prone to failure during operation due to factors such as material fatigue, inadequate lubrication, and external impacts. Empirical data show that nearly 50% of bearing failures are caused by insufficient lubrication, which is characterized by an increase in temperature and vibration within the bearings [3]. If

these issues are not detected and addressed in a timely manner, they could lead to severe accidents and significant economic losses [4,5]. Therefore, the importance of effective health monitoring and fault diagnosis of bearings cannot be overstated, as it is vital for the prevention of accidents, reduction of downtime, and enhancement of production efficiency [6,7].

Deep learning technology has seen rapid advancements, leading to the extensive application of deep learning-based diagnosis methods in the field of motor bearing fault diagnosis [8,9]. Effective discriminant features are automatically extracted by these methods, enabling accurate fault diagnosis without the need for expert experience knowledge. Tang et al. [10] proposed a deep learning-based approach, which utilizes a trusted multiscale quadratic attention-embedded Convolutional Neural Network for bearing fault diagnosis. Zhao et al. [11] also introduced a method based on the Joint Distributional Adaptive (JDA) and Deep Belief Network (DBN) for bearing fault diagnosis, employing a modified Sparrow Search Algorithm (CWTSSA) to optimize the DBN parameters, thereby enhancing diagnostic accuracy under varying operating conditions. Wang et al. [12] further developed a novel framework that combines a variational autoencoder (VAE) with a Convolutional Neural Network (CNN), aiming to improve the performance of rolling bearing fault diagnosis. However, the effectiveness of these methods is heavily dependent on the availability of ample high-quality labeled data. In real-world industrial settings, the typical operation of motor systems under normal conditions results in a scarcity of labeled fault data. Limited data, or few-shot conditions, can hinder the diagnostic performance of deep learning-based fault diagnosis methods, particularly in complex and extreme industrial scenarios [13]. As a result, the development of effective fault diagnosis methods under few-shot conditions has become a critical area of research focus.

In recent decades, a multitude of approaches have been proposed to tackle the challenges associated with few-shot problems, categorizing them into three main groups: transfer learning-based methods, meta-learning-based methods, and data augmentation-based methods. Transfer learning (TL) has risen as a potent machine learning technique, addressing few-shot learning issues by transferring knowledge from a source domain to a related yet distinct target domain [14]. Across various domains such as text recognition, sentiment analysis, image recognition, software defect detection, and fault diagnosis, this method has shown significant success. TL has been noted to substantially enhance model performance on target tasks, allowing for efficient and precise learning even when data availability is limited. A fusion framework (SLTL) that integrates a Self-Attentive Lightweight Convolutional Neural Network (SLCNN) with transfer learning for bearing fault diagnosis was introduced by Kim et al. [15]. This framework, designed for low-performance hardware, achieves state-of-the-art results with a minimal amount of labeled training data. Fu et al. [16] also investigated the combination of wavelet transform and Convolutional Neural Networks to enhance bearing fault diagnosis through deep transfer learning techniques. An algorithm for rolling bearing fault diagnosis, based on improved alternating transfer learning, was proposed by Peng et al. [17]. This algorithm reduces feature distribution discrepancies between the source and target domains and utilizes a SoftMax classifier at the fully connected layer for fault classification. Despite its benefits, transfer learning continues to confront challenges due to data distribution differences between source and target domains, and the acquisition of ample labeled source domain data remains a considerable obstacle in practical applications [18].

Meta-learning, when compared to transfer learning, has the advantage of rapidly adapting to new tasks by learning across multiple tasks to develop universal learning strategies. This approach reduces the dependence on large datasets and is especially effective in situations where data are scarce or tasks are rapidly changing [19]. Zeng et al. [20] proposed

a deep residual shrinkage network for motor rolling bearing fault classification, which improves model generalization during training through a meta-learning paradigm. This approach has achieved higher diagnostic accuracy and robustness under conditions of limited data and diverse scenarios. Li et al. [21] introduced a model-agnostic meta-learning fault diagnosis (MLFD) method that converts raw signals under varying operating conditions into time-frequency images, creating multi-level fault diagnosis tasks through random sampling based on meta-learning protocols. The MLFD method, by utilizing the acquired knowledge, enables rapid and accurate few-shot bearing fault diagnosis in unknown operating conditions. Su et al. [22] also proposed a novel method, Data Reconstruction Hierarchical Recurrent Meta-Learning (DRHRML), for few-shot bearing fault diagnosis across different working conditions. This method encompasses both data reconstruction and meta-learning phases, trained through a hierarchical recurrent meta-learning strategy. However, models based on meta-learning may be susceptible to overfitting due to a lack of task diversity. Additionally, the computation of high-order gradients can lead to stability issues, and challenges including high computational costs and insufficient generalization persist [23].

Compared to meta-learning methods, the ADSRN method is primarily distinguished by the following aspects: (1) ADSRN enhances the accuracy of few-shot recognition through dynamic threshold adjustment, whereas traditional meta-learning methods (such as MAML) rely on fixed task distributions; (2) it incorporates the Dynamic Time Warping (DTW) method, which, unlike time-alignment-based approaches, aligns periodic features in the frequency domain, making it more suitable for vibration signal analysis; (3) by introducing residual modules, ADSRN alleviates the gradient vanishing problem in few-shot scenarios, while most meta-learning models have not optimized network depth for signal feature extraction.

Data augmentation is recognized as an effective strategy for mitigating the challenges associated with small dataset sizes. It simulates a variety of operating conditions and fault patterns through operations such as image rotation, scaling, and cropping, which leads to the generation of new vibration data. Data augmentation, when compared to transfer learning and meta-learning, circumvents the pitfalls of negative transfer and the need for a substantial amount of labeled data. Furthermore, it mitigates the risk of overfitting and bolsters the model's generalization capability [24]. A novel algorithm based on Conditional Generative Adversarial Networks (CGAN) was introduced by Ahang et al. [25], which is trained on normal and faulty data under real fault conditions to generate faulty data from normal data in target conditions. The use of Generative Adversarial Networks (GAN) for data augmentation was employed by Shao et al. [26] to enhance machine fault diagnosis performance. Hou et al. [27] proposed a new network architecture known as the Siamese Multiscale Residual Feature Fusion Network (SMSRFFN). This architecture initially pairs training samples based on matching rules to augment the sample size and subsequently utilizes a Multiscale Residual Feature Extraction Network (MSRFEN) for extracting fault features at various scales, which also expedites network convergence.

Although the aforementioned methods have demonstrated their effectiveness in few-shot fault diagnosis, they rely heavily on the design of deep network structures to extract sufficient fault features from limited labeled data. Under few-shot conditions, deep networks, which require a large number of training parameters, are prone to overfitting and gradient explosion issues [28]. These problems can significantly impair the model's generalization capability. To address these challenges, this study proposes an Adaptive Deep Siamese Residual Network (ADSRN) for diagnosing motor bearing faults under few-shot conditions. The proposed ADSRN includes the following key components:

(1) An Adaptive Siamese Network is designed to fully leverage the discriminative features within limited labeled fault data. A method for constructing sample pairs is also

developed. This method creates cross samples using the training and testing datasets, thereby enhancing the effectiveness of training with limited labeled samples.

(2) A Residual Network Module is embedded and integrated into the Siamese network to enhance feature representation. This integration effectively prevents overfitting and gradient explosion issues that may arise from scarce samples, thus improving the model's performance in diagnosing bearing faults under few-shot conditions.

(3) Adaptive thresholds are introduced to ensure the preservation of key features in the signal. Additionally, Dynamic Time Warping (DTW) is incorporated as a measurement module. This module not only aligns periodic features in the frequency domain but also more accurately captures structural changes in the signal, thereby accelerating model convergence. The proposed method exhibits strong robustness against outliers by seeking the best overall match rather than comparing points one by one.

Overall, the Adaptive Deep Siamese Residual Network (ADSRN) proposed in this paper is capable of efficiently diagnosing bearing faults under few-shot conditions without relying on a large amount of auxiliary data. To evaluate the effectiveness of the proposed ADSRN, the study employed two challenging public datasets to quantitatively assess its diagnostic performance and compare it with several state-of-the-art intelligent diagnostic methods, including ResNet [29], WDCNN [30], and CNN-LSTM [31]. To enhance the credibility of the experimental results, the study conducted 5 independent repeated experiments using 4 small sample training datasets (with 10, 15, 20, and 30 samples, respectively) to calculate the standard deviation of accuracy. Additionally, confusion matrices and feature visualization plots were generated after classification. This research lays the foundation for bearing fault diagnosis methods under few-shot conditions.

The remainder of this paper is organized as follows. Section 2 describes the principle of the basic model, the fault diagnosis framework with the embedded DTW, and the overall ADSRN model. Section 3 validates the performance of the proposed method using two case studies and several comparative methods. Section 4 concludes the paper and details the future work.

2. Proposed ADSRN Diagnostic Model Framework

In this section, the construction of sample pair labels is detailed, where labels are formed based on criteria defined within the ADSRN diagnostic model. Additionally, the embedding of DTW within the ADSRN diagnostic model is described, which enhances the model's ability to handle sequence data and align periodic features in the frequency domain.

2.1. The Construction of Sample-Label Pairs

Constructing positive and negative sample pairs effectively is crucial for subsequent model training. Initially, vibration sample data are processed using the Fast Fourier Transform (FFT) to obtain their frequency-domain information. Generally, it is sufficient to consider only the magnitude spectrum (absolute values), especially when it comes to Dynamic Time Warping (DTW). In our experiments, emphasis was placed on the fault characteristic frequencies of the bearing signals, with less focus given to the phase information of the signals. If more precise signal reconstruction or capturing of temporal structures is required, the complex form, including phase information, can be utilized. The Fast Fourier Transform (FFT) is illustrated as shown in Equation (1):

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-i2\pi kn/N}, k = 0, 1, \ldots i, \ldots j, N-1 \quad (1)$$

where x_n represents the time domain signal, X_k represents the frequency-domain signal, and N represents the signal length.

A pairing strategy is then applied to create sample pairs from both the same and different fault categories. This method enriches the information extracted from a few-shot dataset, laying a solid foundation for comparing features between different and similar categories during the feature extraction process. The similarity between two feature vectors obtained from the extraction process is measured using DTW. Samples from the same category are labeled as similar, while those from different categories are labeled as distinct. The process of pairing and labeling samples during data processing is as follows:

$$(x_i, x_j, Y) = P(x_i, x_j) \tag{2}$$

The total number of training samples is N, with each sample represented by x, i and j representing the indices of the samples, which are paired sequentially. After pairing, there are C_N^2 pairs of samples, and each sample pair can be represented by (x_i, x_j), where $i, j = 1, 2, \ldots, N$, and $i \neq j$. If x_i and x_j are fault sample pairs of the same type, then assign $Y = 0$; otherwise, $Y = 1$, where $p(\cdot)$ represents the marking process of the sample pairs. After the samples are paired, all sample pairs are sequentially input into the ADSRN for feature extraction. This process can be represented as:

$$F_1 = [BN(\mathrm{Re}lu(Conv(x, w)))]_n \tag{3}$$

where, x represents the input sample, w represents the weight parameters of the ADSRN, $Conv$ represents the convolution operation. ReLu refers to the activation function, which is used for the nonlinear transformation of the neuron outputs in the neural network. refers to batch normalization; indicates that the same operation is repeated times.

The construction process of the sample pairs is shown in Figure 1.

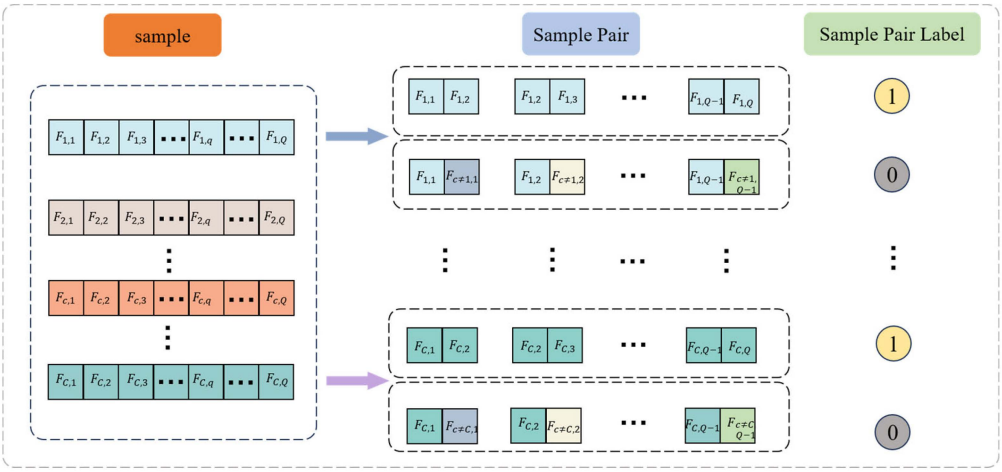


Figure 1. Pairing strategy for few-shot datasets.

The sample x_i is brought into F_1 for feature extraction, as shown below:

$$x_i = F_1(x_i, w)_2 \tag{4}$$

In the above Equation (4), "2" indicates that the operation F_1 is performed twice, w represents the weight parameters.

After feature extraction, a threshold is determined through the shrinkage module operation to minimize noise interference. The expression for the threshold determination process is

$$F_2 = FC(\mathrm{Re}LU(BN(FC(x, w)))) \tag{5}$$

$$\alpha_i = \frac{1}{1 + e^{-F_2(average|x_i|)}} \tag{6}$$

$$\tau_i = \alpha_i * average|x_i| \tag{7}$$

where FC refers to the fully connected layer in the neural network, α_i is the output of the activation function sigmoid and τ_i is the computed threshold.

The above is the threshold calculation formula. In this network, a shrinkage operation is used to contract the output of the residual module to reduce the impact of noise. and ADSRN is also used for feature extraction of sample x_i here. The expressions for the contraction operation (soft threshold function) and feature extraction are

$$y_i = \begin{cases} x_i - \tau_i & x_i > \tau_i \\ 0 & -\tau_i > x_L > \tau_i \\ x_i + \tau_i & x_i < -\tau_i \end{cases} \tag{8}$$

Here, y_i is the output after the threshold calculation.

$$H(x_i) = x_i + y_i \tag{9}$$

Among them, $H(x_i)$ denotes the output feature of x_i after feature extraction from one side of the ADSRN. Similarly, the output feature of x_j after feature extraction from the other side of the ADSRN can be defined as follows:

$$H(x_j) = x_j + y_j \tag{10}$$

The output feature vectors from both sides of the ADSRN method are compared using Dynamic Time Warping (DTW), which measures the similarity between two sequences that may differ in periodic features. First, a cost matrix C is defined.

The output feature vectors from both sides of the ADSRN method are compared using DTW, which measures the similarity between two sequences that may differ in periodic features. Therefore, the cost matrix C is defined and initialized as follows:

$$C = \left[C_{i,j}\right]_{n \times m} \tag{11}$$

$$\begin{cases} C_{0,0} = 0 \\ C_{i,0} = \infty \quad for\ i = 0 \\ C_{0,j} = \infty \quad for\ j = 0 \end{cases} \tag{12}$$

where n is the length of the sequence $H(x_i)$, and m is the length of the sequence $H(x_j)$. Each element $C_{i,j}$ in the matrix C represents the distance between the i-th element of $H(x_i)$ and the j-th element of $H(x_j)$. Then, the cost matrix is initialized, and its first row and first column are initialized to infinity, except for the starting point $C_{0,0}$, which is initialized to 0.

The cost matrix is then filled in, where d is the distance function, using the Euclidean distance.

$$C_{i,j} = d\left(H(x_i)_i, H(x_j)_j\right) + \min\left(C_{i-1,j}, C_{i,j-1}, C_{i-1,j-1}\right) \tag{13}$$

Finally, the DTW distance is calculated. The DTW distance is the value in the bottom right corner of the cost matrix C, representing the optimal match distance between the two sequences.

$$d_{DTW}\left(H(x_i), H(x_j)\right) = C_{n,m} \tag{14}$$

In this way, DTW can quantify the similarity between two sequences, even if they are not perfectly aligned in terms of periodic features. The smaller the DTW distance, the more similar the two sequences are. Conversely, the larger the DTW distance, the less similar the two sequences are. The contrastive loss function L_1 is represented as follows:

$$L_1 = \frac{1}{2N}\sum_{n=1}^{N} Yd^2 + (1-Y)\max(\text{margin} - d, 0)^2 \tag{15}$$

where N represents the number of samples, and Y indicates whether the input pair has the same type of fault. The expression $Y = 1$ signifies that the input pair is similar, while $Y = 0$ signifies dissimilarity. The margin represents the threshold, with a default value set to 2.

In order to implement fault diagnosis, the extracted features $H(x_i)$ and $H(x_j)$ are fed back into the SoftMax classifier. The output y_n of the SoftMax classifier is defined as:

$$y_n = \frac{e^{H_n(x_i)}}{\sum_{n=1}^{Nclass} e^{H_n(x_i)}} \tag{16}$$

where n represents the n-th class, y_n is the predicted probability that the sample belongs to the n-th class, $Nclass$ is the number of classes, ensuring that $\sum y_n = 1$. In the above formula, $Nclass = 10$. After predicting the probabilities of belonging to the 10 classes, the class with the maximum predicted probability value is selected as the result of the fault diagnosis. The cross-entropy loss function is commonly used as the objective function for multi-class classification problems. In this paper, this function serves as the loss function for fault diagnosis and is represented as

$$L_2 = -\sum_{n=1}^{Nclass} t_n \log y_n \tag{17}$$

where t_n represents the actual fault type. When the sample belongs to the n-th class, $t_n = 1$; otherwise, $t_n = 0$.

2.2. ADSRN Diagnostic Model Embedded with DTW

The first core component of the ADSRN model proposed in this study is a unique signal feature extraction module called the FC (Fully Connected) Block. This FC Block consists of two fully connected layers (Linear Layers), a batch normalization layer (BatchNorm1d), a ReLU activation function, and a Sigmoid activation function arranged sequentially.

For dual-channel input signals, the module first computes the modulus of each channel's signals and applies average pooling, reducing the signal length from 1024 to a single scalar value. The compressed features are then transformed using fully connected layers and expanded back to the original number of channels. Next, the difference between the original signals and the expanded features is calculated, with any negative values in the difference set to zero. Finally, the non-negative differences are multiplied element-wise with the original signals to produce the final output features. After undergoing absolute value processing, pooling, linear transformation, and sign correction within the FC Block, the input data are output as a tensor with the same shape as the original input. This signal processing module effectively reduces noise while extracting essential features, enabling strong learning capabilities for bearing fault diagnosis in few-shot scenarios. It facilitates better learning and recognition of key information within the signals. Average pooling,

linear transformation, feature reconstruction, and reconstructed signal adjustment are as shown in Equations (18)–(20):

$$z_c = \frac{1}{T}\sum_{t=1}^{T}|x_c(t)|(c = 1, 2, \ldots, C) \tag{18}$$

where $x_c(t)$ is the value of the c-th channel of the input signal at time step t, T is the total number of time steps of the signal, which is 1024 here. $|x_c(t)|$ is the modulus value of the signal at each time step. z_c is the scalar feature after compression of the c-th channel, representing the global intensity of the signal of that channel.

$$\hat{z}_c = \mathrm{Re}LU(W_2 \cdot Sigmoid(W_1 \cdot z_c + b_1) + b_2) \tag{19}$$

where W_1, b_1 is the weight matrix and bias of the first fully connected layer, which maps the scalar z_c to the hidden dimension. *Sigmoid* is the activation function that restricts the output to the $[0, 1]$ range, controlling the flow of information. W_2, b_2 is the weight matrix and bias of the second fully connected layer, which expands the hidden features back to the original channel dimension. Re*LU* is the activation function that enhances the nonlinear expression ability and retains positive features. $\hat{z}_c$ is the reconstructed channel feature vector (with the same dimension as the original signal channel).

$$\widetilde{x}_c(t) = x_c(t) \cdot \max(0, \hat{z}_c - x_c(t)) \tag{20}$$

where $\hat{z}_c$ is the reconstructed channel feature, $\max(0, \hat{z}_c - x_c(t))$ is the dynamic threshold operation, which only retains part of the signal $\hat{z}_c > x_c(t)$. $\widetilde{x}_c(t)$ is the adjusted output signal.

The second core component of the ADSRN model proposed in this study is a fundamental convolutional block, referred to as the Basic Block. This Basic Block consists of two 1D convolutional layers (Conv1d), two batch normalization layers (BatchNorm1d), a shortcut connection (residual connection), and an FC Block module. In the Basic Block, the input data are processed sequentially through Conv1d, BatchNorm1d, activation functions, and the FC Block. The skip connection combines the original input features with the transformed features, preserving some of the original characteristics while incorporating new feature representations. The final output is a tensor with a shape similar to the input, ensuring that both the original and new features are retained. The network structure is illustrated in Figure 2.

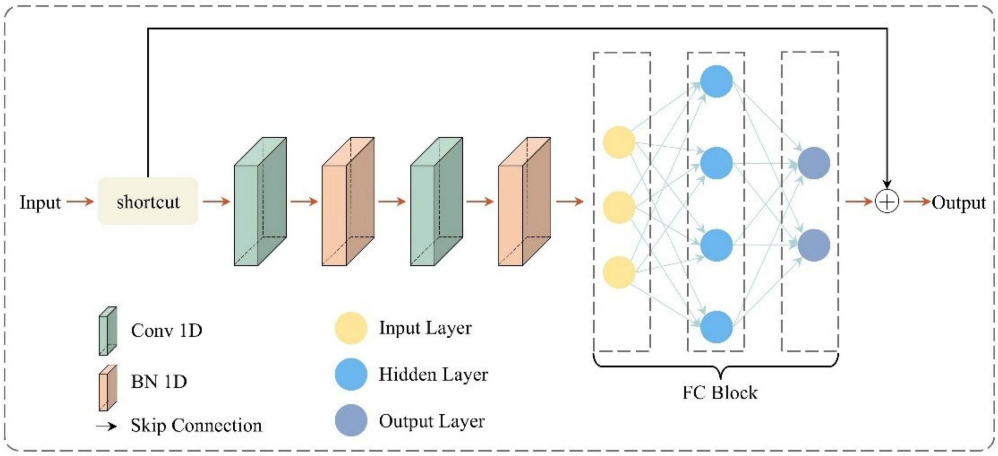


Figure 2. Basic Block network structure in ADSRN.

The detailed parameters of the Basic Block module are shown in the Table 1.

Table 1. The detailed parameters of the Basic Block module.

Layer Type	Kernel Size	Stride	Padding	Feature Channel
Conv1d	3×3	1	1	Adaptive
BN1d	/	/	/	Adaptive
Conv1d	3×3	1	1	Adaptive
BN1d	/	/	/	Adaptive
FC Block	/	/	/	Adaptive

The Basic Block employs a classical residual connection structure, allowing the model to learn the residual between the input and output. This effectively addresses the vanishing gradient problem in deep networks during feature extraction, ensuring smooth gradient flow, facilitating deeper feature learning, and helping to prevent overfitting. Additionally, the integration of a custom FC Block extends the traditional residual block's functionality. The combination of the FC Block and the residual module equips the model with a specialized feature extraction mechanism. Together, they synergistically enhance the model's ability to process and optimize signal features, improving its capacity to represent signal characteristics more effectively.

The construction of the ADSRN model begins with fundamental modules, which serve as the cornerstone for complex network architectures. Utilizing these modules, we have developed a model that consists of two key components: the formation of sample-label pairs and the ADSRN diagnostic model integrated with DTW. The visualization of the model is presented in Figure 3, while the detailed structural parameters are listed in Table 2. By integrating DTW into the ADSRN model, we achieved an efficient diagnosis of few-shot bearing fault data. It has been found during the research that applying DTW in the frequency domain can capture the periodic characteristics of signals and accelerate the convergence speed of the model. This approach of embedding DTW not only enhances the model's robustness to sequential data but also boosts its analytical capabilities in the frequency domain. As a result, the model can more accurately capture the frequency-domain features of bearing faults, thereby improving the accuracy and reliability of fault diagnosis. Although DTW is typically used in the time domain, combining it with the Fast Fourier Transform (FFT) can enhance its effectiveness. FFT transforms the signal into the frequency domain to extract frequency features, which can be used as inputs for DTW to better capture the periodicity or frequency variations of the signal.

The overall framework of our proposed ADSRN model is shown in Figure 4. The model is trained and optimized using paired samples, ultimately mapping the extracted features into a 100-dimensional space through a linear classifier, followed by a Sigmoid activation function to produce the final classification output. The proposed ADSRN network, based on a Siamese architecture, combines the FC Block and Basic Block modules to determine the similarities and differences between two input signals. Each input signal is processed by identical sub-networks that output feature vectors, which are then compared. By leveraging the strengths of both residual and fully connected blocks, the model enhances its feature learning capabilities, particularly in few-shot scenarios, mitigating the challenges posed by limited data during feature extraction. Moreover, the model's design effectively addresses the vanishing gradient problem in deep networks, using convolutional and classifier modules to achieve deep feature extraction and accurate classification of input signals.

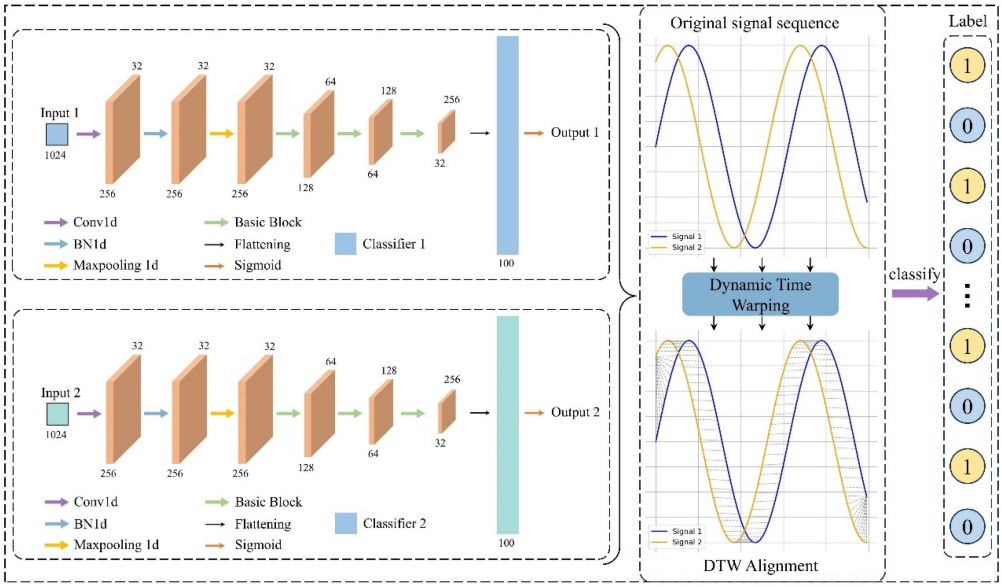


Figure 3. Visualization results of the ADSRN model framework.

Table 2. The detailed structural parameters of the ADSRN model.

Layer Type	Kernel Size/Stride	Kernel Number	Padding	Input/Output
Conv 1d	7 × 7/4	32	3	2/32
BN 1d	/	/	/	32
Max Pooling	3 × 3/1	/	1	32
Basic Block	/2	Number: 2	/	32/64
Basic Block	/2	Number: 2	/	64/128
Basic Block	/2	Number: 2	/	128/256
Conv 1d	7 × 7/4	32	3	256/32
BN 1d	/	/	/	32
Max Pooling	3 × 3/1	/	1	32
Basic Block	/2	Number: 2	/	32/64
Basic Block	/2	Number: 2	/	64/128
Basic Block	/2	Number: 2	/	128/256
Classifier Linear	/	/	/	256/100
Classifier Linear	/	/	/	100/100

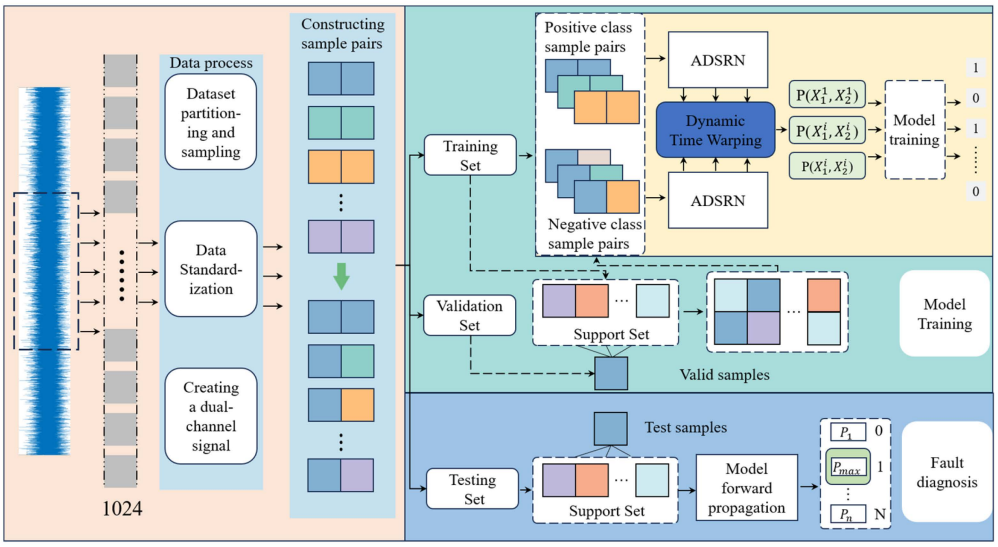


Figure 4. The comprehensive framework of the proposed ADSRN.

3. Case Study

To quantitatively verify the effectiveness and superiority of the proposed ADSRN, two challenging bearing cases under different few-shot scenarios are used to test in this section. Besides, several state-of-the-art deep learning methods are compared, including ResNet, WDCNN, and CNN-LSTM.

3.1. Case 1

3.1.1. Description of the Dataset (Case Western Reserve University)

The Case Western Reserve University (CWRU) bearing dataset is studied in this case. The testing bearings used are SKF 6205 models, with three fault states and a normal state. The fault states consist of inner race fault, outer race fault, and roller fault. Further, each fault state contains three crack sizes, namely 0.1778 mm, 0.3556 mm, and 0.5334 mm. Therefore, a total of nine fault states and a normal state. The experimental setup for the CWRU bearing tests is shown in Figure 5. It is primarily composed of the following five components: Fan end bearing, Electric motor, Drive end bearing, Torque transducer and encoder, Dynamometer. The speeds are set to 1797 rpm, respectively, with a sampling frequency of 12 kHz. For each normal state, 80 samples were selected for testing, and the performance of four neural networks under few-shot conditions was compared.

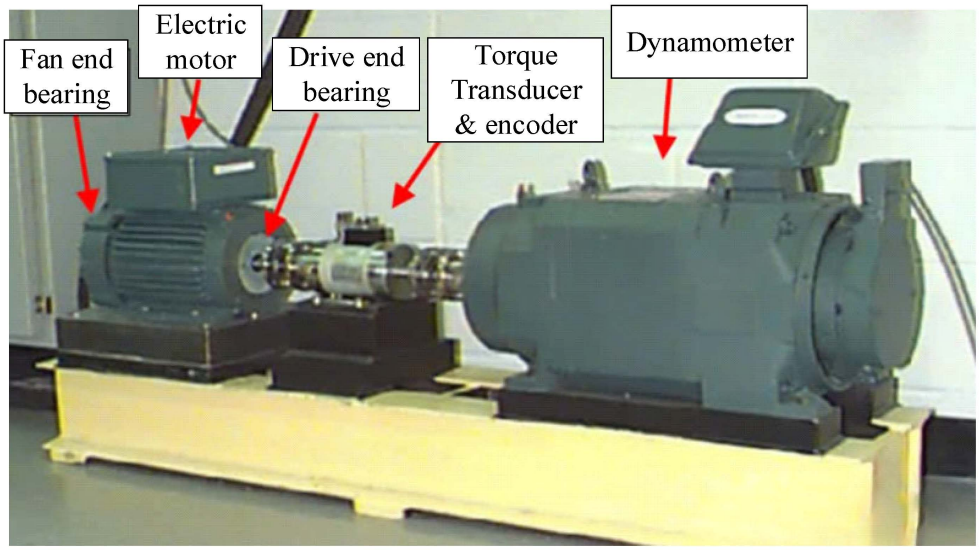


Figure 5. The bearing data test rig of CWRU.

For the training process, few-shot training datasets are formed by randomly selecting combinations of 10, 15, 20, and 30 samples. Each sample consists of 1024 sampling points. The detailed information about the experimental dataset is presented in Table 3.

The spectrum of the original vibration signal, obtained throughthe Fast Fourier Transform (FFT), can reveal more fault-related information. We discuss the data preprocessing and hyperparameter settings. Specifically, for the FFT, the settings include an overlap rate of 0, a sample pairing strategy (with a random seed set), and a sample length of 1024. During the experimental stage, the hyperparameter settings are as follows: an initial learning rate of 1×10^{-4}, a batch size of 32, and 50 training epochs. This transformation process is beneficial for subsequent operations, such as convolution in neural network models. To mitigate the effects of random errors, each experiment is conducted five times.

Table 3. CWRU experimental dataset.

Label	Fault Size (mm)	Fault Location	Number of Training Samples in Four Few-Shot Datasets
0	0.1778	Ball	10, 15, 20, 30
1	0.3556	Ball	10, 15, 20, 30
2	0.5334	Ball	10, 15, 20, 30
3	0.1778	Inner	10, 15, 20, 30
4	0.3556	Inner	10, 15, 20, 30
5	0.5334	Inner	10, 15, 20, 30
6	0.1778	Outer	10, 15, 20, 30
7	0.3556	Outer	10, 15, 20, 30
8	0.5334	Outer	10, 15, 20, 30
9	-	Normal	10, 15, 20, 30

3.1.2. Experiment Analysis (Case 1)

Figures 6 and 7, along with Table 4, present the diagnostic outcomes derived from four training datasets. During the training phase, samples are randomly paired and fed into the model. The comparative results displayed in Figures 6 and 7, and Table 4 indicate that the average accuracy of the proposed ADSRN method across the four low-sample datasets is shown in Table 4, with values of 97.15%, 98.45%, 98.90%, and 99.60%, and standard deviations of 0.34, 0.75, 0.64, and 0.35, respectively. This level of accuracy significantly surpasses that of ResNet (84.25 $\pm$ 1.55%), WDCNN (89.24 $\pm$ 2.54%), and CNN-LSTM (98.5 $\pm$ 1.32%). Furthermore, when only ten training samples are available for each normal state, the average classification accuracy of the ADSRN method is 23.78% higher than that of ResNet, 33.29% higher than that of WDCNN, and 8.77% higher than that of CNN-LSTM. Similarly, for other few-shot training datasets, ADSRN outperforms other comparative methods. The results under the ADSRN method shown in Figure 7 are clustered, maintaining a high range of accuracy, and demonstrating its good stability and accuracy. However, in the first three low-sample datasets (10, 15, 20), the fault diagnosis accuracy of the WDCNN method is consistently lower than that of the other methods. This further proves that the proposed ADSRN has a significant advantage in bearing fault diagnosis.

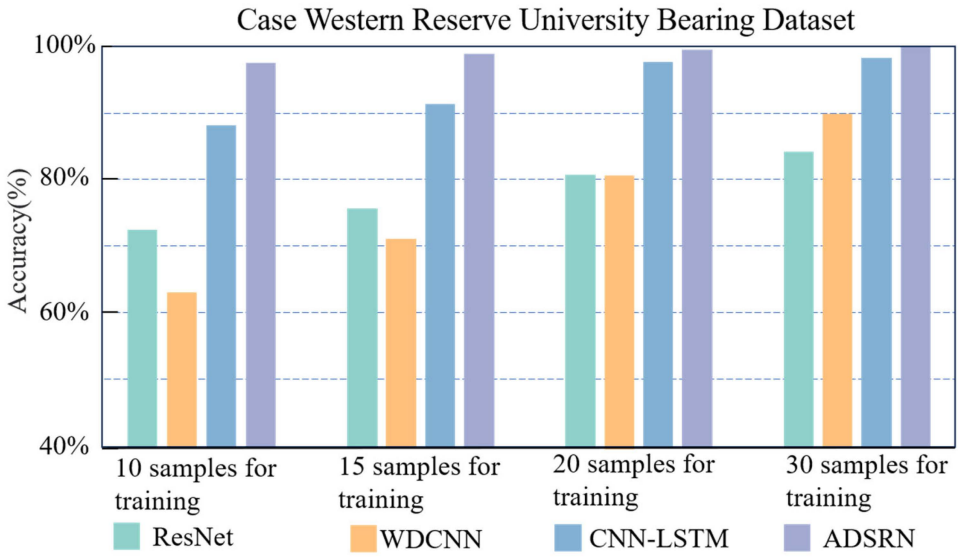


Figure 6. Diagnosis accuracy for different models under four few-shot datasets (Case 1).

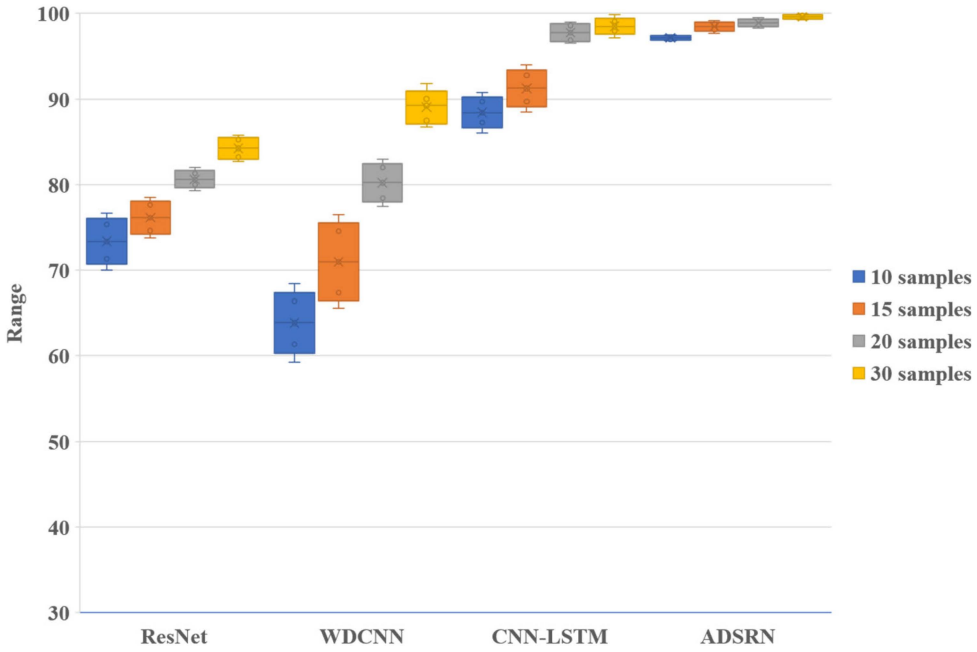


Figure 7. Box plot of fault diagnostic accuracy distribution of different models under the four types of small sample datasets (Case 1).

Table 4. Experimental results using four few-shot datasets (Case 1).

Model	10 Training Samples	15 Training Samples	20 Training Samples	30 Training Samples
	Accuracy (%)	Accuracy (%)	Accuracy (%)	Accuracy (%)
ResNet	73.37 ± 3.34	76.13 ± 2.35	80.63 ± 1.35	84.25 ± 1.55
WDCNN	63.86 ± 4.59	71.00 ± 5.47	80.22 ± 2.75	89.24 ± 2.54
CNN-LSTM	88.38 ± 2.35	91.25 ± 2.77	97.75 ± 1.25	98.5 ± 1.32
ADSRN	**97.15 ± 0.34**	**98.45 ± 0.75**	**98.90 ± 0.64**	**99.60 ± 0.35**

To further visualize the diagnosis results and explore the diagnostic effectiveness for each normal state, the confusion matric of different diagnosis models is shown in Figure 8 where the training samples are set to 20 for each normal state. It can be seen that the proposed ADSRN model only misclassifies one testing sample out of a total of 800 samples. For each normal state, the ADSRN can achieve nearly perfect prediction, which significantly outperforms other deep learning methods.

In order to more intuitively illustrate the superior feature extraction capability of ADSRN, T-SNE technology was utilized to visualize the feature extraction results in a two-dimensional space. Different colors were assigned to represent different normal states. As depicted in Figure 9, a higher degree of clustering within the same color distribution and a lower degree of overlap with other color distributions suggest that better feature extraction has been achieved for a specific category by the model. Figure 9a–d displays the two-dimensional visualizations of the feature extraction results obtained using a Residual Network (ResNet), Wide and Deep Convolutional Neural Network (WDCNN), Convolutional Long Short-Term Memory Network (CNN-LSTM), and ADSRN.

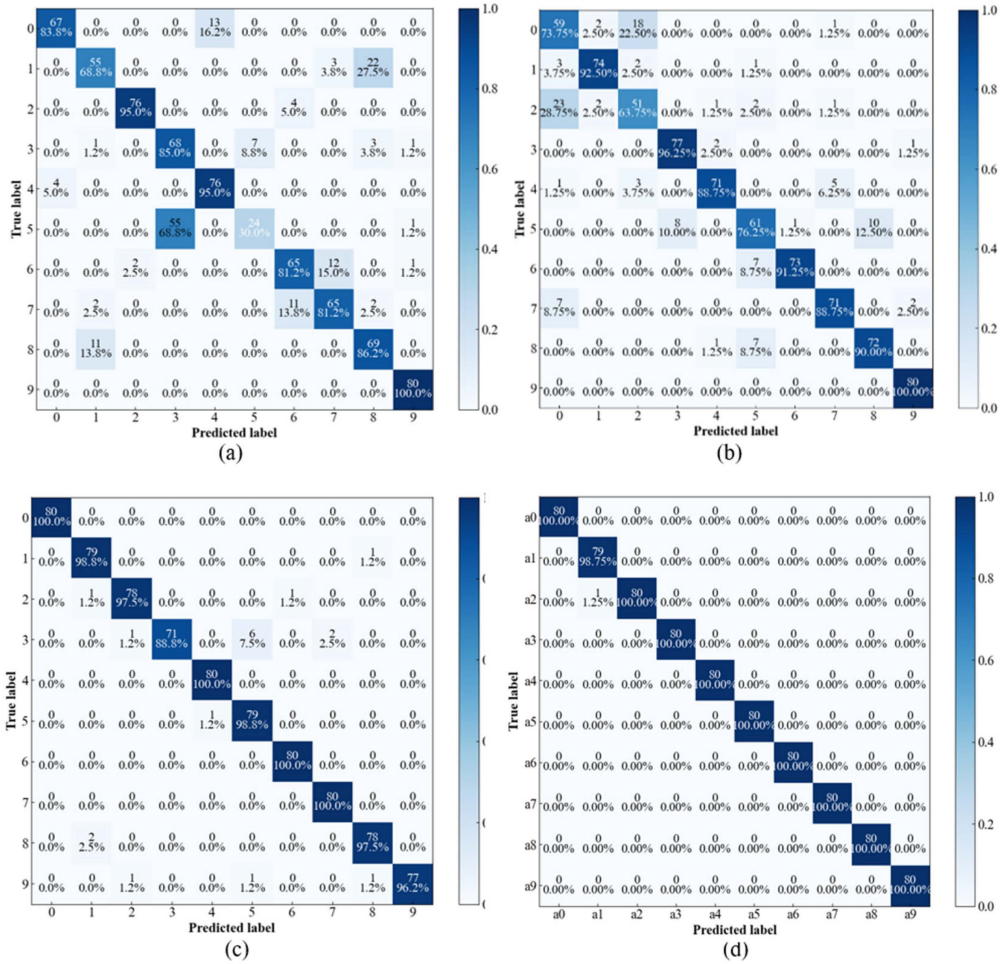


Figure 8. Confusion matrix for various models when the number of training samples of each type of fault is 20 (Case 1). (**a**) ResNet; (**b**) WDCNN; (**c**) CNN-LSTM; (**d**) ADSRN.

Unclear boundaries between categories 4 and 8 were observed in the proposed model, yet its overall classification accuracy was higher than that of the other comparative methods. The unclear boundaries may have resulted from the similarity in features between categories 4 and 8. In Figure 9b, unclear boundaries between categories 3 and 5 were observed, and a few samples of category 6 were misclassified into category 3. Unclear boundaries were also observed between categories 2, 7, 0, and 4. A few samples of category 9 were misclassified into category 1. Confusion between categories 5 and 3 occurred in Figure 9a, and the majority of category 8 were misclassified into category 1, with unclear boundaries. Additionally, a few samples of category 0 were misclassified into category 4, and a few samples of category 2 were misclassified into category 9. In Figure 9c, a few samples of category 2 were misclassified into category 7, a few samples of category 7 were misclassified into category 6, and a few samples of category 6 were misclassified into category 9.

In summary, the visualizations in Figure 9 clearly demonstrate that, compared to the Residual Network (ResNet), Wide and Deep Convolutional Neural Network (WDCNN), and Convolutional Long Short-Term Memory Network (CNN-LSTM), the ADSRN model possesses a superior feature extraction capability.

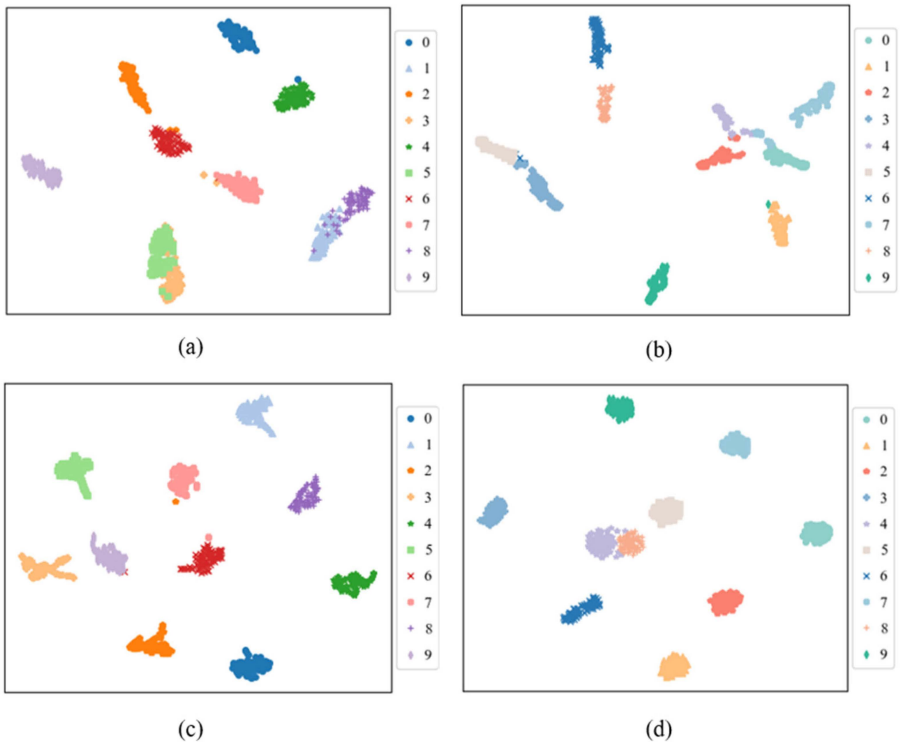


Figure 9. Feature visualization using various models when the number of training samples of each type of fault is 20 (Case 1). (**a**) ResNet; (**b**) WDCNN; (**c**) CNN-LSTM; (**d**) ADSRN.

3.2. Case 2

3.2.1. Description of the Dataset (Southeast University)

In this case, the effectiveness of the ADSRN method is further validated using the bearing dataset from Southeast University. This dataset includes two operating conditions for each normal state: one with a rotation speed of 1200 rpm and a load of 0 Nm, and the other with a rotation speed of 1800 rpm and a load of 7.32 Nm. A total of four fault types including Ball (rolling element fault), Comb (compound fault, combining rolling element, outer race, and inner race faults), Inner (inner race fault), Outer (outer race fault), and a normal state are considered. The detailed settings of the dataset are shown in Table 5. The data are collected using a Drive System Dynamics Simulator (DSDS), with signals recorded from eight channels for each normal state, the experimental bench is shown in Figure 10, which consists of a motor controller, a motor, a planetary gearbox, a parallel gearbox, a brake, and a brake controller. To gain a clearer understanding of the test bench structure, one can refer to the simplified structural diagram of the test bench, as shown in Figure 11.

Table 5. Southeast University experimental dataset.

Label	Speed (rpm)	Fault Location	Number of Training Samples in Four Few-Shot Datasets
0	1200	Ball	10, 15, 20, 30
1	1800		10, 15, 20, 30
2	1200	Comb	10, 15, 20, 30
3	1800		10, 15, 20, 30
4	1200	Inner	10, 15, 20, 30
5	1800		10, 15, 20, 30
6	1200	Outer	10, 15, 20, 30
7	1800		10, 15, 20, 30
8	1200	Normal	10, 15, 20, 30
9	1800		10, 15, 20, 30

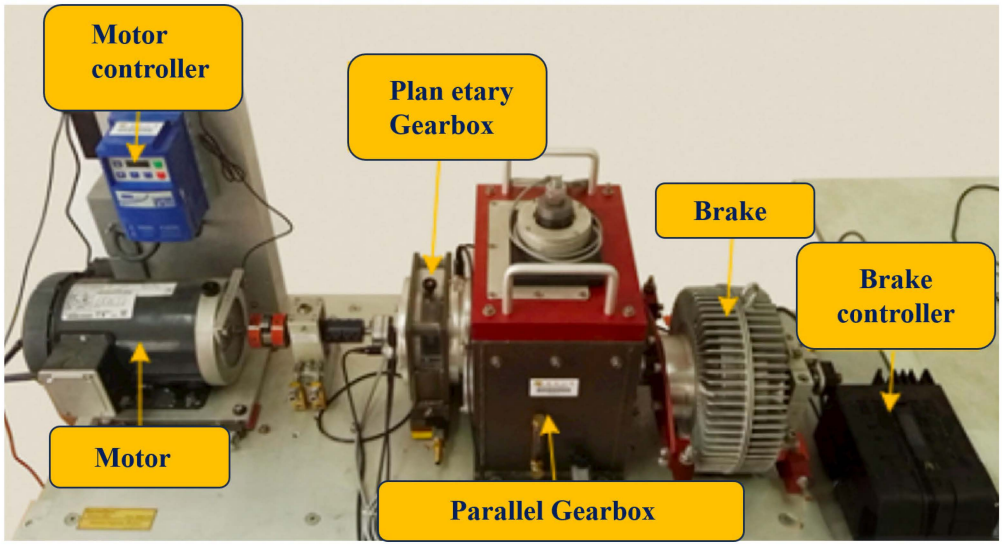


Figure 10. The bearing data test rig of Southeast University.

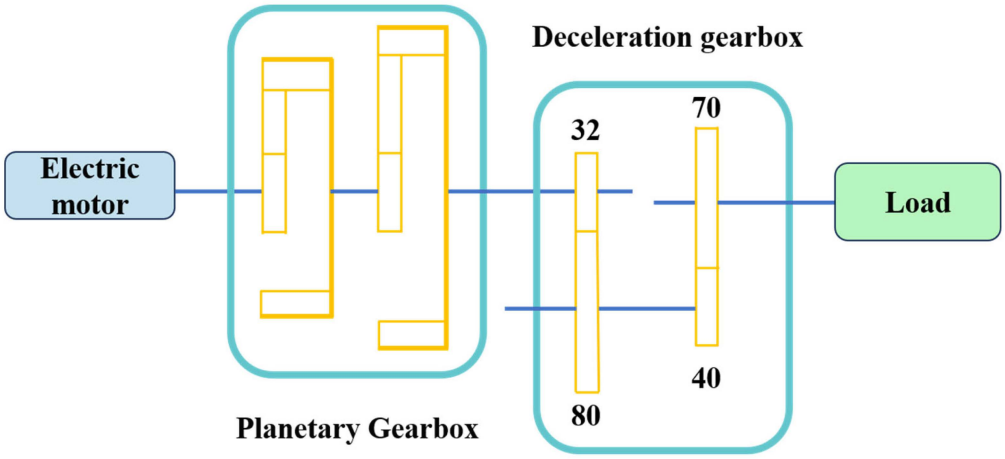


Figure 11. Schematic diagram of the bearing test bench at Southeast University.

In this case, during signal preprocessing, the vibration signals of the planetary gearbox in the x and y directions are selected to form dual-channel signal samples for the experiment. The performance of the proposed model is compared with models in Case 1, including ResNet, WDCNN, and CNN-LSTM. To ensure robustness and reduce the chance of random errors, each experiment is repeated five times.

3.2.2. Experiment Analysis (Case 2)

The experimental results under the four few-shot scenarios are presented in Figures 12 and 13, and Table 6. It can be observed that under each few-shot condition, the proposed ADSRN method achieves the highest average classification accuracy. Particularly, when the number of training samples for each normal state is set to 10, the average diagnostic accuracy of the proposed method reaches 96.61%, which is significantly higher than that of ResNet (75.4%), WDCNN (46.42%), and CNN-LSTM (83.55%). As shown in Figure 13, under the four low-sample datasets, the test results under the ADSRN method are clustered and maintain a high range of accuracy, indicating that it maintains good stability and accuracy. The proposed ADSRN also demonstrates a significant advantage in classification accuracy. This further proves that the proposed ADSRN method possesses excellent generalization performance and strong robustness in intelligent bearing fault diagnosis, even with a limited number of training samples.

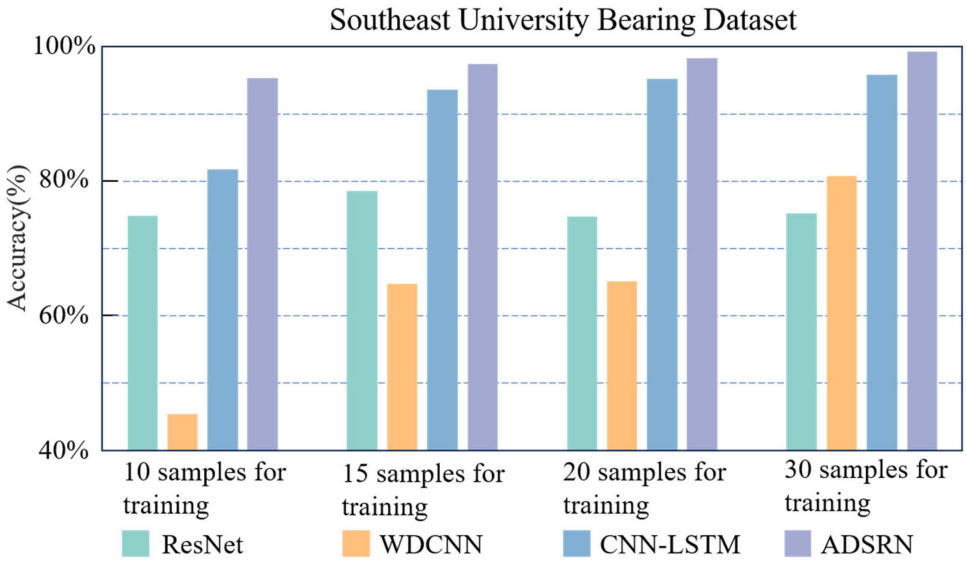


Figure 12. Diagnosis accuracy for different models under four few-shot datasets (Case 2).

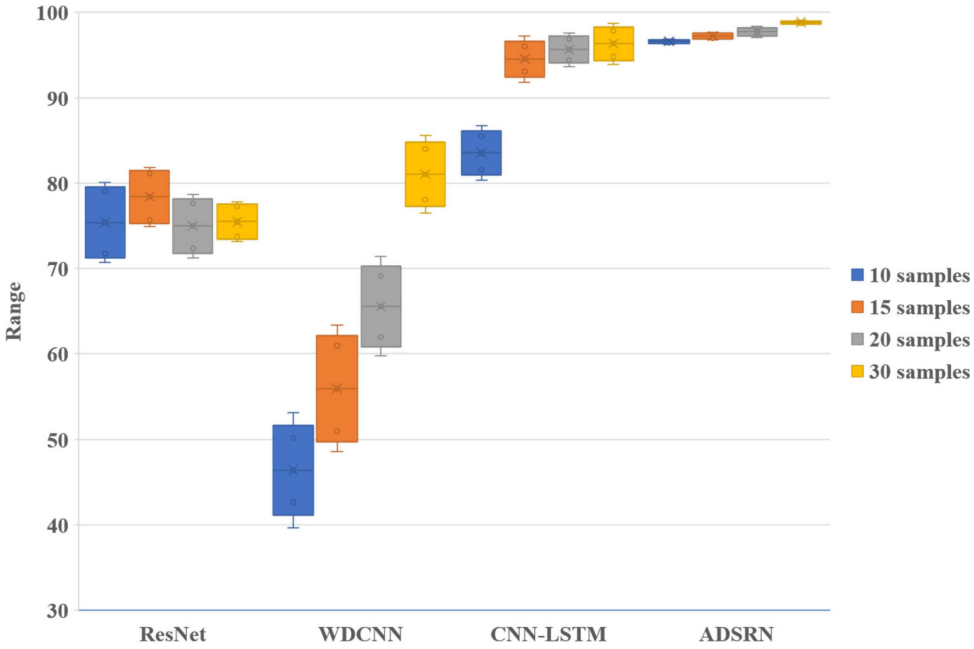


Figure 13. Box plot of fault diagnostic accuracy distribution of different models under the four types of small sample datasets (Case 2).

Table 6. Experimental results using four few-shot datasets (Case 2).

Model	10 Training Samples	15 Training Samples	20 Training Samples	30 Training Samples
	Accuracy (%)	Accuracy (%)	Accuracy (%)	Accuracy (%)
ResNet	75.40 ± 4.72	78.40 ± 3.46	75.00 ± 3.72	75.50 ± 2.32
WDCNN	46.42 ± 6.75	55.96 ± 7.42	65.56 ± 5.82	81.06 ± 4.55
CNN-LSTM	83.55 ± 3.21	94.55 ± 2.71	95.65 ± 1.98	96.35 ± 2.41
ADSRN	**96.61 ± 0.32**	**97.25 ± 0.42**	**97.72 ± 0.62**	**98.85 ± 0.23**

To further visualize the diagnosis result of each normal state, Figure 14 shows the confusion matrix of all methods, in which the training number of each normal state is set to

30. It can be observed that the recognition accuracy of each normal state of the proposed ADSRN can reach above 90%. Even for a compound fault (class 2 and class 3), there are only 4 testing samples misclassified by our method. As for comparison methods, there are a lot of misclassified samples. The overall diagnosis accuracy of ResNet (75.50 $\pm$ 2.32%), WDCNN (81.06 $\pm$ 4.55%), and CNN-LSTM (96.35 $\pm$ 2.41%) is lower than that of the ADSRN model (98.85 $\pm$ 0.23%).

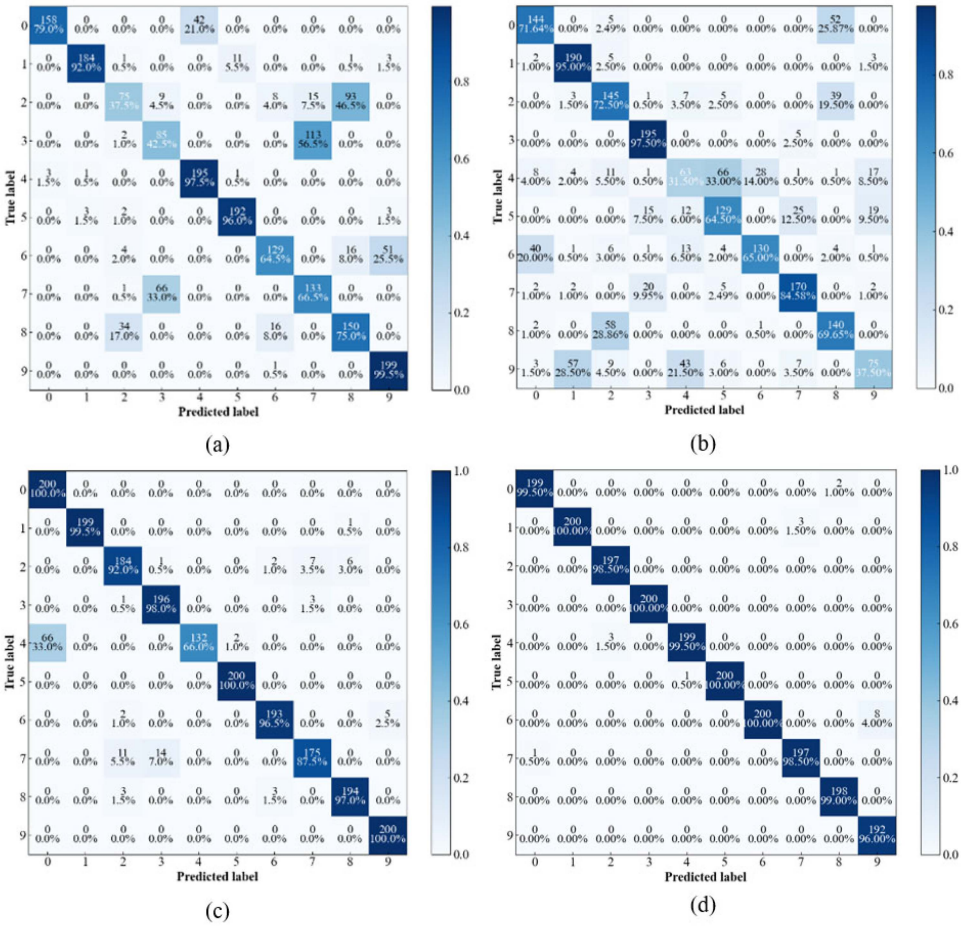


Figure 14. Confusion matrix for various models when the number of training samples of each type of fault is 20 (Case 2). (**a**) ResNet; (**b**) WDCNN; (**c**) CNN-LSTM; (**d**) ADSRN.

To further provide a more intuitive illustration of the superior feature extraction capability of ADSRN, T-SNE technology is employed to visualize the feature extraction results in a two-dimensional space. Different colors represent different normal states. As shown in Figure 15, a higher degree of clustering within the same color distribution and a lower degree of overlap with other color distributions indicate that better feature extraction is achieved for a specific category by the model. Figure 15a–d presents the two-dimensional visualizations of the feature extraction results obtained using Residual Network (ResNet), Wide and Deep Convolutional Neural Network (WDCNN), Convolutional Long Short-Term Memory Network (CNN-LSTM), and ADSRN methods.

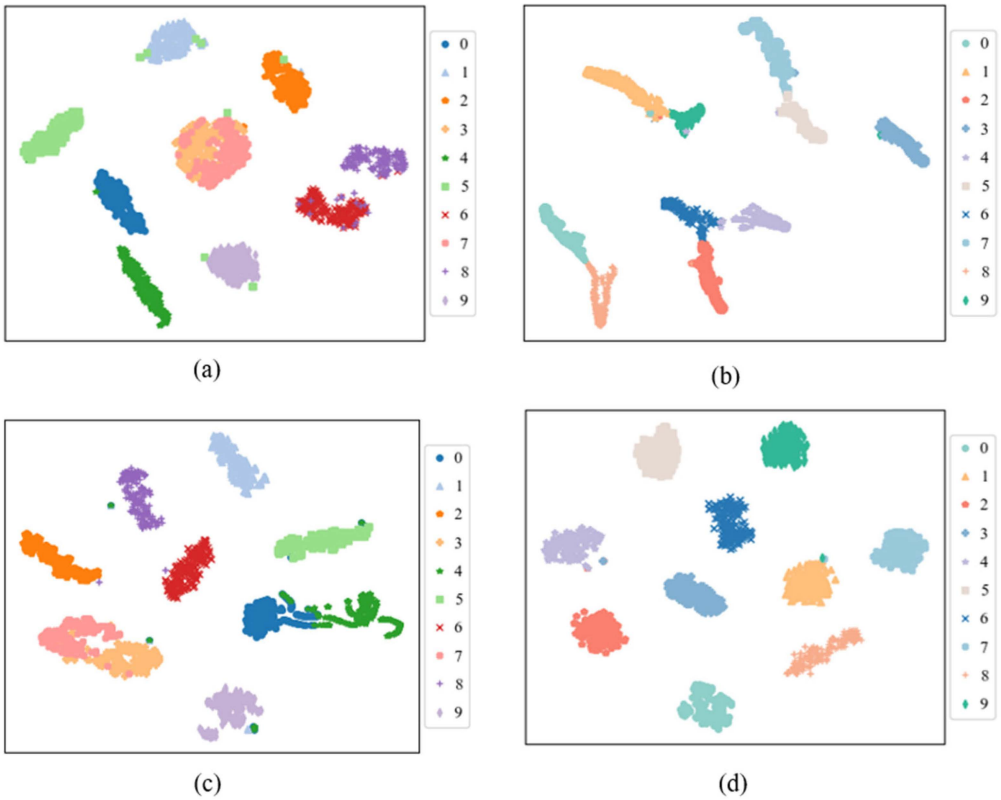


Figure 15. Feature visualization using various models when the number of training samples of each type of fault is 20 (Case 2). (**a**) ResNet; (**b**) WDCNN; (**c**) CNN-LSTM; (**d**) ADSRN.

In the proposed model, samples of individual categories 2 (1200, 0 Nm, comb) and 3 (1800, 7.32 Nm, comb) are mistakenly classified into category 4 (1200, 0 Nm, Inner), while categories 9 (1800, 7.32 Nm, normal) and 7 (1800, 7.32 Nm, outer) are erroneously classified into category 1 (1800, 7.32 Nm, ball). In Figure 15b, individual categories 7 (1800, 7.32 Nm, outer), 22 (1200, 0 Nm, comb), and 6 (1200, 0 Nm, outer) are incorrectly classified into category 1 (1800, 7.32 Nm, ball). Additionally, a significant overlap occurs between categories 1 (1800, 7.32 Nm, ball) and 9 (1800, 7.32 Nm, normal), and the gaps between various categories are not distinct, with some categories even overlapping. In Figure 15a, a clear confusion is observed between categories 7 (1800, 7.32 Nm, outer) and 3 (1800, 7.32 Nm, comb), and various other samples are also misclassified. Similarly, in Figure 15c, a clear confusion is seen between categories 7 (1800, 7.32 Nm, outer) and 3 (1800, 7.32 Nm, comb), and an unclear boundary exists between categories 4 (1200, 0 Nm, inner) and 0 (1200, 0 Nm, ball). Various other samples are also misclassified.

In summary, the visualizations in Figure 15 clearly demonstrate that, compared to the Residual Network (ResNet), Wide and Deep Convolutional Neural Network (WDCNN), and Convolutional Long Short-Term Memory Network (CNN-LSTM), the ADSRN model possesses a superior feature extraction capability.

4. Conclusions

In this paper, a novel Adaptive Deep Siamese Residual Network (ADSRN) method is proposed to address the few-shot problem in bearing fault diagnosis. Our approach enhances diagnostic performance through the following key steps:

(1) By randomly pairing positive training samples and negative training samples, the proposed model can mine more features from limited labeled data. Then, the ADSRN model is designed to enhance feature representation, in which the generalization

performance is improved by embedding a Residual Network Module, addressing issues of overfitting and gradient explosion.

(2) A novel DTW is designed as an output feature matching technique, which aligns these periodic characteristics in the frequency domain to more accurately capture the structural changes in signals, rather than relying solely on the absolute temporal alignment in the time domain. Furthermore, DTW facilitates a more precise measurement of distances between samples. This approach expedites model convergence and enhances robustness against outliers.

(3) The diagnostic performance of the proposed method is evaluated quantitatively on two challenging public datasets and benchmarked against state-of-the-art models, including ResNet, WDCNN, and CNN-LSTM. The experimental results indicate that our method surpasses existing approaches in terms of classification performance and computational efficiency. Specifically, using only 30 training samples, the model achieves accuracy rates of 99.60 $\pm$ 0.35% and 98.85 $\pm$ 0.23% on the two datasets, respectively, highlighting its robust diagnostic capability under few-shot conditions with stable performance across different scenarios. In summary, the ADSRN method effectively diagnoses bearing faults with limited training data and without requiring large amounts of auxiliary information. This study proposes a novel deep learning framework for bearing fault diagnosis, offering promising potential for practical applications. However, the scope of experimental validation remains limited. Future research could focus on integrating this method with other network architectures to overcome the inherent limitations of Siamese neural networks.

Author Contributions: Methodology, Y.J. and C.T.; Validation, M.L.; Formal analysis, J.S.; Investigation, Z.D.; Resources, Z.X.; Data curation, W.J.; Visualization, Z.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 51405449 and Grant 52405139 and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LZ22E050001. The Natural Science Foundation of Zhejiang Province (Youth Project, Grant No. LQN25E050023), the Department of Education Zhejiang Province (General Project, Grant No. Y202455248), and the Major Projects of the Jinhua Science and Technology Program (Grant No. 2024A11054).

Data Availability Statement: The data cannot be made publicly available upon publication because they are owned by a third party and the terms of use prevent public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Conflicts of 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.

References

1. Cao, W.; Mecrow, B.C.; Atkinson, G.J.; Bennett, J.W.; Atkinson, D.J. Overview of electric motor technologies used for more electric aircraft (MEA). *IEEE Trans. Ind. Electron.* **2011**, *59*, 3523–3531.
2. Bose, B.K. Power electronics and motor drives recent progress and perspective. *IEEE Trans. Ind. Electron.* **2008**, *56*, 581–588. [CrossRef]
3. Pastukhov, A.; Timashov, E.; Stanojević, D. Temperature Conditions and Diagnostics of Bearings. *Appl. Eng. Lett. J. Eng. Appl. Sci.* **2023**, *8*, 45–51. [CrossRef]
4. Xu, F.; Ding, N.; Li, N.; Liu, L.; Hou, N.; Xu, N.; Guo, W.; Tian, L.; Xu, H.; Wu, C.-M.L.; et al. A Review of Bearing Failure Modes, Mechanisms and Causes. *Eng. Fail. Anal.* **2023**, *152*, 107518. [CrossRef]
5. Dong, Z.; Jiang, Y.; Jiao, W.; Zhang, F.; Wang, Z.; Huang, J.; Wang, X.; Zhang, K. Double attention-guided tree-inspired grade decision network: A method for bearing fault diagnosis of unbalanced samples under strong noise conditions. *Adv. Eng. Inform.* **2025**, *64*, 103004. [CrossRef]

6. Wang, F.; Liu, C.; Su, W.; Xue, Z.; Li, H.; Han, Q. Condition monitoring and fault diagnosis methods for low-speed and heavy-load slewing bearings: A literature review. *J. Vibroeng.* **2017**, *19*, 3429–3444. [CrossRef]
7. Divya, D.; Marath, B.; Kumar, M.S. Review of fault detection techniques for predictive maintenance. *J. Qual. Maint. Eng.* **2023**, *29*, 420–441. [CrossRef]
8. Raouf, I.; Kumar, P.; Kim, H.S. Deep learning-based fault diagnosis of servo motor bearing using the attention-guided feature aggregation network. *Expert Syst. Appl.* **2024**, *258*, 125137. [CrossRef]
9. Wang, X.; Li, A.; Han, G. A deep-learning-based fault diagnosis method of industrial bearings using multi-source information. *Appl. Sci.* **2023**, *13*, 933. [CrossRef]
10. Tang, Y.; Zhang, C.; Wu, J.; Xie, Y.; Shen, W.; Wu, J. Deep Learning-Based Bearing Fault Diagnosis Using a Trusted Multiscale Quadratic Attention-Embedded Convolutional Neural Network. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 3513215. [CrossRef]
11. Zhao, H.; Yang, X.; Chen, B.; Chen, H.; Deng, W. Bearing fault diagnosis using transfer learning and optimized deep belief network. *Meas. Sci. Technol.* **2022**, *33*, 065009. [CrossRef]
12. Wang, Y.; Li, D.; Li, L.; Sun, R.; Wang, S. A novel deep learning framework for rolling bearing fault diagnosis enhancement using VAE-augmented CNN model. *Heliyon* **2024**, *10*, e35407. [CrossRef]
13. Hakim, M.; Omran, A.A.B.; Ahmed, A.N.; Al-Waily, M.; Abdellatif, A. A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations. *Ain Shams Eng. J.* **2023**, *14*, 101945. [CrossRef]
14. Li, L.; Wang, Z.-S. Compression Quality Prediction Model for JPEG2000. *IEEE Trans. Image Process.* **2009**, *19*, 384–398.
15. Kim, J.; Yoon, J.; Byun, J.-Y.; Kim, H.; Han, S.; Kim, J.; Lee, J.H.; Jo, H.-S.; Chung, S. Nano-interstice driven powerless blood plasma extraction in a membrane filter integrated microfluidic device. *Sensors* **2021**, *21*, 1366. [CrossRef]
16. Fu, S.; Liu, K.; Xu, Y.; Liu, Y. Rolling Bearing Diagnosing Method Based on Time Domain Analysis and Adaptive Fuzzy C-Means Clustering. *Shock. Vib.* **2016**, *2016*, 9412787.
17. Peng, W.; Danqing Li Heng, W. Rolling bearing fault diagnosis algorithm based on improved alternating transfer learning. *J. Vib. Shock* **2024**, *43*, 239–249.
18. Wang, J.; Chen, Y.; Feng, W.; Yu, H.; Huang, M.; Yang, Q. Transfer Learning with Dynamic Distribution Adaptation. *ACM Trans. Intell. Syst. Technol. TIST* **2020**, *11*, 1–25. [CrossRef]
19. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic Meta-learning for Fast Adaptation of Deep Networks. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; Volume 70, pp. 1126–1135.
20. Zeng, L.; Jian, J.; Chang, X.; Wang, S. A meta-learning method for few-shot bearing fault diagnosis under variable working conditions. *Meas. Sci. Technol.* **2024**, *35*, 056205. [CrossRef]
21. Li, C.; Li, S.; Zhang, A.; He, Q.; Liao, Z.; Hu, J. Meta-learning for few-shot bearing fault diagnosis under complex working conditions. *Neurocomputing* **2021**, *439*, 197–211. [CrossRef]
22. Su, H.; Xiang, L.; Hu, A.; Xu, Y.; Yang, X. A novel method based on meta-learning for bearing fault diagnosis with small sample learning under different working conditions. *Mech. Syst. Signal Process.* **2022**, *169*, 108765. [CrossRef]
23. Chen, L.; Lu, S.; Chen, T. Understanding benign overfitting in gradient-based meta learning. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 19887–19899.
24. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [CrossRef]
25. Ahang, M.; Jalayer, M.; Shojaeinasab, A.; Ogunfowora, O.; Charter, T.; Najjaran, H. Synthesizing rolling bearing fault samples in new conditions: A framework based on a modified CGAN. *Sensors* **2022**, *22*, 5413. [CrossRef] [PubMed]
26. Shao, S.; Wang, P.; Yan, R. Generative adversarial networks for data augmentation in machine fault diagnosis. *Comput. Ind.* **2019**, *106*, 85–93. [CrossRef]
27. Hou, Z.-G.; Wang, H.-W.; Lv, S.-L.; Xiong, M.-L.; Peng, K. Siamese multiscale residual feature fusion network for aero-engine bearing fault diagnosis under few-shot condition. *Meas. Sci. Technol.* **2022**, *34*, 035109. [CrossRef]
28. Xu, K.; Kong, X.; Wang, Q.; Han, B.; Sun, L. Intelligent fault diagnosis of bearings under small samples: A mechanism-data fusion approach. *Eng. Appl. Artif. Intell.* **2023**, *126*, 107063. [CrossRef]
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
30. Zhang, W.; Peng, G.; Li, C.; Chen, Y.; Zhang, Z. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors* **2017**, *17*, 425. [CrossRef] [PubMed]
31. Sun, H.; Zhao, S. Fault diagnosis for bearing based on 1DCNN and LSTM. *Shock. Vib.* **2021**, *2021*, 1221462. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.