Open Journal of Applied Sciences, 2026, 16(3), 950-974
https://www.scirp.org/journal/ojapps
ISSN Online: 2165-3925
ISSN Print: 2165-3917

Application of a Projection Filtering Enhanced Convolutional Network in Fault Diagnosis of Variable-Speed Rolling Bearings

Lijie Tang*, Weiwen Tian

Hunan Provincial Key Laboratory of Health Maintenance for Mechanical Equipment, Hunan University of Science and Technology, Xiangtan, China
Email: *15367134059@163.com, 2506149770@qq.com

How to cite this paper: Tang, L.J. and Tian, W.W. (2026) Application of a Projection Filtering Enhanced Convolutional Network in Fault Diagnosis of Variable-Speed Rolling Bearings. *Open Journal of Applied Sciences*, **16**, 950-974.
https://doi.org/10.4236/ojapps.2026.163056

Received: March 6, 2026
Accepted: March 21, 2026
Published: March 24, 2026

Abstract

Under variable-speed conditions in modern industry, rolling bearing vibration signals are highly non-stationary, and fault features are easily obscured by speed interference and noise. Traditional methods and existing models, lacking systematic interference suppression and targeted feature screening, often result in low feature discrimination and insufficient diagnostic accuracy, limiting industrial applications. To address this, a projection-filtering-based convolutional network (NDACN) is proposed for rolling bearing fault diagnosis. First, a projection matrix is constructed via non-redundant attribute projection (NAP) to effectively strip variable-speed interference, converting non-stationary features into approximately constant-speed stationary features for de-conditioning. Second, a dynamic filtering threshold (DFT) module is adopted, using gradient retention and dynamic threshold adjustment to adaptively select effective fault features, mitigating key information loss and gradient degradation caused by traditional hard thresholding. Then, a parameterized attention convolution enhancement mechanism is introduced to focus on critical fault bands and feature regions through adaptive weight allocation, strengthening core fault representation while suppressing irrelevant interference, and reducing information loss via gated residual fusion. Finally, validation on three variable-speed bearing datasets shows that the proposed method achieves superior diagnostic accuracy and stability under variable-speed conditions.

Keywords

Rolling Bearing, Convolutional Network, Projection Filtering, Fault Diagnosis

1. Introduction

With the continuous improvement of the intelligent level of modern industrial equipment, higher requirements are put forward for the accuracy and adaptability of condition monitoring and fault diagnosis [1]. As the core transmission component of major equipment such as rotating machinery, aerospace, and rail transit, the operating state of rolling bearings directly determines the stability and safety of the whole system. Faults related to bearings account for 30% - 60%, and are regarded as the main cause of equipment failure and even major safety accidents [2]. In practical complex industrial scenarios, bearings mostly operate under non-stationary working conditions such as variable speed, frequent start-stop and load fluctuation. Fault samples are difficult to obtain, and vibration signals present obvious non-stationary and nonlinear characteristics [3]-[5]. Meanwhile, strong interference caused by speed variation easily covers weak fault features, so that traditional diagnosis methods based on the stationary signal assumption can hardly extract high-discrimination fault information [6], resulting in insufficient generalization ability and greatly reduced diagnosis accuracy of the model [7] [8]. Therefore, fault diagnosis of rolling bearings under variable speed conditions has become a key research hotspot and an important technical difficulty in the field of mechanical fault diagnosis.

To address the problem of bearing fault diagnosis under variable speed conditions, existing studies are mainly divided into two technical routes: signal processing and intelligent diagnosis. In terms of signal processing, methods such as order tracking, empirical mode decomposition and wavelet decomposition are widely used in feature extraction of non-stationary vibration signals, and the preliminary separation of fault features and noise can be realized [9] [10]. However, such methods highly rely on manual experience and parameter adjustment, and show poor adaptability to complex variable speed conditions with wide range and strong fluctuation. In the field of intelligent diagnosis, deep learning models including convolutional neural network (CNN) [11], long short-term memory network (LSTM), CNN-LSTM hybrid model [12], wavelet domain convolutional neural network (WDCNN) [13], domain adversarial neural network (DANN) [14], residual convolutional neural network (ResCNN) [15] have gradually become research hotspots due to their powerful adaptive feature learning ability. Tong *et al.* [16] proposed a domain adaptation method based on feature transfer learning to solve the performance degradation of diagnosis models caused by data distribution difference under variable working conditions, by which the accuracy of bearing fault identification is effectively improved and superior performance is achieved compared with competitive methods. Although some studies have alleviated the influence of working condition differences to a certain extent with strategies such as domain adaptation and residual structure [17], obvious shortcomings still exist in existing models. Redundant information related to speed is easily retained in the feature extraction process, and a targeted screening and enhancement mechanism for fault-sensitive features is lacked, which greatly restricts the diagnosis accuracy and generalization ability under actual variable speed conditions.

In bearing fault diagnosis under variable speed, affected by the non-stationarity of vibration signals caused by time-varying speed, fault features are easily dispersed in different time-frequency domains and superimposed with noise and redundant information. Traditional feature extraction methods treat all frequency band features equally, and problems such as low discrimination of key fault features, insufficient distinction between different frequency band features, and inadequate fusion of global and local features are widely encountered. Li *et al.* [18] proposed an intelligent diagnosis framework based on empirical mode decomposition and multi-modal feature collaborative optimization for the difficulty of gearbox fault diagnosis under complex working conditions, such as variable speed and strong noise. A fault diagnosis model integrating multi-scale convolutional neural network and lightweight convolutional attention is constructed, by which the adaptability to complex working conditions and diagnosis efficiency are effectively improved. Aiming at the problem of accurate fault diagnosis of rolling bearings, Wang *et al.* [19] proposed a bearing fault diagnosis method based on empirical mode decomposition and improved Chebyshev distance. Effective bearing fault diagnosis is realized through signal equal division, IMF decomposition, SDP image transformation and feature distance calculation. To solve the problems of strong noise interference and difficulty in extracting discriminative domain-invariant features in cross-domain fault diagnosis of rolling bearings, Li *et al.* [20] proposed an intelligent diagnosis model based on DANN and attention mechanism, in which EEMD denoising and attention mechanism are adopted to optimize feature extraction. In addition, to solve the problems that existing graph attention networks do not consider the characteristics of vibration signals and the attention coefficients cannot effectively reflect the importance of edges, Li *et al.* [21] proposed a recursive multi-head graph attention residual network. Feature representation is enhanced through recursive graph construction, multi-kernel Gaussian symmetric graph attention mechanism and graph encoding module, and the effectiveness and superiority of the method under strong noise are verified by Data. Furthermore, the attention mechanism can adaptively adjust the focus according to vibration signals without manual feature screening, and the adaptability of the model under variable speed and strong noise conditions is improved. Combined with the gated residual fusion strategy, adaptive feature enhancement is realized and the loss of fault information is avoided, thus providing high-discrimination features for high-precision fault diagnosis [22] [23].

As an effective method for working condition interference suppression, non-redundant attribute projection has been gradually applied in bearing fault diagnosis. By constructing a projection matrix, working condition interferences such as speed in the feature space can be effectively eliminated, and the purification and enhancement of fault features are realized, which provides a new approach for solving the non-stationarity of variable speed signals [24] [25]. To solve the problem that traditional decomposition methods are prone to errors when processing broadband signals with sharp harmonics, He *et al.* [26] realized effective separation of broadband modes, narrowband modes and noise from complex non-stationary signals. Higher decomposition accuracy is obtained compared with EEMD and VMD, and the method is suitable for the quality evaluation of welding inverter power supply signals. However, conventional feature extraction methods treat all frequency band features equally, and can hardly focus on the key frequency bands and feature regions related to faults, which fails to meet the requirements of high-precision diagnosis under complex variable speed conditions.

To solve the problems including strong non-stationarity of vibration signals, easy coverage of fault features by interference, easy loss of key information in traditional feature screening methods, and insufficient ability of deep learning models to focus on fault features under variable speed conditions, a convolutional network based on projection filtering (Nuisance Attribute Projection-Dynamic Filtering Threshold-Attention-CNN, NDACN) is proposed in this chapter for the intelligent fault diagnosis of rolling bearings under variable speed. The main innovations and contributions of this study are summarized as follows:

1) Non-redundant attribute projection (NAP) is introduced as a pre-processing method for eliminating working condition interference. A special projection matrix is constructed to effectively remove irrelevant working condition information such as speed fluctuation and load variation. The original non-stationary variable speed features are gradually transformed into stationary features under approximately constant speed, and the influence of working condition difference on subsequent feature learning is reduced from the source.

2) A dynamic filtering threshold (DFT) module is introduced. The screening and enhancement of effective fault features are completed by means of adaptive threshold adjustment and a gradient retention mechanism. The problems of key feature loss and gradient vanishing caused by traditional hard threshold screening are avoided, and the saliency and discrimination of fault features are further improved.

3) A parameterized attention convolution enhancement mechanism is embedded. The model is enabled to automatically focus on frequency band information and local feature regions that contribute more to fault diagnosis through adaptive weight assignment. The core fault representation is strengthened while the interference of noise and redundant information is suppressed. Combined with the gated residual fusion strategy, the information loss in the process of feature transmission is reduced, and high-precision and high-robustness bearing fault identification under variable speed environment is finally realized.

2. Theoretical Backgrounds

2.1. Nuisance Attribute Projection

Nuisance Attribute Projection (NAP) was originally employed to suppress interferences in tasks such as face recognition and image recognition. In this study, interferences induced by variable operating conditions are regarded as redundant attributes. NAP is introduced to eliminate such interference in the feature space, mitigate speed-related effects, and enhance feature representation for improved bearing fault diagnosis.

The schematic of the NAP principle is depicted in **Figure 1**. The vector S denotes redundant attributes from varying operating conditions, while the target vector C contains the essential information related to bearing fault modes and severity.

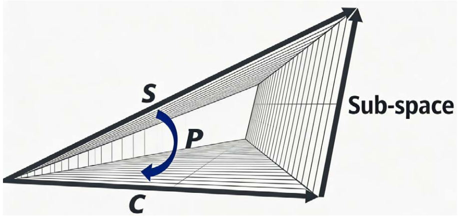


Figure 1. Nuisance attribute projection.

Consider an N-dim feature space, n samples from various operating conditions can be represented as a $N \times n$ sample matrix $F = [f_1, f_2, \cdots, f_n]$, and the feature matrix F' is calculated from F and the projection matrix P as follows:

$$F' = P \times F \tag{1}$$

where P is an $N \times N$ matrix.

$$P = I - \sum_i^d w_i w_i^{\mathrm{T}} \tag{2}$$

where I is an $N \times N$ unit matrix, i is the i_{th} NAP direction, w_i is an orthonormal vector, and d is the number of major directions extracted from the original feature space with the $d \le N$.

The parameter d is selected to balance computational complexity and projection performance. A larger d improves the projection effect but increases computational cost, which is evaluated using the sum of distances between projected features. NAP aims to minimize the projection error projection effect via the projection matrix *P*, where a smaller projection effect indicates better performance. PE is computed as follows.

$$\mathrm{PE} = \sum_{i,j} W_{ij} \left\| P \cdot x_i - P \cdot x_j \right\|^2 \tag{3}$$

where W_{ij} is the weight matrix representing the relation between every feature sample.

W_{ij} is positive when x_i and x_j come from different conditions or zero when these pairs come from the same condition. In this paper, different operating conditions refer to different rotational speeds or channels, which require removing interfering attributes in the feature vectors. W_{ij} in NAP algorithm is shown in Equation (4):

$$W_{ij} = \begin{cases} 1 & \text{if } wc(x_i) \ne wc(x_j) \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where the wc is operation condition.

The main feature vectors of the following feature problems can be obtained by transforming the result P in Equation (3):

$$FZ\left(W\right)F^{\mathrm{T}}w_i = \lambda_i w_i\left(i = 1,\cdots,d\right) \tag{5}$$

where the matrix $Z\left(W\right) = \mathrm{diag}\left(W \cdot L\right) - W$, $\mathrm{diag}\left(\cdot\right)$ is a factor that forms a diagonal matrix by vectors, and L is the length column vector of every vector.

The matrix P can be calculated by solving for w_i in Equation (5) to remove nuisance attributes from the characteristic space.

In this study, the samples for constructing the working condition matrix W are the multi-speed vibration feature samples of healthy bearings in the training set. The parameter d is determined by traversing integers from 1 to N, calculating the projection error (PE) for each integer, and selecting the value that minimizes the PE. The solution of the projection matrix P is performed using only the training set data throughout the process, with no test set data involved, thus avoiding data leakage.

2.2. Based on Invertible Gaussian Reparameterization Encoding

Invertible Gaussian Reparameterization (IGR, Invertible Gaussian Reparameterization) [27] is a distribution technique that maps Gaussian noise onto a simplex, providing a continuous relaxation of discrete distributions. The IGR defined on the simplex $S^{(K-1)}$ is determined by parameters $\left(\mu,\sigma\right)$, where $\mu \in R^{K-1}$ and $\sigma \in \left(0,\infty\right)^{K-1}$, $\sigma > 0$. A Gaussian noise $\varepsilon \sim N\left(0, I_{K-1}\right)$ is used for the affine transformation:

$$y = \mu + \mathrm{diag}\left(\sigma\right)\varepsilon \tag{6}$$

Then, perform an invertible smooth mapping (where $\tau > 0$ is a temperature hyperparameter):

$$\tilde{z} = g\left(y,\tau\right)_k = \frac{\exp\left(y_k/\tau\right)}{\sum_{j-1}^{K-1}\left(\dfrac{y_j}{\tau}\right) + \delta} \tag{7}$$

Obtain the density function through variable transformation:

$$\tilde{q}_{\mu,\sigma,\tau}\left(\tilde{z}\right) = N\left(y \mid \mu,\sigma\right)\left|\det J_g\left(y,\tau\right)\right|^{-1} \tag{8}$$

where $J_g\left(y,\tau\right)$ is the Jacobian matrix of $g\left(y,\tau\right)$. The Jacobian term cancels out in the KL divergence ratio, simplifying to the Gaussian KL divergence:

$$\mathrm{KL}\left(\tilde{q}_{\mu,\sigma,\tau}\tilde{z} \,||\, \tilde{q}_{\mu_0,\sigma_0,\tau_0}\left(\tilde{z}\right)\right) = \mathrm{KL}\left(N\left(\mu,\sigma^2\right) \,||\, N\left(\mu_0,\sigma_0^2\right)\right) \tag{9}$$

Due to the lack of parameter interpretability, solve for the reference parameters by optimization (where p_0 is the target distribution):

$$\left(\mu_0,\sigma_0\right) = \arg\min_{(\mu,\sigma)} E_{\mu,\sigma,\tau}\left(\tilde{z}\right)\left[\left\|\tilde{z} - p_0\right\|_2^2\right] \tag{10}$$

3. The Proposed Method

3.1. Framework of the NDACN Method

The NDACN method achieves accurate fault diagnosis of variable-speed rolling bearings through five progressive steps, which are closely connected in accordance with the flowchart logic as shown in **Figure 2**. The specific implementation is as follows:

Step 1: NAP Projection Processing. Twelve time-domain features, six frequency-domain features, eight node energies from three-layer wavelet packet decomposition and one envelope spectrum entropy are extracted from the vibration signals of variable-speed bearings to construct a 27-dimensional original feature matrix F. A working condition matrix W is built using multi-speed healthy signals, which are all derived from the training set and correspond to fault-free feature samples under different rotational speeds.

The orthogonal projection directions are obtained by solving the characteristic equation to construct the projection matrix $p = I - \sum_i^d w_i w_i^{\mathrm{T}}$. The stripping of rotational speed interference is achieved via $K = P \times F$, converting the non-stationary variable-speed features into approximately stationary features under constant speed.

Step 2: Data Preprocessing. Z-score standardization is adopted to normalize the de-conditioned feature matrix, eliminating the influence of different dimensions. The standardized data are divided into a training set and a validation set A_i, which are used for model training, hyperparameter tuning, and diagnostic performance evaluation respectively to ensure the rationality of data utilization.

Step 3: DFT Dynamic Feature Screening. Neighborhood features $[L_0, L_1] = \mathrm{Conv1}(A_i)$ are extracted via one-dimensional convolution, and Gaussian noise is injected to adapt the dimensions through parameter broadcasting. A temperature hyperparameter is introduced to construct a discrete distribution, and effective features are screened via Softmax soft filtering and argmax hard threshold decision. Invertible Gaussian reparameterization (IGR) is utilized to retain gradients, and the sparse degree is controlled via the loss filtering function L_{SF}, outputting a high-discrimination feature matrix B_i.

Step 4: Attention Enhancement Mechanism. The feature sparsity metrics M_i and Ω_i are calculated based on high-order statistics, and a dynamic sparsity control loss is constructed to guide the model to focus on key fault features. Q, K, and V vectors are extracted via one-dimensional convolution, and the similarity is calculated to generate an attention weight matrix, which is weighted and aggregated to obtain enhanced features. A learnable scaling coefficient is introduced, and gated residual fusion $G_i = B_i + r_i \odot E_i$ is adopted to strengthen the representation ability of fault features.

Step 5: Fault Diagnosis via Convolutional Network. The attention-enhanced features are input into the convolutional network, where local features are captured via convolutional layers, the generalization ability is improved pooling layers, and the features are mapped to the fault category space via fully connected layers. The model is trained using the cross-entropy loss function and the Adam optimizer, and an adaptive learning rate strategy is set to avoid overfitting. Finally, four types of diagnostic results are output: healthy, inner race fault, outer race fault, and cage fault.

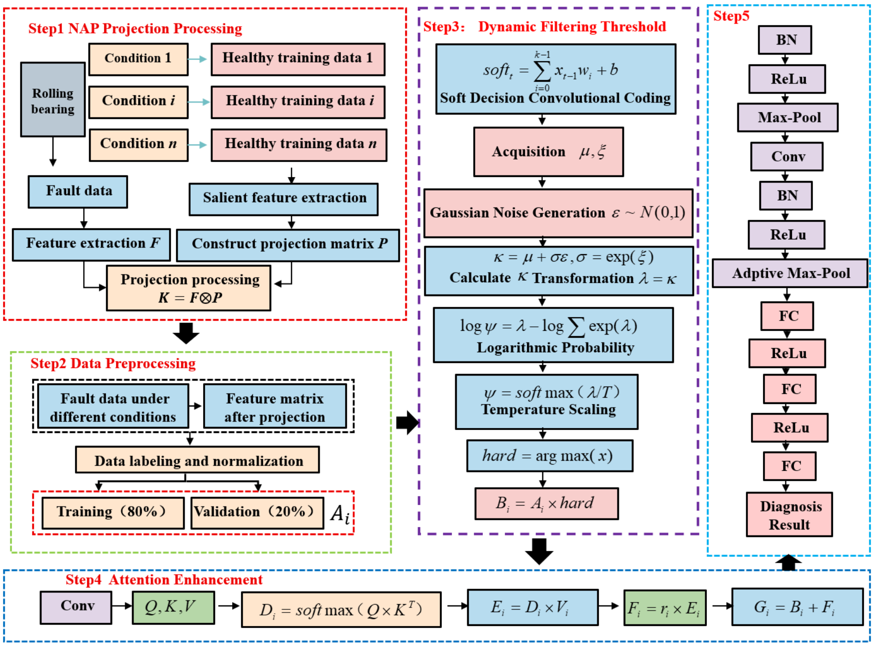


Figure 2. Framework of the NDACN method.

3.2. Dynamic Filtering Threshold

After NAP deconditioning, noise interference and redundant information may still remain in the feature matrix. Moreover, traditional hard threshold screening methods face challenges in sparsity control, which tend to cause the loss of key fault features or excessive residual noise. In addition, the direct application of hard threshold operations leads to gradient loss, which impairs the stability and convergence speed of subsequent model training. Dynamic filtering dynamically adjusts the feature retention rate through a gradient retention mechanism and a loss function.

In the establishment of threshold decision-making, the core of threshold processing is to effectively remove noise interference while preserving key fault information, which transforms feature selection into a binary mask operation:

$$soft_i = \sum_{j=0}^{k-1} x_{i-1,j} w_j + b \tag{11}$$

where $soft_i$ denotes the convolutional coding value of the *i*-th feature, $x_{i-1,j}$ represents the *j*-th feature value of the ($i-1$)-th layer, w_j is the convolution kernel weight, and b is the bias term.

After obtaining the convolutional coding, the feature information μ (feature mean) and ξ (feature standard deviation) are extracted, and Gaussian noise $\varepsilon \sim N(0,1)$ is generated. Then:

$$k = \mu + \sigma\varepsilon, \sigma = \exp(\xi) \tag{12}$$

After calculating *k*, transform $\lambda = k$ (feature scaling value) and perform probability calculation:

$$\log\psi = \lambda - \log\sum\exp(\lambda) \tag{13}$$

Parameter scaling is performed via the temperature parameter T (temperature hyperparameter):

$$\psi = \text{softmax}\left(\lambda/T\right) \tag{14}$$

Finally, feature screening is realized through a hard threshold operation:

$$Hard = \text{argmax}\left(x\right) \tag{15}$$

$$B_i = A_i \times Hard \tag{16}$$

where *Hard* denotes the hard mask value, B_i is the output high-discrimination feature matrix, and A_i is the input standardized feature matrix.

During training, the corresponding hard-optimized coding results are back-propagated for gradients via the IGR technique to preserve the differentiability of feature screening. In this study, the hard mask (*Hard*) is applied to each feature dimension; specifically, argmax *hard* threshold decision-making is performed dimension by dimension on the 27-dimensional feature matrix after soft filtering, enabling adaptive screening of feature dimensions.

3.3. Parameterized Attention Convolution Enhancement

The features screened by the DFT module still suffer from problems such as a low proportion of key information and insufficient discrimination of fault features across different frequency bands. Traditional feature extraction methods treat all frequency band features equally, making it difficult to focus on the key frequency bands related to faults, thus leading to insufficient sensitivity of diagnostic models to weak faults. The parameterized attention convolution enhancement can strengthen the representation of key fault features and suppress the interference of irrelevant features. The core idea is to reallocate attention weights, enabling the model to automatically focus on frequency bands and feature regions that contribute more to fault diagnosis, thereby improving the discriminability of weak faults. Meanwhile, combined with a gated residual enhancement strategy, it realizes the adaptive fusion of original features and attention-enhanced features, avoiding the loss of key information. Its core components include: defining sparsity metrics based on high-order statistics: $M_i = \sum_{l=1}^{L} B_i\left(l\right)^4$, $\Omega_i = \sum_{l=1}^{L} B_i\left(l\right)^2$, where B_i is the input feature vector, and $\mathcal{L}_{sk}$ is the feature sparsity loss. The dynamic sparsity control loss is:

$$\mathcal{L}_{sk} = -\frac{\Upsilon\left(e\right)L}{N}\sum_{i=1}^{N}\frac{M_i}{\Omega_i^2} \tag{17}$$

where $\Upsilon\left(e\right) \in \left(0,1\right)$ is the proportional factor decaying with training epochs, N is the number of feature channels, and $\frac{M_i}{\Omega_i^2}$ is the kurtosis of the feature, reflecting sparsity. Then, one-dimensional convolution is used to extract attention elements (where convolution kernels act independently on each channel):

$$\left[Q_i, K_i, V_i\right] = \text{Conv1}\left(B_i\right) \tag{18}$$

To ensure the consistency of feature information before and after, and to ensure that key information is attended to, the similarity matrix is calculated as:

$$D_i = \text{softmax}\left(\frac{Q_i K_i^{\mathrm{T}}}{\sqrt{d_k}}\right) \tag{19}$$

and the features are aggregated by weighting: $E_i = D_i V_i$. To achieve gated residual enhancement, a learnable scaling coefficient r_i is introduced to realize adaptive feature fusion. The final implementation effect can be expressed as:

$$G_i = B_i + r_i \odot E_i \tag{20}$$

This structure suppresses noise while retaining key fault features, as shown in Figure 3.

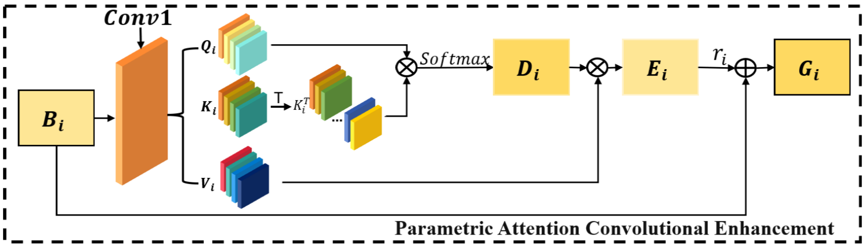


Figure 3. Parameterized attention convolution enhancement.

3.4. Data Evaluation Methods

The "imbalanced data scenario" mentioned in this study refers to the slight quantitative differences among various types of samples (non-extreme imbalance) caused by the varying difficulties in collecting fault samples under variable-speed operating conditions in actual industrial settings. In contrast to the samples collected in equal proportions in the laboratory, this scenario is more consistent with practical engineering applications. To comprehensively and objectively evaluate the fault diagnosis performance of the NDACN method, especially its ability to identify minority-class fault samples under the variable-speed and imbalanced data scenario, accuracy (Acc), precision, recall and F1-score are selected as the core evaluation metrics. All metrics are derived based on the four-class confusion matrix, covering both the overall diagnosis performance and the recognition performance of minority-class samples. Their specific definitions and calculation methods are as follows:

1) Confusion Matrix Definition

To evaluate the algorithm performance more effectively, the confusion matrix is commonly adopted. The evaluation method based on the four-class confusion matrix is shown in Table 1, and the calculation of related parameters is given in Equations (21) to (234).

2) Core Evaluation Metric Calculation

To comprehensively and objectively evaluate the model's fault diagnosis performance and avoid the one-sidedness of a single metric, classification accuracy, precision, recall, and F1-score are selected as the core classification indicators. Meanwhile, the classification result confusion matrix and feature clustering visualization are adopted as qualitative evaluation indicators. The calculation formulas for each quantitative metric are as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{21}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{22}$$

$$F_1 = \frac{\left(1+\beta^2\right)\times \text{precision} \times \text{recall}}{\beta^2 \times \text{precision} + \text{recall}} \tag{23}$$

$$F_1\text{-score} = \frac{2\times \text{Precision} \times \text{Recall}}{\text{Precision} \times \text{Recall}} \tag{24}$$

In the multi-class scenario, the F1-score is calculated as the weighted average of the F1-scores of each class, and precision and recall are also computed using the weighted average method, with the weights being the proportion of samples in each class.

Table 1. Four-class confusion matrix.

	Predicted Healthy	Predicted Inner Race Fault	Predicted Outer Race Fault	Predicted Cage Fault
Actual Healthy	TP_1	FN_{12}	FN_{13}	FN_{14}
Actual Inner Race Fault	FN_{21}	TP_2	FN_{23}	FN_{24}
Actual Outer Race Fault	FN_{31}	FN_{32}	TP_3	FN_{34}
Actual Cage Fault	FN_{41}	FN_{42}	FN_{43}	TP_4

Note: TP_k denotes the number of true positive samples for the k-th class, and FN_{kl} denotes the number of false negative samples where the k-th class is misclassified as the l-th class.

Classification accuracy reflects the model's overall recognition ability; precision and recall reflect the recognition accuracy and coverage of the model for various types of faults; F1-score is the harmonic mean of precision and recall, which is more suitable for performance evaluation on imbalanced Dataset. The confusion matrix intuitively displays the model's recognition results for each type of fault. For feature clustering visualization, the t-SNE algorithm is used to reduce high-dimensional fault features to a two-dimensional space, and the model's feature extraction ability is verified through the degree of aggregation and separation of feature clusters.

4. Experimental Verification and Analysis

4.1. Variable-Speed Test Data Acquisition

The energy recovery bearing dataset contains experimental data from simulated bearing fault tests. As illustrated in Figure 4, the test rig is primarily composed of a speed-regulating motor, test bearings with implanted faults, a B&K accelerometer, and a supporting data acquisition system. Acceleration data under eight health states of bearings were collected in the experiment, including normal state, inner race fault, outer race fault, and cage fault. The experimental platform for data acquisition consists of core components: a variable-speed AC motor, a coupling, support bearings, a housing for mounting test bearings, an accelerometer, and a data acquisition system. During operation, the test bearings are driven to rotate by the motor; if faults exist in the bearings, periodic shocks are generated. These shocks are captured as vibration signals by the accelerometer and digitized by the data acquisition system.

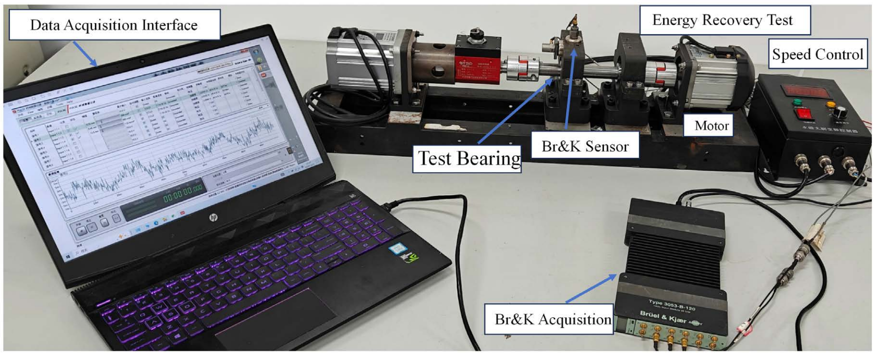


Figure 4. Data acquisition experimental platform.

The motor speed was regulated by a PLC (Programmable Logic Controller) to achieve increase, decrease, and increase-then-decrease variations, from which vibration datasets under three operating conditions were acquired. The sampling duration ranged from 10 seconds to 30 seconds, and data of healthy bearings, inner race faulty bearings, outer race faulty bearings, and cage faulty bearings were recorded respectively. The corresponding sampling frequency was 8192 Hz, and the bearing model was 6206. The faulty components of the tested rolling bearing and its disassembled parts are shown in **Figure 5**.

(a) Inner race fault bearing (b) Outer race fault bearing (c) Cage fault bearing

(d) Rolling bearing components

Figure 5. Rolling bearing fault components and configuration.

To verify the effectiveness of the NDACN method, three types of variable operating condition data were selected in the experiments, namely increasing speed, decreasing speed, and increasing-then-decreasing speed, which correspond to Experiment 1, Experiment 2, and Experiment 3, respectively. Based on the 27 dimensions of feature extraction and selection, and to prevent model overfitting, 80% of the original data was extracted as the training set and 20% as the test set after extracting significant features, with a total of 352 samples. Detailed information is presented in Table 2.

Table 2. Bearing feature dataset.

Data	Condition	Dimension	Number of Samples (Training/Validation)
Data 1 (Increasing speed)	Health Inner, Outer, Cage	27	352 282/70
Data 2 (Decreasing speed)	Health Inner, Outer, Cage	27	352 282/70
Data 3 (Increasing then decreasing speed)	Health Inner, Outer, Cage	27	352 282/70

In this study, the original data was divided into a training set and a test set at a ratio of 80:20 in the feature screening stage, with an independent validation set further split from the training set at a ratio of 20% for hyperparameter tuning and early stopping determination. In the model performance verification stage, the training-test set division ratios of 4:6, 3:7 and 2:8 were adopted to further verify the generalization ability of the model under different sample sizes. All division schemes ensured that the validation set was independent of the training and test sets, and the validation set was only used for hyperparameter tuning without participating in model training. A total of 352 samples were generated from the original vibration data through non-overlapping framing with a frame length of 1024 sampling points and an overlap rate of 0%, including 88 samples for each state of healthy condition, inner race fault, outer race fault and cage fault. The data division strictly followed the principle of "source data independence", where the framed windows from the same original collected data were assigned to either the training set or the test set exclusively to avoid data leakage. As shown in Figures 6-8 are the 10-second vibration data under the three operating conditions, corresponding to the time-domain waveforms of the healthy bearing, inner race faulty bearing, outer race faulty bearing and cage faulty bearing in sequence.

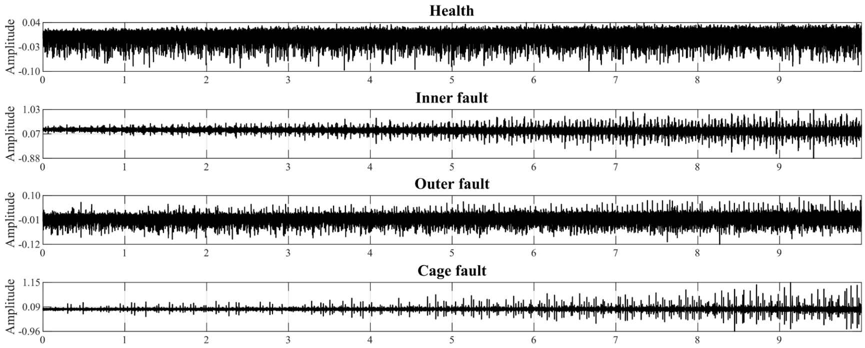


Figure 6. Time-domain diagram of experiment 1.

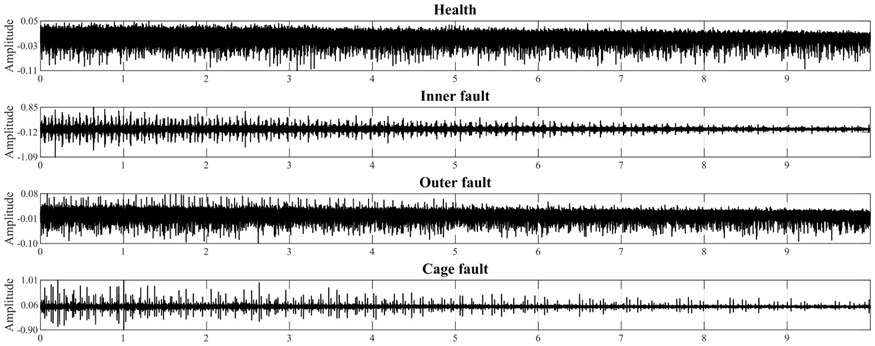


Figure 7. Time-domain diagram of experiment 2.

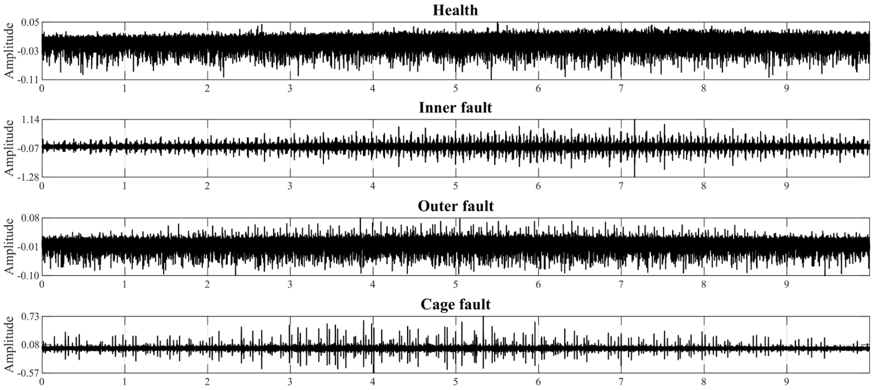


Figure 8. Time-domain diagram of experiment 3.

4.2. Feature Extraction and Selection

In the fault feature extraction, 27 features which are frequently used in bearing fault diagnosis are extracted in this section, including 12 time-domain features, 6 frequency-domain features, and 8 node energy and amplitude spectral entropy of wavelet packet three-layer decomposition. The detailed information of 18 time-domain and frequency-domain features and their computation methods are shown in Table 3.

The calculation of various features is based on the original vibration signals and spectrum data. Among them, time-domain features are directly extracted from the original signals through statistical analysis, frequency-domain features are calculated after converting the signals to the frequency domain via Fourier transform, wavelet packet features are obtained by acquiring the energy of each node through three-layer wavelet packet decomposition, and the envelope spectrum entropy is obtained by calculating the entropy of the envelope signal, comprehensively covering the time-domain, frequency-domain, and nonlinear characteristics of fault signals.

4.3. Analysis of NAP Projection Processing Effect

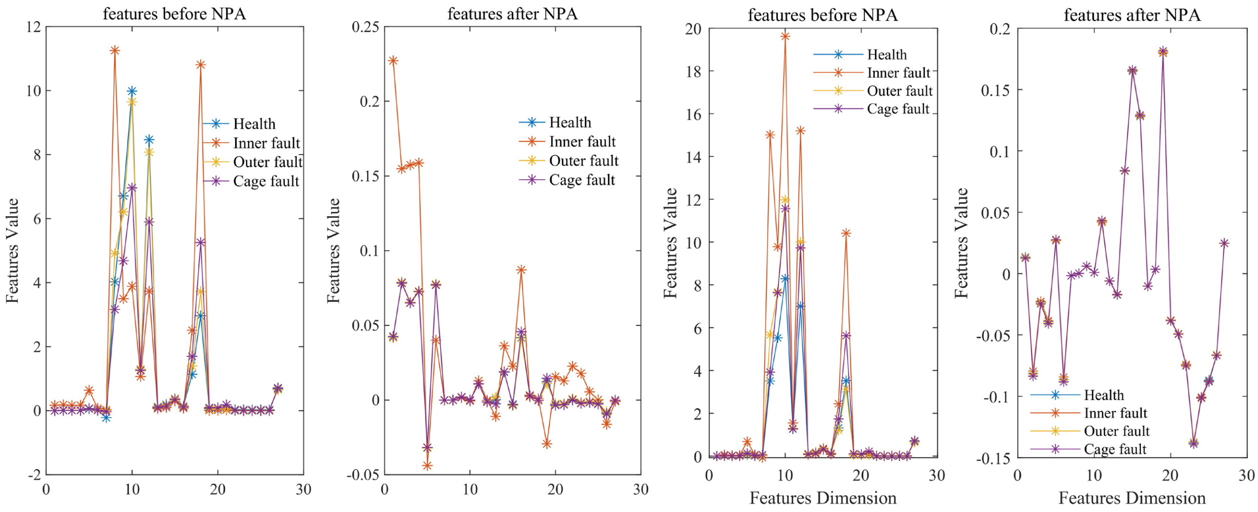
To verify the purification effect of the NAP (Noise-Attenuation Projection) method on vibration signals under variable speed conditions, three comparative experiments were performed under increasing speed, decreasing speed, and increasing-then-decreasing speed conditions. A 27-dimensional feature set was adopted, and the feature distributions before and after NAP projection are compared in **Figure 9**, aiming to eliminate speed interference and enhance fault features.

In all three experiments, consistent improvements were obtained after NAP projection. Before projection, features of healthy and faulty bearings showed obvious fluctuations and overlaps, which made different bearing conditions hard to distinguish. After projection, feature curves were smoothed, speed interference was suppressed, and the boundaries between health and fault classes became clearer.

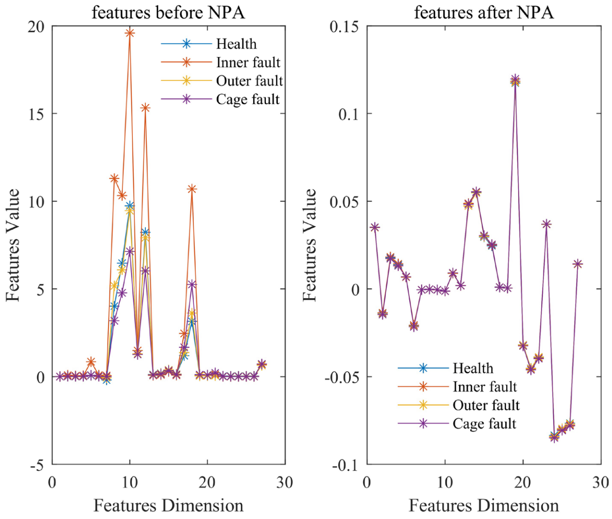
Table 3. Time-domain and frequency-domain features.

Frequency domain parameters		Time domain parameters
Mean	Skewness	Spectral gravity frequency
$P_1=\frac{1}{N}\sum_{n=1}^{N}x(n)$	$P_7=\frac{1}{N}\sum_{n=1}^{N}\left(\frac{x(n)-P_1}{P_6}\right)$	$P_{13}=\frac{\sum_{k=1}^{M}f(k)s(k)}{\sum_{k=1}^{M}s(k)}$
Root mean square	Kurtosis	Spectral root mean square frequency
$P_2=\sqrt{\frac{1}{N}\sum_{n=1}^{N}x(n)}$	$P_8=\frac{1}{N}\sum_{n=1}^{N}\left(\frac{x(n)-P_1}{P_6}\right)^4$	$P_{14}=\sqrt{\frac{\sum_{k=1}^{M}f(k)s(k)}{\sum_{k=1}^{M}s(k)}}$
Square root amplitude	Peak factor	Spectral root 4/2-moment ratio
$P_3=\left(\frac{1}{N}\sum_{n=1}^{N}\sqrt{\lvert x(n)\rvert}\right)^2$	$P_9=\frac{P_5}{P_2}$	$P_{15}=\sqrt{\frac{\sum_{k=1}^{M}f^4(k)s(k)}{\sum_{k=1}^{M}f^2(k)s(k)}}$
Mean amplitude	Clearance factor	Spectral standard deviation frequency
$P_4=\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert$	$P_{10}=\frac{P_5}{P_3}$	$P_{16}=\sqrt{\frac{\sum_{k=1}^{M}\left(f(k)-P_{17}\right)^2 s(k)}{\sum_{k=1}^{M}s(k)}}$
Maximum peak	Shape factor	Spectral frequency skewness
$P_5=\frac{\max\left(x(n)\right)-\min\left(x(n)\right)}{2}$	$P_{11}=\frac{P_2}{P_4}$	$P_{17}=\sqrt{\frac{\sum_{k=1}^{M}\left(\frac{f(k)-P_{17}}{P_{20}}\right)^3 s(k)}{\sum_{k=1}^{M}s(k)}}$
Standard deviation	Crest factor	Spectral frequency kurtosis
$P_6=\sqrt{\frac{1}{N-1}\sum_{n=1}^{N}\left(x(n)-P_1\right)^2}$	$P_{12}=\frac{P_5}{P_4}$	$P_{18}=\sqrt{\frac{\sum_{k=1}^{M}\left(\frac{f(k)-P_{17}}{P_{20}}\right)^4 s(k)}{\sum_{k=1}^{M}s(k)}}$

Note: $x(n)(n=1,2,\cdots,N)$ is the original signal, N is the number of data points; $s(k)(k=1,2,\cdots,M)$ is the amplitude of the original signal spectrum, M is the number of spectral lines; $f(k)$ is the amplitude of the k_{th} spectral line frequency.



(a) Data1 (b) Data2

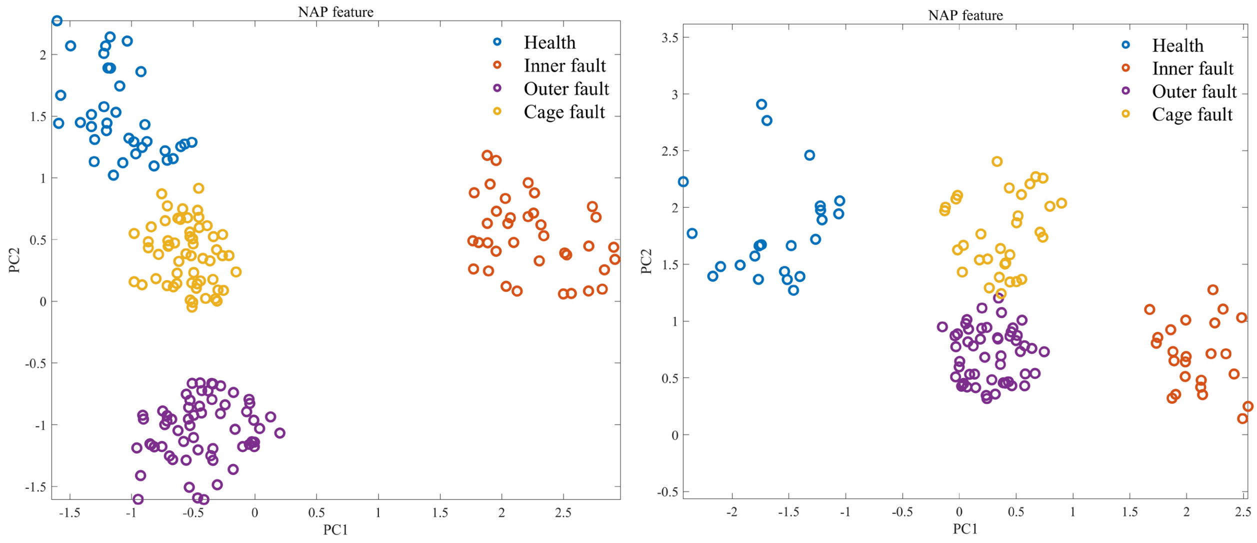


(c) Data3

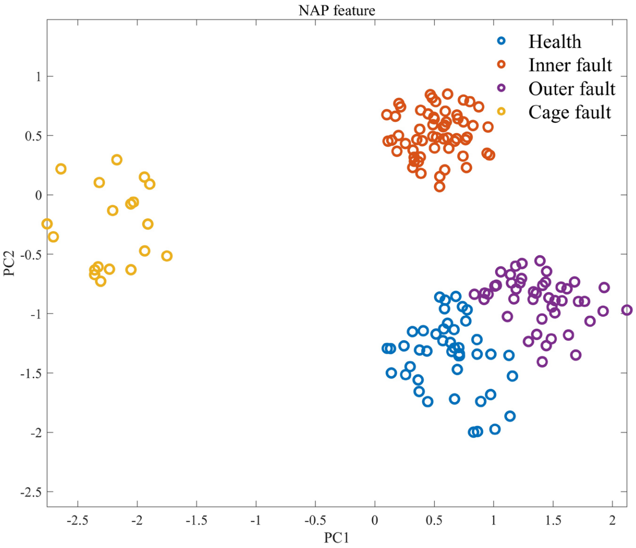
Figure 9. Feature comparison before and after NAP projection.

Experimental results show that NAP exhibits strong adaptability to different operating conditions. In the three experiments, feature overlap was greatly reduced, weak fault features were effectively enhanced, and key fault characteristics were well preserved. NAP can suppress speed interference without reducing feature dimensions, thus ensuring the integrity of input features for fault diagnosis.

To intuitively verify the optimization effect of NAP projection on feature distribution, PCA dimensionality reduction combined with K-means clustering (number of clusters $k = 3$) is adopted to project the 27-dimensional features into a two-dimensional space for visual analysis. The clustering results under three different variable-speed conditions are shown in **Figure 10**, focusing on verifying the improvement effect of NAP on the intra-class aggregation and inter-class separation of features.



(a) Data1 (b) Data2



(c) Data3

Figure 10. Visualization of feature projection matrix.

After projection, the four types of samples from independent clusters with clear boundaries and compact distributions, showing strong intra-class aggregation and little inter-class overlap. NAP projection effectively improves intra-class similarity and amplifies inter-class differences, providing highly discriminative input features for subsequent feature screening and attention enhancement modules.

The ReliefF algorithm is adopted to quantitatively evaluate the importance of the 27-dimensional features after NAP projection, with the results presented in Figure 11. Higher feature weights indicate greater contributions to fault identification, while lower weights suggest redundancy or interference. In Data-1, high-weight fault features are distinct, with the 18th-dimensional feature reaching 0.38. In Data-2, feature weights vary gently, but core features remain effective. In Data-

3, the 23rd-25th-dimensional features show significantly higher weights, capturing the key information of weak faults.

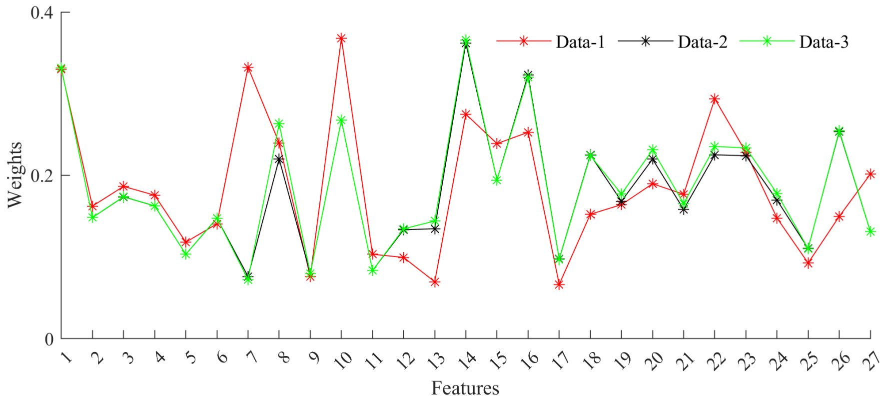


Figure 11. Feature dimension weight distribution.

The results of feature weight analysis verify the effectiveness of NAP in filtering features. It eliminates speed interference, significantly improves the contribution of core features, and maintains stability under different working conditions. Based on these results, subsequent steps can further select high-weight features to construct a simplified feature set, reducing the computational complexity of the model while retaining key fault information, and improving diagnostic efficiency and generalization ability.

4.4. Comparative Analysis of Experimental Results of Different Models

To comprehensively verify the effectiveness and superiority of the NDACN method in variable-speed bearing fault diagnosis, this paper conducts comparative Data based on three gradient working conditions of Data 1, Data 2, and Data 3, with training set and test set ratios of 4:6, 3:7, and 2:8 respectively. F1-score, precision, recall, and accuracy are taken as the main evaluation indicators. The NDACN is compared with seven diagnostic methods: CNN, LSTM, CNN-LSTM, WDCNN, DANN, ResCNN, and NAP-CNN. The results are shown in **Tables 4-6**. Overall, under various working conditions and different data division conditions, all diagnostic indicators of the NDACN method are superior to those of the comparison models, and it also shows stronger stability, which effectively verifies the reliability and advantages of the proposed method in variable-speed bearing fault diagnosis.

Specifically, the network architecture of NDACN is configured as follows: 3 layers of 1D convolutional layers (with the number of channels being 32, 64, and 128 in sequence, kernel size of 3 for all layers, stride of 1, and padding of 1); 2 layers of max-pooling layers (pooling kernel of 2 and stride of 2); 2 layers of fully connected layers (with 256 and 4 neurons respectively); and a Dropout layer (dropout rate of 0.2). For the comparative models (CNN/LSTM/CNN-LSTM/WDCNN/DANN/ResCNN), the network structure parameters adopted are classic and publicly available in the same research field, consistent with those reported in the original literature, to ensure the fairness of comparison.

Table 4. Experimental results of different models under different ratios in data 1.

Method	4:6				3:7				2:8			
	F1	Precision	Recall	Acc	F1	Precision	Recall	Acc	F1	Precision	Recall	Acc
CNN	0.7100	0.7082	0.7118	0.7056	0.6962	0.6915	0.6984	0.6892	0.6779	0.6759	0.6801	0.6724
LSTM	0.7841	0.7825	0.7858	0.7812	0.7704	0.7648	0.7731	0.7665	0.7513	0.7492	0.7535	0.7458
CNN-LSTM	0.6801	0.6791	0.6812	0.6745	0.6659	0.6627	0.6678	0.6593	0.6492	0.6452	0.6502	0.6418
WDCNN	0.8512	0.8501	0.8523	0.8486	0.8374	0.8328	0.8396	0.8342	0.8198	0.8187	0.8209	0.8135
DANN	0.8737	0.8726	0.8749	0.8712	0.8593	0.8553	0.8621	0.8568	0.8426	0.8415	0.8437	0.8382
ResCNN	0.8974	0.8963	0.8985	0.8951	0.8836	0.8791	0.8863	0.8807	0.8663	0.8652	0.8674	0.8619
NAP-CNN	0.8670	0.8659	0.8681	0.8643	0.8529	0.8487	0.8559	0.8504	0.8358	0.8347	0.8369	0.8305
NDACN	**0.9896**	**0.9895**	**0.9898**	**0.9892**	**0.9781**	**0.9732**	**0.9760**	**0.9745**	**0.9676**	**0.9681**	**0.9681**	**0.9663**

Table 5. Experimental results of different models under different ratios in data 2.

Method	4:6				3:7				2:8			
	F1	Precision	Recall	Acc	F1	Precision	Recall	Acc	F1	Precision	Recall	Acc
CNN	0.7058	0.7041	0.7075	0.7012	0.6915	0.6869	0.6940	0.6848	0.6732	0.6712	0.6753	0.6680
LSTM	0.7795	0.7779	0.7812	0.7767	0.7658	0.7603	0.7684	0.7621	0.7468	0.7447	0.7489	0.7414
CNN-LSTM	0.6756	0.6746	0.6767	0.6701	0.6615	0.6583	0.6634	0.6549	0.6449	0.6410	0.6458	0.6375
WDCNN	0.8468	0.8457	0.8479	0.8443	0.8331	0.8286	0.8353	0.8300	0.8156	0.8145	0.8167	0.8093
DANN	0.8693	0.8682	0.8705	0.8668	0.8549	0.8510	0.8577	0.8524	0.8384	0.8373	0.8395	0.8330
ResCNN	0.8931	0.8920	0.8942	0.8908	0.8793	0.8749	0.8819	0.8765	0.8621	0.8610	0.8632	0.8577
NAP-CNN	0.8627	0.8616	0.8638	0.8600	0.8487	0.8445	0.8516	0.8461	0.8317	0.8306	0.8328	0.8263
NDACN	**0.9854**	**0.9853**	**0.9855**	**0.9850**	**0.9739**	**0.9691**	**0.9720**	**0.9703**	**0.9635**	**0.9640**	**0.9640**	**0.9622**

Table 6. Experimental results of different models under different ratios in data 3.

Method	4:6				3:7				2:8			
	F1	Precision	Recall	Acc	F1	Precision	Recall	Acc	F1	Precision	Recall	Acc
CNN	0.7142	0.7125	0.7159	0.7100	0.7004	0.6958	0.7028	0.6936	0.6818	0.6798	0.6838	0.6768
LSTM	0.7887	0.7871	0.7904	0.7856	0.7749	0.7693	0.7770	0.7709	0.7558	0.7537	0.7579	0.7502
CNN-LSTM	0.6846	0.6835	0.6857	0.6789	0.6698	0.6665	0.6719	0.6642	0.6535	0.6495	0.6546	0.6460
WDCNN	0.8556	0.8545	0.8568	0.8520	0.8412	0.8365	0.8436	0.8383	0.8238	0.8227	0.8250	0.8176
DANN	0.8781	0.8770	0.8793	0.8744	0.8637	0.8597	0.8664	0.8610	0.8468	0.8457	0.8480	0.8414
ResCNN	0.9017	0.9006	0.9029	0.8982	0.8879	0.8833	0.8900	0.8846	0.8705	0.8694	0.8717	0.8650
NAP-CNN	0.8713	0.8702	0.8725	0.8674	0.8568	0.8525	0.8590	0.8535	0.8398	0.8387	0.8410	0.8343
NDACN	**0.9938**	**0.9937**	**0.9940**	**0.9933**	**0.9823**	**0.9774**	**0.9799**	**0.9788**	**0.9717**	**0.9722**	**0.9722**	**0.9700**

The training settings are as follows: batch size = 32; the optimizer is Adam with an initial learning rate of 0.001, and the cosine annealing learning rate scheduler is adopted ($T_{\max} = 50$, $\eta_{\min} = 0.0001$); the number of training epochs = 100; the early stopping strategy is implemented (patience = 10, *i.e.*, training stops if the validation set loss does not decrease for 10 consecutive epochs); L_2 regularization is applied (weight decay = 0.0005); all models are implemented based on the PyTorch framework.

Under the 4:6 training-test set division ratio, the comprehensive diagnostic performance of NDACN can reach the relatively optimal level in all three data: In Data 1, the F1-score, precision, recall, and accuracy are 0.9896, 0.9895, 0.9898, and 0.9892, respectively. In Data 2, the values are 0.9854, 0.9853, 0.9855, and 0.9850, respectively. In Data 3, the values are 0.9938, 0.9937, 0.9940, and 0.9933, respectively. Compared with the suboptimal model ResCNN, all indicators of NDACN are improved by more than 9%. Compared with models such as DANN, WDCNN, and NAP-CNN, the advantages are more obvious, and the performance gap with basic models such as CNN, LSTM, and CNN-LSTM is further expanded.

As the training set ratio gradually decreases to 3:7 and 2:8, the diagnostic performance of all models declines to varying degrees, but the performance degradation of NDACN is always the smallest: In Data 1, its accuracy decreases from 0.9892 to 0.9663, with a decline of only 2.31%.In Data 3, under the 2:8 small-sample training ratio, it can still maintain a high accuracy of 0.9700. In contrast, among the comparison models, the performance degradation of CNN and CNN-LSTM exceeds 4.5%, that of LSTM, DANN, and WDCNN is above 3.5%, and the accuracy of NAP-CNN under the 2:8 ratio is only 0.8343, which is 13.57% lower than that of NDACN.

To intuitively analyze the diagnostic performance of the 8 models under the 4:6 training-test set ratio, the specific results are shown in **Figure 12**.

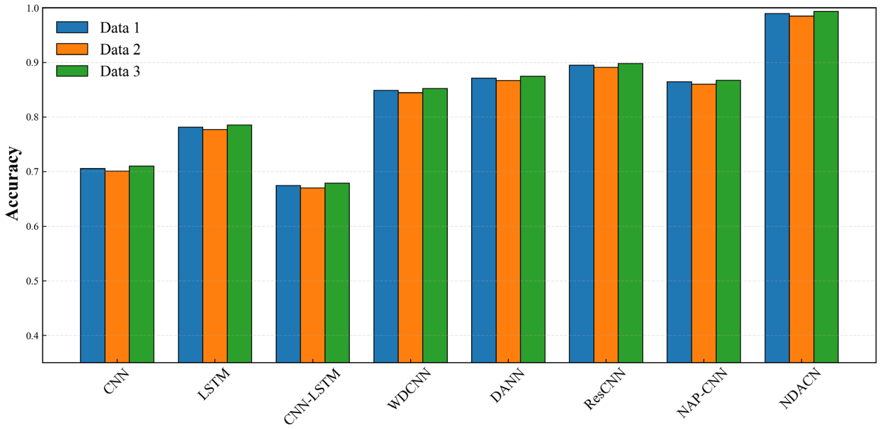


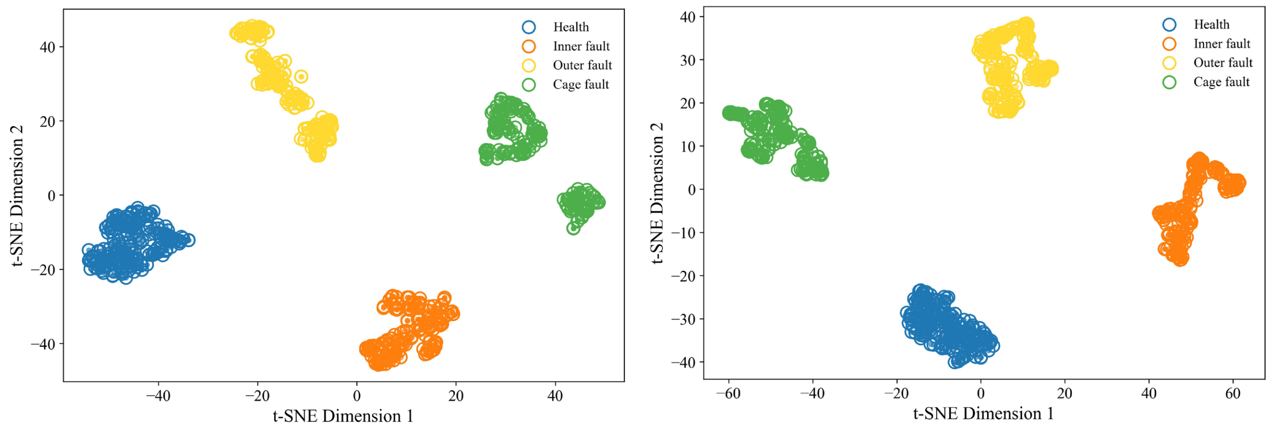
Figure 12. Model accuracy under 4:6 training-test ratio.

Comparison with other models shows that NDACN has better performance stability. NAP-CNN (incorporating NAP de-conditioning) generally outperforms

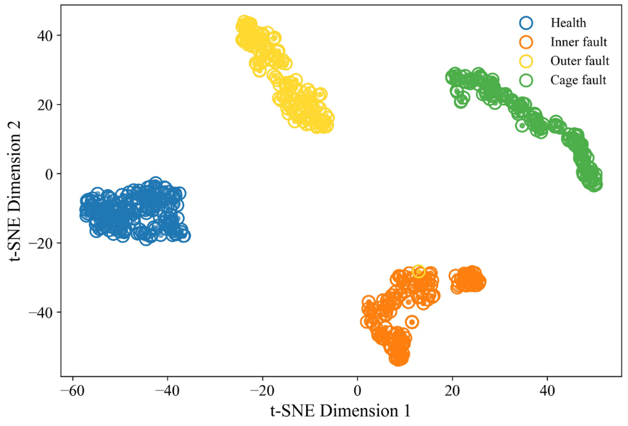
CNN, LSTM, and CNN-LSTM without interference processing, confirming that variable-speed interference is critical to diagnostic accuracy. Although DANN alleviates condition differences better than WDCNN, it lacks targeted fault feature screening and enhancement. NDACN achieves a good balance between precision and recall, with an F1-score above 0.96 and a precision-recall difference of less than 0.5% across all three datasets.

Through the collaborative framework of NAP de-conditioning, DFT dynamic feature screening, attention enhancement, and convolutional network diagnosis, NDACN effectively addresses key issues (interference and low feature discrimination) in variable-speed bearing fault diagnosis, and its performance is superior to other comparison models under different variable-speed conditions and training set ratios.

To verify the enhancement and discrimination effect of the attention enhancement module in NDACN on fault features, the fault feature channel information processed by this module is reduced to a two-dimensional space using the t-SNE algorithm for visual analysis. Feature visualization is completed in Data 1, Data 2, and Data 3 respectively, and the results are shown in **Figure 13**.



(a) Data1 (b) Data2

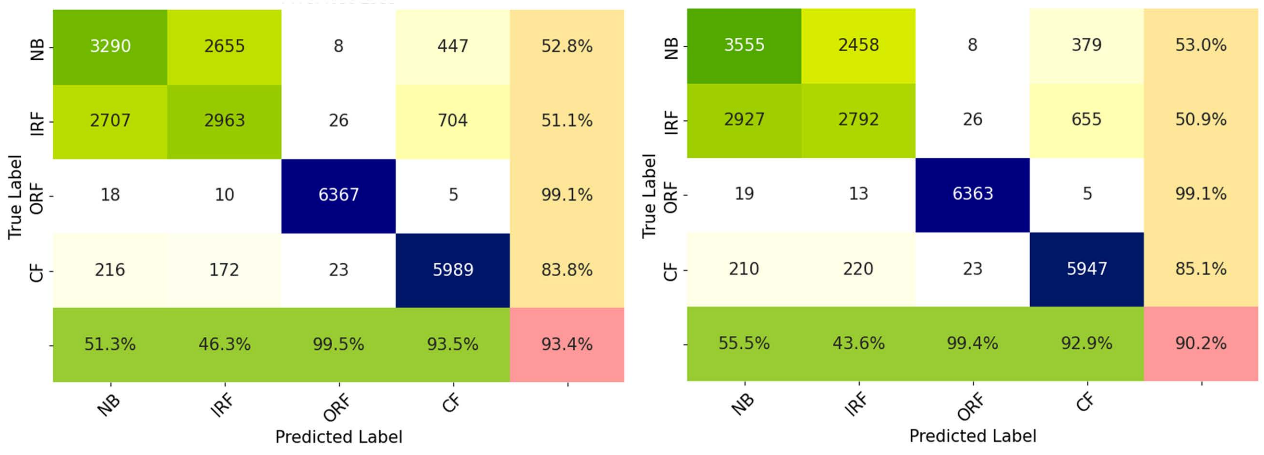


(c) Data3

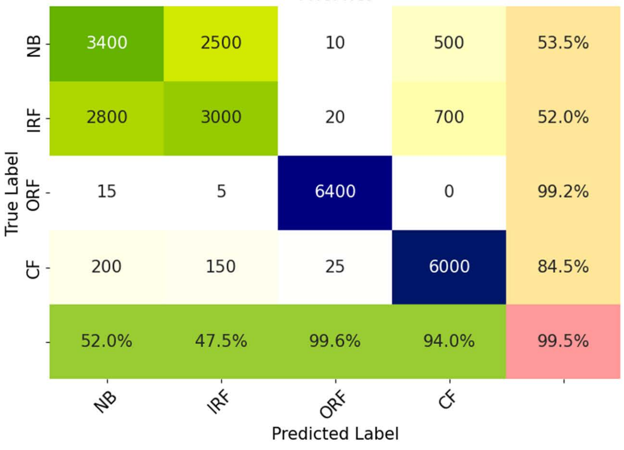
Figure 13. Visualization of NDACN after attention enhancement under different data.

In the three datasets, healthy samples are clustered in distinct regions and far from fault clusters, while different fault samples exhibit regular distributions. This demonstrates that the attention enhancement module can effectively capture fault specificity, enhance inter-class differences via adaptive weight assignment, and reduce the influence of speed fluctuations and noise. Compared with NAP projection only, the addition of the attention mechanism significantly improves intra-class compactness and inter-class separation, with clearer clustering boundaries and more concentrated distributions. On the basis of NAP-based interference elimination, the module further focuses on key fault features and improves the quality of feature representation.

As shown in **Figure 14**, the confusion matrices of the three Data all show high classification accuracy, with the proportion of samples on the diagonal exceeding 84%. Among them, the classification effect of outer race fault (ORF) is relatively optimal: the recall rate is 99.1% in Data 1, 99.1% in Data 2, and 99.2% in Data 3, with only a few misclassified samples. In terms of precision, the precision of ORF is higher than 99.4% in the three Data, and the precision of cage fault (CF) exceeds



(a) Data1 (b) Data2



(c) Data3

Figure 14. Confusion matrices of NDACN under different data.

93.5%, indicating that the model has high prediction reliability for these two types of faults, and the probability of misclassification into other categories is low. The precision of normal state (NB) and inner race fault (IRF) is relatively low, mainly because samples from other categories are easily misclassified into these two types. This phenomenon is basically consistent with the recall analysis results and conforms to the actual characteristics of fault diagnosis under variable-speed conditions.

From the confusion matrix results of the three Data, it can be seen that NDACN can achieve accurate identification of various faults under different working conditions. In particular, it exhibits nearly perfect classification ability for outer race faults with clear features, and can maintain high recognition accuracy for weak cage faults. Although there is a certain proportion of misclassification between the healthy state and inner race faults, the overall diagnostic performance is still superior to the comparison models.

5. Conclusions and Future Work

Aiming at the problems of non-stationary vibration signals, fault features susceptible to interference, and low feature discrimination in rolling bearing fault diagnosis under variable-speed conditions in modern industry, this paper proposes a Projection Filtering-based Convolutional Network (NDACN) for rolling bearing fault diagnosis. Through the synergistic effect of Non-redundant Attribute Projection (NAP), Dynamic Filtering Threshold (DFT), and parameterized attention convolution enhancement mechanism, the proposed method effectively strips working condition interference, screens effective features, strengthens core information representation, and alleviates the shortcomings of traditional methods. Comparative Data on three different variable-speed bearing Datasets show that the proposed NDACN method exhibits better diagnostic accuracy and stability than existing mainstream diagnostic models under variable-speed conditions, verifying the effectiveness and practicability of the method. The research in this paper provides a feasible technical approach for rolling bearing fault diagnosis in complex variable-speed scenarios, and also provides useful references for similar rotating machinery fault diagnosis research.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Li, S., Peng, Y., Shen, Y., Zhao, S., Shao, H., Bin, G., *et al.* (2024) Rolling Bearing Fault Diagnosis under Data Imbalance and Variable Speed Based on Adaptive Clustering Weighted Oversampling. *Reliability Engineering & System Safety*, **244**, Article ID: 109938. https://doi.org/10.1016/j.ress.2024.109938

[2] Hu, C.S., Li, G.L., Zhao, Y. and Cheng, F.J. (2022) Summary of Fault Diagnosis Methods for Rolling Bearings Under Variable Working Conditions. *Journal of Computer Engineering & Applications*, **58**, 26.

[3] Ding, Y., Jia, M., Zhuang, J., Cao, Y., Zhao, X. and Lee, C. (2023) Deep Imbalanced Domain Adaptation for Transfer Learning Fault Diagnosis of Bearings under Multiple Working Conditions. *Reliability Engineering & System Safety*, **230**, Article ID: 108890. https://doi.org/10.1016/j.ress.2022.108890

[4] Luo, Y., Shi, J., Tan, J., Ren, Z., Wan, J., Safran, M., *et al.* (2025) An Ensemble Data-Model-Label Three-Level Regularization Framework for Imbalanced Intelligent Fault Diagnosis. *IEEE Transactions on Reliability*, **74**, 3884-3896. https://doi.org/10.1109/tr.2024.3415117

[5] Ren, Z., Lin, T., Feng, K., Zhu, Y., Liu, Z. and Yan, K. (2023) A Systematic Review on Imbalanced Learning Methods in Intelligent Fault Diagnosis. *IEEE Transactions on Instrumentation and Measurement*, **72**, 1-35. https://doi.org/10.1109/tim.2023.3246470

[6] Tiboni, M., Remino, C., Bussola, R. and Amici, C. (2022) A Review on Vibration-Based Condition Monitoring of Rotating Machinery. *Applied Sciences*, **12**, Article 972. https://doi.org/10.3390/app12030972

[7] Samanta, B., Al-Balushi, K.R. and Al-Araimi, S.A. (2003) Artificial Neural Networks and Support Vector Machines with Genetic Algorithm for Bearing Fault Detection. *Engineering Applications of Artificial Intelligence*, **16**, 657-665. https://doi.org/10.1016/j.engappai.2003.09.006

[8] Lu, S., Yan, R., Liu, Y. and Wang, Q. (2019) Tacholess Speed Estimation in Order Tracking: A Review with Application to Rotating Machine Fault Diagnosis. *IEEE Transactions on Instrumentation and Measurement*, **68**, 2315-2332. https://doi.org/10.1109/tim.2019.2902806

[9] Sun, Y., Li, S. and Wang, X. (2021) Bearing Fault Diagnosis Based on EMD and Improved Chebyshev Distance in SDP Image. *Measurement*, **176**, Article ID: 109100. https://doi.org/10.1016/j.measurement.2021.109100

[10] Kankar, P.K., Sharma, S.C. and Harsha, S.P. (2011) Fault Diagnosis of Ball Bearings Using Continuous Wavelet Transform. *Applied Soft Computing*, **11**, 2300-2312. https://doi.org/10.1016/j.asoc.2010.08.011

[11] Joseph, D.S., Pawar, P.M. and Pramanik, R. (2022) Intelligent Plant Disease Diagnosis Using Convolutional Neural Network: A Review. *Multimedia Tools and Applications*, **82**, 21415-21481. https://doi.org/10.1007/s11042-022-14004-6

[12] Petmezas, G., Haris, K., Stefanopoulos, L., Kilintzis, V., Tzavelis, A., Rogers, J.A., *et al.* (2021) Automated Atrial Fibrillation Detection Using a Hybrid CNN-LSTM Network on Imbalanced ECG Datasets. *Biomedical Signal Processing and Control*, **63**, Article ID: 102194. https://doi.org/10.1016/j.bspc.2020.102194

[13] Li, T., Zhao, Z., Sun, C., Cheng, L., Chen, X., Yan, R., *et al.* (2022) WaveletKernelNet: An Interpretable Deep Neural Network for Industrial Intelligent Diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, **52**, 2302-2312. https://doi.org/10.1109/tsmc.2020.3048950

[14] Zong, X., Yang, R., Wang, H., Du, M., You, P., Wang, S., *et al.* (2022) Semi-Supervised Transfer Learning Method for Bearing Fault Diagnosis with Imbalanced Data. *Machines*, **10**, Article 515. https://doi.org/10.3390/machines10070515

[15] Chen, Y., Yang, Y., Zhang, J., Yang, J., Huang, Y. and Wang, W. (2025) A Fault Prediction Method Based on SG-ResCNN for the Sliding Bearings of Fire Pumps. *Proceedings of the* 2025 2*nd International Conference on Image Processing, Intelligent Control and Computer Engineering*, Zhengzhou, 25-27 July 2025, 104-112. https://doi.org/10.1145/3768184.3768204

[16] Tong, Z., Li, W., Zhang, B., Jiang, F. and Zhou, G. (2018) Bearing Fault Diagnosis

under Variable Working Conditions Based on Domain Adaptation Using Feature Transfer Learning. *IEEE Access*, **6**, 76187-76197. https://doi.org/10.1109/access.2018.2883078

[17] Han, S., Zhong, X., Shao, H., Xu, T., Zhao, R. and Cheng, J. (2021) Novel Multi-Scale Dilated CNN-LSTM for Fault Diagnosis of Planetary Gearbox with Unbalanced Samples under Noisy Environment. *Measurement Science and Technology*, **32**, Article ID: 124002. https://doi.org/10.1088/1361-6501/ac1b43

[18] Yuan, B., Li, Y. and Chen, S. (2025) Efficient Gearbox Fault Diagnosis Based on Improved Multi-Scale CNN with Lightweight Convolutional Attention. *Sensors*, **25**, Article 2636. https://doi.org/10.3390/s25092636

[19] Wang, M. and Wang, Z. (2026) A Novel CNN-Transformer Integrated with Cross-Attention Mechanism for Intelligent Diagnosis of Rolling Bearing Faults. *Digital Signal Processing*, **175**, Article ID: 105996. https://doi.org/10.1016/j.dsp.2026.105996

[20] Wu, H., Li, J., Zhang, Q., Tao, J. and Meng, Z. (2022) Intelligent Fault Diagnosis of Rolling Bearings under Varying Operating Conditions Based on Domain-Adversarial Neural Network and Attention Mechanism. *ISA Transactions*, **130**, 477-489. https://doi.org/10.1016/j.isatra.2022.04.026

[21] Yuan, Z., Li, X., Liu, S. and Ma, Z. (2023) A Recursive Multi-Head Graph Attention Residual Network for High-Speed Train Wheelset Bearing Fault Diagnosis. *Measurement Science and Technology*, **34**, Article ID: 065108. https://doi.org/10.1088/1361-6501/acb609

[22] Khandelwal, D., Anand, P., Ray, M. and R. G., S. (2025) Fault Detection in Electrical Power Systems Using Attention-Gru-Based Fault Classifier (AGFC-Net). *Scientific Reports*, **15**, Article No. 24133. https://doi.org/10.1038/s41598-025-06493-w

[23] Guo, H., Ling, X., Zhang, X. and Ren, X. (2026) A Novel Hierarchical Graph Neural Network with Heterogeneous Multisensor Data for Fault Diagnosis of Complex Mechanical Equipment. *IEEE Transactions on Instrumentation and Measurement*, **75**, 1-10. https://doi.org/10.1109/tim.2026.3666015

[24] Jiang, H., Chen, J. and Dong, G. (2016) Hidden Markov Model and Nuisance Attribute Projection Based Bearing Performance Degradation Assessment. *Mechanical Systems and Signal Processing*, **72**, 184-205. https://doi.org/10.1016/j.ymssp.2015.10.003

[25] Yang, D., Lv, Y., Yuan, R., Yang, K. and Zhong, H. (2022) A Novel Vibro-Acoustic Fault Diagnosis Method of Rolling Bearings via Entropy-Weighted Nuisance Attribute Projection and Orthogonal Locality Preserving Projections under Various Operating Conditions. *Applied Acoustics*, **196**, Article ID: 108889. https://doi.org/10.1016/j.apacoust.2022.108889

[26] Peng, Y., Li, Z., He, K., Liu, Y., Li, Q. and Liu, L. (2019) Broadband Mode Decomposition and Its Application to the Quality Evaluation of Welding Inverter Power Source Signals. *IEEE Transactions on Industrial Electronics*, **67**, 9734-9746. https://doi.org/10.1109/tie.2019.2955429

[27] Potapczynski, A., Loaiza-Ganem, G. and Cunningham, J.P. (2020) Invertible Gaussian Reparameterization: Revisiting the Gumbel-Softmax. *Advances in Neural Information Processing Systems*, **33**, 12311-12321.