DOI: 10.23977/jemm.2026.110117
ISSN 2371-9133 Vol. 11 Num. 1

Multi-Source Information Fault Diagnosis Method for Rolling Bearing Based on Multi-View Clustering

Jie Shen

School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai, 200093, China

Keywords: variational autoencoder; multi-view clustering; rolling bearing; fault diagnosis

Abstract: This paper addresses the difficulty that decision-level fusion in unsupervised fault diagnosis of rolling bearing multi-source data fails to effectively exploit the complementary information among multi-source data. A rolling bearing fault diagnosis method based on multi-view clustering (MVC) is proposed. First, a variational autoencoder (VAE) is employed to extract latent features from multi-source signals. On this basis, a joint matrix that integrates multi-source data features and a high-confidence matrix containing consistency information are constructed. A low-rank approximation method is then used to mine complementary information within multi-source information and enhance structural consistency. Finally, accurate identification of unsupervised fault diagnosis is achieved. Experimental results show that the proposed method can effectively utilize the complementary information among multiple views, overcome dependence on a single data source, and exhibit excellent robustness and stability.

1. Introduction

With the rapid emergence and vigorous development of the Industrial Internet and artificial intelligence, the operation and maintenance mode of mechanical equipment is gradually transforming toward data-driven intelligent operation and maintenance. Meanwhile, due to the widespread application of sensor technology, the Internet of Things, and cloud computing, multi-source state data can be collected in real time during equipment operation. This enables fault diagnosis technology of mechanical equipment to enter the era of big data, providing a substantial data foundation and technical support for comprehensive health monitoring of mechanical equipment[1]. In this context, unsupervised learning methods based on multi-source information have become an important technical approach in health monitoring and fault diagnosis of mechanical equipment, as they do not rely on large amounts of labeled data for training[2].

Unsupervised learning methods based on multi-source information fusion can be classified into data-level fusion, feature-level fusion, and decision-level fusion according to the hierarchy and stage of information processing[3]. Data-level fusion directly integrates raw data. Its advantage is minimal information loss, while its disadvantages include high requirements for data quality, high computational cost, and sensitivity to noise. For example, Fu et al. [4]proposed a shaft system fault diagnosis method for hydropower units based on dual-channel data main vibration feature extraction. In this method, principal component analysis is first used to perform data-level fusion on dual-channel time-domain vibration signals of the shaft system to obtain the synthesized vibration in the direction of maximum vibration intensity. Feature extraction is then conducted on the synthesized vibration, followed by classification using a support vector machine, thereby achieving fault diagnosis. Feature-level fusion extracts features independently from each data source and then fuses these features. Its advantage is high computational efficiency, while its disadvantage is partial information loss. For instance, Zhang et al. [5]adopted a parallel structure in which convolutional neural networks and gated recurrent units simultaneously extract fault features from raw vibration signals of a gearbox. The extracted feature vectors from the two channels are then merged into a fused feature vector and input into SoftMax for fault classification. Decision-level fusion independently models each data source to obtain preliminary results, and then fuses these results according to certain rules to provide collaborative diagnosis or decision-making. Its advantages include good fault tolerance, minimal impact from failure or performance degradation of a single channel, and very high computational efficiency. However, it suffers from significant information loss. For example, Liu et al. [6]proposed a knowledge-driven method to extract vibration signal features from a motor bearing fault dataset. An optimal feature subset is selected using a recursive feature elimination algorithm. Six classifiers are then employed, and decision fusion methods including voting, stacking, and fusion strategies are combined to establish a fault diagnosis model.

Among these methods, decision-level fusion can avoid problems caused by heterogeneity of multi-source data, but it tends to ignore the potential relationships among different data sources, resulting in insufficient diagnostic accuracy. In the field of image recognition, the commonly used multi-view clustering (MVC) method can effectively mine the potential relationships among different data sources and obtain consistent results from different views. The main idea of MVC is to find a fused graph that contains the principal information of all views. A graph partitioning algorithm is then applied to the fused graph to produce the final clustering results[7]. For example, Lu et al. [8]used an autoencoder to extract low-level features for each view. On this basis, a contrastive learning method is adopted to maximize the consistency of information among views, thereby learning high-level features and semantic labels for each view. A multi-attention mechanism is then introduced to fuse high-level features from each view, resulting in more accurate multi-view feature representations. Finally, clustering based on three types of decisions is performed on the fused features and semantic labels, yielding three soft clustering results corresponding to core, boundary, and trivial regions, which addresses the uncertainty between objects and clusters. Yang et al. [9]proposed a global hybrid alignment and cross-contrastive learning framework. This framework adaptively learns and fuses features from different views through a view-specific autoencoding network and an attention fusion layer, thereby obtaining a more comprehensive consensus representation. By combining a global hybrid alignment strategy with cross-contrastive learning, the framework fully utilizes view-specific information during clustering. It further strengthens common features across multiple views, achieving an effective balance between view consistency and specificity.

In decision-level fusion, the final result depends on high-quality features extracted from each view. Therefore, effective extraction of key features is particularly important. Deep learning models have achieved great success in fields such as image recognition because they can autonomously learn data features[10]. Among them, the variational autoencoder (VAE) can map high-dimensional and complex data into a low-dimensional space, thereby extracting key features, and has attracted extensive attention [11]. For example, Yin et al. [12]proposed a variational autoencoder-based model for bearing health condition assessment oriented to high-entropy feature data. Wang et al. [13]proposed a defect classification method for contact networks based on a variational autoencoder, named DefVAE. This method uses latent features output by the VAE encoder to determine the feature distribution of known defect samples. A large amount of defect data is generated through distribution space resampling and decoding, thereby compensating for insufficient samples.

To address the above issues in decision-level fusion, and inspired by MVC and the VAE, this paper proposes a fault diagnosis model for rolling bearings based on multi-view clustering and multi-source information fusion. The method first employs a VAE to independently perform feature extraction on multi-source signals of rolling bearings, such as vibration, temperature, and rotational speed, thereby obtaining high-quality features. These features are then input into a multi-view clustering framework. Complementary information among different views is mined through low-rank approximation, which effectively resolves the difficulty in utilizing potential information among different signal sources in decision-level fusion. As a result, collaborative diagnosis of multi-source signals in an unsupervised manner is achieved, and fault recognition accuracy is improved. This provides an efficient and reliable technical approach for health management of rolling bearings under complex operating conditions.

2. Fundamental Theory

2.1. VAE

The VAE originates from the autoencoder (AE) and is an unsupervised learning method. It typically consists of a three-layer neural network, including an input layer, a hidden layer, and an output layer[14].

Based on Bayesian variational inference theory, for independently and identically distributed data $P(x)$ in the observation space, a low-dimensional feature z is introduced. The model can then be expressed as[15]:

$$P(x)=\int_{Z}P(x\mid z)P(z)dz \qquad (1)$$

The VAE employs a neural network to train a simple distribution $Q(z)$ to approximate $P(x\mid z)$, and the Kullback-Leibler divergence (KL Divergence) is used to measure the similarity between the two distributions.

The structure of the VAE model is shown in Fig. 1. The original data X are mapped by the encoder E to the mean μ and variance σ of the latent variable Z. Then, combined with a random vector e that follows a standard normal distribution, the latent variable Z is generated. Finally, the decoder D reconstructs the approximate original data Y.

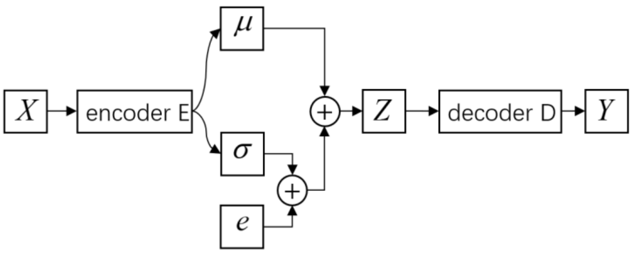

Fig. 1 Structure of VAE

2.2. MVC

Multi-view clustering (MVC) is an unsupervised learning method. It does not rely on labels. By mining the consistency and complementarity among multiple views, the data are partitioned into different clusters, such that samples within the same cluster have high similarity, while samples in different clusters have low similarity [16].

Given a set of multi-view data $X=\{X^1,X^2,\cdots,X^V\}$, where X^v, $v=1,2,\cdots,V$, represents the data of the v-th view, the structure of the multi-view clustering model is shown in Fig. 2. The MVC model first obtains the similarity matrix of each single view. A commonly used method is to employ a

Gaussian kernel function to measure the similarity A between two data points [17]. For the v-th view $X^v = \{x_1^v, x_2^v, \ldots, x_n^v\} \in \mathbb{R}^{d\times n}$, where n denotes the number of samples and d denotes the dimensionality of the sample data, the similarity between sample points x_i^v and x_j^v is defined as:

$$S_{ij}^v = \exp\left(-\frac{\left\|x_i^v - x_j^v\right\|_2^2}{\delta}\right) \quad (2)$$

Then, based on the multi-view similarity matrix $S = \{S^1, S^2, \cdots, S^V\}$, the joint matrix A and the high-confidence matrix B are constructed. A commonly used construction method is as follows:

$$\begin{aligned} &A = \sum_{v=1}^{V} S^v, \\ &B_{ij} = \begin{cases} 1 & S_i^v = S_j^v, \forall v = 1,\ldots,V \\ 0 & otherwise \end{cases} \end{aligned} \quad (3)$$

Finally, the joint matrix A and the high-confidence matrix B are stacked to form a third-order tensor $\mathcal{P} \in \mathbb{R}^{n\times n\times 2}$, where $\mathcal{P}(:,:,1) = B$, $\mathcal{P}(:,:,2) = A$. The final result C is then obtained by solving the model.

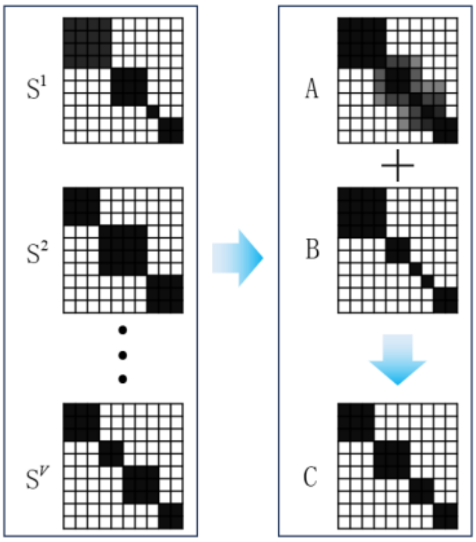

Fig. 2 Multi-view clustering model

3. Proposed Method

3.1. Fault Feature Extraction Method for Rolling Bearing Based on VAE

The fault feature extraction method for rolling bearings based on VAE is a model that mines deep latent features from raw vibration signals through a probabilistic generative framework. The structure of the VAE is shown in Fig. 3. First, multi-source signal samples of rolling bearings are collected as $X = \{X^1, X^2, \cdots, X^V\}$, and feature extraction is performed separately for each single signal source.

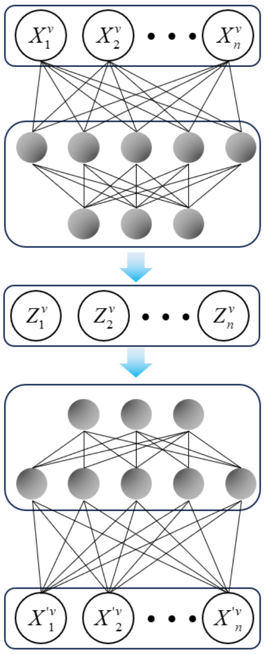

Fig. 3 VAE structure

After the collected multi-source signal samples are input into the VAE, the encoder maps the high-dimensional input data into a low-dimensional latent space, obtaining latent variables $Z^v=\{Z_1^v,Z_2^v,\cdots Z_n^v\}$. In order to retain key fault information while reducing feature dimensionality and avoiding redundancy, partial components of Z^v are selected as the extracted fault features. These components contain essential information about the bearing health condition in multi-source signals. After extracting features from all signal sources, a feature set of multi-source data is formed as $F=\{F^1,F^2,\cdots,F^V\}$.

3.2. Multi-Source Information Fusion Method Based on MVC

Spectral clustering is first performed separately on each source data feature set $F=\{F^1,F^2,\cdots,F^V\}$ to obtain preliminary clustering results $\{C_1^v,C_2^v,\ldots,C_n^v\}$, where n is the number of samples. The optimal number of clusters is determined by the Calinski–Harabasz index.

The similarity matrix within each source data is calculated according to Equation (2), where δ is an adjustable parameter. However, in multi-view data, the measurement differs among views. Therefore, based on the Gaussian kernel function, the similarity matrix between samples in a single view is calculated as follows:

$$S_{ij}^v=\exp\left(-\frac{\left\| f_i^v-f_j^v\right\|_2^2}{\sigma\max\{\left\| f_i^v-f_j^v\right\|_2^2\}}\right) \quad (4)$$
$$i=1,\cdots,n,\ j=1,\cdots,n,v=1,\cdots V$$

The proposed method requires the construction of a joint matrix A and a high-confidence matrix B. The joint matrix integrates feature matrices of multi-source information and contains abundant information, but its reliability is relatively low. The high-confidence matrix represents the consensus information of multi-source data for the same sample. It has high reliability but contains limited information. Then, based on the low-rank approximation method, the high-confidence matrix is used to constrain the joint matrix toward higher reliability, while the joint matrix is used to enrich the high-confidence matrix. In this paper, the joint matrix and the high-confidence matrix are constructed based on the feature matrices and clustering results obtained from spectral clustering of each source data.

The joint matrix is constructed based on the similarity matrices of different views as follows:

$$\mathrm{A} = \delta^v \sum_{v=1}^{V} \mathrm{S}^v \tag{5}$$

where δ^v denotes the weight of the v-th channel. Signals from different channels have different importance, and the weight is calculated as follows:

$$\delta^v = \frac{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} \| f_i^v - f_j^v \| \mathrm{S}_{ij}^v}}{\sum_{v=1}^{V}\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} \| f_i^v - f_j^v \| \mathrm{S}_{ij}^v}} \tag{6}$$

The high-confidence matrix is calculated based on the clustering results of a single channel. When all signal sources assign two samples to the same cluster, the result is considered highly reliable. Otherwise, it is considered unreliable. The high-confidence matrix is defined as:

$$\mathrm{B}_{ij} = \begin{cases} 1 & c_i^v = c_j^v, \forall v = 1, \ldots, V \\ 0 & otherwise \end{cases} \tag{7}$$

The main idea of the proposed method is to utilize the high reliability of the high-confidence matrix to refine the joint matrix and improve its reliability. At the same time, the rich information in the joint matrix is propagated to the high-confidence matrix, guiding the process of refining the joint matrix.

The joint matrix A and the high-confidence matrix B are combined to form a third-order tensor $\mathcal{P} \in \mathbb{R}^{n \times n \times 2}$, where $\mathcal{P}(:, :, 1) = \mathrm{B}$, $\mathcal{P}(:, :, 2) = \mathrm{A}$. The joint matrix measures the similarity between samples based on the extracted feature matrix, while the high-confidence matrix reflects the relationships between samples obtained after clustering the feature matrix. The difference between these two matrices lies in the amount of information and reliability. Therefore, the third-order tensor $\mathcal{P}$ composed of A and B exhibits a low-rank structure. The proposed method can be mathematically formulated as a constrained low-rank approximation problem:

$$\begin{aligned} &\min_{\mathcal{P},\mathrm{E}} \| \mathcal{P} \|_* + \lambda \| \mathrm{E} \|_F^2 \\ &\text{s.t. } \mathcal{P}(i,j,1) = \mathrm{A}(i,j), \text{ if } \mathrm{A}(i,j) = 1, \\ &\quad \mathcal{P}(:,:,2) + \mathrm{E} = \mathrm{B}, \\ &\quad 0 \le \mathcal{P}(:,:,1) \le 1, \forall i, j \\ &\quad 0 \le \mathcal{P}(:,:,2) \le 1, \forall i, j \end{aligned} \tag{8}$$

where $\lambda > 0$ is a coefficient that balances the error matrix, and E adopts the Frobenius norm to avoid trivial solutions. By optimizing Equation (8), the limited but highly reliable information in B is expected to be propagated into the joint matrix A, while the information in A is used to supplement B. Finally, an optimized joint matrix $\mathcal{P}^*(:,:,2)$ is obtained. Then, $\mathcal{P}^*(:,:,2)$ is used as the joint matrix for spectral clustering to perform secondary clustering.

3.3. Problem-Solving Methodology

Equation (8) is solved based on the alternating direction method of multipliers (ADMM). An auxiliary tensor $\mathcal{T}$ is introduced to separate the tensor in the objective function. Then, Equation (8) can be rewritten as:

$$
\begin{aligned}
&\min_{\mathcal{P},\mathcal{T},\mathrm{E}} \|\mathcal{T}\|_{*} + \lambda \|\mathrm{E}\|_F^2 \\
&\text{s.t. } \mathcal{P}(i,j,1) = \mathrm{A}(i,j), \text{ if } \mathrm{A}(i,j)=1, \\
&\quad \mathcal{P}(:,:,2) + \mathrm{E} = \mathrm{B}, \\
&\quad 0 \le \mathcal{P}(:,:,1) \le 1, \forall i,j \\
&\quad 0 \le \mathcal{P}(:,:,2) \le 1, \forall i,j \\
&\quad \mathcal{T} = \mathcal{P}
\end{aligned} \tag{9}
$$

To handle the equality constraints, two Lagrange multipliers are introduced. The augmented Lagrangian function of Equation (9) is expressed as:

$$
\begin{aligned}
&L_\mu(\mathcal{P},\mathcal{T},\mathrm{E},\mathcal{L}、\mathrm{L}) = \|\mathcal{P}\|_{\circledast} + \lambda \|\mathrm{E}\|_F^2 \\
&+\langle \mathcal{L}, (\mathcal{P}-\mathcal{T}) \rangle + \frac{\mu}{2}\|\mathcal{P}-\mathcal{T}\|_F^2 \\
&+\langle \mathrm{L}, (\mathcal{P}(:,:,2) + \mathrm{E} - \mathrm{B}) \rangle \\
&+\frac{\mu}{2}\|\mathcal{P}(:,:,2) + \mathrm{E} - \mathrm{B}\|_F^2 \\
&\text{s.t. } \mathcal{P}(i,j,1) = \mathrm{A}(i,j), \text{ if } \mathrm{A}(i,j)=1, \\
&\quad 0 \le \mathcal{P}(:,:,1) \le 1, \forall i,j \\
&\quad 0 \le \mathcal{P}(:,:,2) \le 1, \forall i,j
\end{aligned} \tag{10}
$$

where $\mu > 0$ is a penalty parameter. The original optimization problem is then transformed into the following form:

$$
\begin{aligned}
&\min_{\mathcal{P},\mathcal{T},\mathrm{E}} \|\mathcal{P}\|_{\circledast} + \lambda \|\mathrm{E}\|_F^2 + \frac{\mu}{2}\left\|\mathcal{P}-\mathcal{T}+\frac{\mathcal{L}}{\mu}\right\|_F^2 \\
&+\frac{\mu}{2}\left\|\mathcal{P}(:,:,2) + \mathrm{E} - \mathrm{B} + \frac{\mathrm{Y}}{\mu}\right\|_F^2 \\
&\text{s.t. } \mathcal{P}(i,j,1) = \mathrm{A}(i,j), \text{ if } \mathrm{A}(i,j)=1, \\
&\quad 0 \le \mathcal{P}(:,:,1) \le 1, \forall i,j \\
&\quad 0 \le \mathcal{P}(:,:,2) \le 1, \forall i,j
\end{aligned} \tag{11}
$$

Then, Equation (11) can be solved by alternately optimizing the subproblems of $\mathcal{P},\mathcal{T}$ and E to obtain the optimal solution.

For the subproblem of $\mathcal{P}$, irrelevant terms are removed, and it can be expressed as:

$$
\min_{\mathcal{P}} \frac{\mu}{2}\left\|\mathcal{P}-\mathcal{T}+\frac{\mathcal{L}}{\mu}\right\|_F^2 + \frac{\mu}{2}\left\|\mathcal{P}(:,:,2) + \mathrm{E} - \mathrm{B} + \frac{\mathrm{Y}}{\mu}\right\|_F^2 \tag{12}
$$

The two slices of $\mathcal{P}$ are solved separately as follows:

$$
\begin{aligned}
&\mathcal{P}(i,j,1) = \begin{cases} \mathrm{B}(i,j) & \text{if } \mathrm{B}(i,j)=1, \\ 0 & \text{if } \mathrm{T}_1(i,j) \le 0 \\ 1 & \text{if } \mathrm{T}_1(i,j) \ge 1 \\ \mathrm{T}_1(i,j) & \text{if } 0 \le \mathrm{T}_1(i,j) \le 1 \end{cases} \\
&\mathcal{P}(i,j,2) = \begin{cases} 0 & \text{if } \mathrm{T}_2(i,j) \le 0 \\ 1 & \text{if } \mathrm{T}_2(i,j) \ge 1 \\ \mathrm{T}_2(i,j) & \text{if } 0 \le \mathrm{T}_2(i,j) \le 1 \end{cases} \\
&\mathrm{T}_1(i,j) = \mathcal{T}(:,:,1) - \mathcal{L}(:,:,1)/\mu \\
&\mathrm{T}_2(i,j) = (\mathcal{T}(:,:,2) - \mathcal{L}(:,:,2)/\mu + \mathrm{B} - \mathrm{E} - \mathrm{L}/\mu)/2
\end{aligned} \tag{13}
$$

For the subproblem of $\mathcal{T}$, irrelevant terms are removed, and it can be expressed as:

$$
\min_{\mathcal{T}} \frac{1}{\mu}\|\mathcal{T}\|_{\otimes} + \frac{1}{2}\|\mathcal{T}-\mathcal{O}\|_F^2 \tag{14}
$$

Fast Fourier Transform (FFT) and singular value decomposition (SVD) are applied to tensor $\mathcal{O}$ along the frontal slices to obtain the tensor SVD representation. The solution is given by:

$$\mathcal{T} = \mathcal{D}_{\mu}(\mathcal{O}) \quad (15)$$

where $\mathcal{D}_{\tau}(\mathcal{O}) = \mathcal{U} * \mathcal{S}_{\tau} * \mathcal{V}^{T}$, $\mathcal{S}_{\tau} = \mathrm{ifft}((\bar{\mathcal{S}} - \tau)_{+}, [], 3)$, and $t_{+} = \max(t, 0)$.

For the subproblem of E, irrelevant terms are removed, and it can be expressed as:

$$\min_{\mathrm{E}} \lambda \| \mathrm{E} \|_F^2 + \frac{\mu}{2} \left\| \mathcal{P}(:,:,2) + \mathrm{E} - \mathrm{A} + \frac{\Lambda_2}{\mu} \right\|_F^2 \quad (16)$$

Since Equation (16) is a quadratic function of E, its derivative can be set to zero to obtain the global minimum, which is:

$$\mathrm{E} = \frac{\mu \mathrm{A} - \Lambda_2 - \mu \mathcal{P}(:,:,2)}{2\lambda + \mu} \quad (17)$$

The main computational procedure is as follows:

Computational Procedure

Input: collected multi-channel signal features $X = \{X^1, X^2, \cdots, X^V\}$

Output: clustering results

1: Compute the spectral clustering feature matrix $\mathrm{F}^v \in \mathbb{R}^{n \times k_v}$ and clustering results $\{C_1^v, C_2^v, \ldots, C_n^v\}$ for each single channel.

2: Calculate the joint matrix A and the high-confidence matrix B using Eqs. (4), (5), and (7).

3: Initialize $\mathcal{P} = 0$, $\mathcal{T} = 0$, $\mathcal{L} = 0$, $\mathrm{L} = 0$, $\mathrm{E} = 0$, and $\mu = 10^{-3}$.

4: While not converged,

update $\mathcal{P}$ using Equation (13)
update $\mathcal{T}$ using Equation (15)
and update E using Equation (17).

end

5: Use $\mathcal{T}(:,:,2)$ as the joint matrix for spectral clustering to perform secondary clustering.

3.4. Process of the Proposed Method

The process of the rolling bearing health monitoring method based on MVC is shown in Fig. 4.

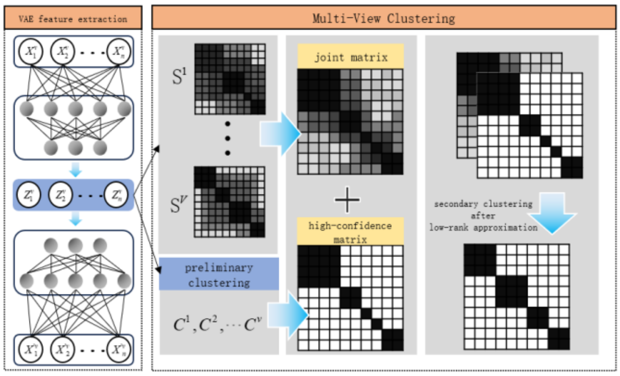

Fig. 4 The proposed method flow

The specific steps are as follows:

Step 1: Collect multi-source state data during the operation of rolling bearings. The proposed method use VAE to extract low-dimensional latent features that characterize the health condition of rolling bearings, and remove redundant information from the original data.

Step 2: Perform preliminary clustering on feature data from each source to obtain clustering results. The proposed method compute the similarity matrix and the joint matrix for each source based on the feature matrix. Construct the high-confidence matrix according to the clustering results of each source.

Step 3: Stack the joint matrix and the high-confidence matrix to form a third-order tensor. The proposed method extract the low-rank structure of the tensor based on the low-rank approximation method. Then perform secondary clustering to obtain the final results, thereby achieving unsupervised fault diagnosis of rolling bearings.

4. Experimental Validation

4.1. Dataset Description

To verify the effectiveness and robustness of the proposed method for fault diagnosis under unsupervised learning conditions, two experimental datasets are selected for analysis and comparison.

Dataset 1: The bearing data from the Case Western Reserve University Bearing Data Center are used for validation. The bearing test bench is shown in Fig. 5, and the tested faulty bearing model is SKF 6205-2RS JEM.

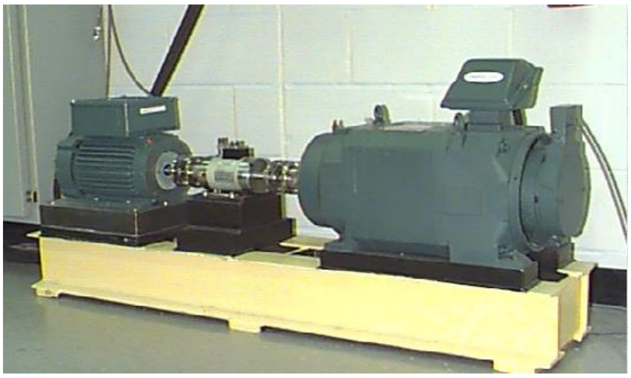

Fig. 5 CWRU bearing fault diagnosis test bench

Vibration signals, including ten categories are selected, which consist of normal condition of the drive-end bearing, inner race faults (007, 014, 021), outer race faults (007, 014, 021), and rolling element faults (007, 014, 021). The sampling frequency is 12 kHz. Each category contains 100 samples, and each sample has a length of 2048 points. Detailed information is shown in Tab. 1.

Tab. 1 CWRU bearing data

Type	Size/inh	Number of samples
Normal	/	100
Inner race	007	
	014	
	021	
Outer race	007	
	014	
	021	
Rolling element	007	
	014	

	021	

Dataset 2: A fixed-axis gearbox fault test bench is constructed for experiments. The layout of the test device is shown in Fig. 6, which mainly consists of a motor, a fixed-axis gearbox, a magnetic powder brake, and a tachometer.

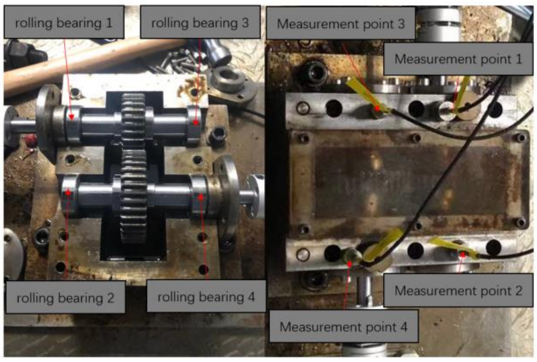

Fig. 6 Fixed-axis gearbox test bench

There are four bearings in the gearbox. Four measurement points are arranged according to the bearing positions in the gearbox, and an acceleration sensor is installed at each measurement point to collect vibration data from the vibration source. Data from different positions are treated as data from different sources. The sampling frequency and duration of the vibration signals are 51.2 kHz and 10 s, respectively. The collected vibration signals include ten categories, which consist of normal condition, inner race faults (small, medium, large), outer race faults (small, medium, large), and rolling element faults (small, medium, large). Each signal group contains four channels according to the sensor arrangement. Each channel contains 512000 data points, resulting in a total of 4×512000 data points. Each signal group is divided into 100 samples, and each sample has a length of 2048 points. Detailed information is shown in Tab. 2.

Tab. 2 Fixed-axis gearbox data

Type	Fault severity	Number of samples
Normal	/	100
Inner race	small	
	medium	
	large	
Outer race	small	
	medium	
	large	
Rolling element	small	
	medium	
	large	

4.2. Experimental Results Analysis

Five metrics are adopted to comprehensively evaluate clustering performance, including the optimal number of clusters, accuracy (ACC), normalized mutual information (NMI), F-score, and adjusted rand index (ARI). For all evaluation metrics except the optimal number of clusters, higher values indicate better clustering performance. The proposed method is compared with independent spectral clustering at each measurement point, as well as LTBPL, t-SVD-MSC, and LTMSC methods, to verify its advantages and applicability under different experimental scenarios.

In the experiments, vibration signals are first transformed by FFT, and then features are extracted using VAE. The feature dimension is set to 16, and the parameter σin Equation (4) of multi-view clustering is set to 0.05. Tab. 3 presents the quantitative evaluation results of the proposed method and comparison methods on Dataset 1 and Dataset 2.

Tab. 3 Quantitative evaluation metrics calculation

Dataset	Fixed-axis gearbox					CWRU				
Method	Optimal clusters	ACC	NMI	Fscore	ARI	Optimal clusters	ACC	NMI	Fscore	ARI
Measurement point 1	5	0.500	0.836	0.667	0.615	**10**	0.891	0.913	0.931	0.841
Measurement point 2	**10**	0.999	0.999	0.999	0.999	**10**	0.999	0.999	0.999	0.999
Measurement point 3	7	0.700	0.905	0.801	0.737	/	/	/	/	/
Measurement point 4	6	0.594	0.871	0.733	0.672	/	/	/	/	/
Proposed method	**10**	**1.000**	**1.000**	**1.000**	**1.000**	10	0.999	0.999	0.999	0.999
LTBPL	8	0.784	0.913	0.843	0.898	**10**	**1.0000**	**1.000**	**1.000**	**1.000**
t-SVD-MSC	9	0.893	0.925	0.943	0.941	**10**	**1.0000**	**1.000**	**1.000**	**1.000**
LTMSC	9	0.852	0.956	0.917	0.898	**10**	**1.0000**	**1.000**	**1.000**	**1.000**

From Tab. 3, it can be observed that all methods obtain the true number of clusters in the CWRU bearing test bench. Both the proposed method and comparison methods achieve excellent performance. Although the proposed method fails to correctly classify a small number of samples, its effectiveness is still validated. In the fixed-axis gearbox dataset, only the proposed method and measurement point 2 obtain the correct number of clusters. The proposed method achieves an accuracy of 100 percent, which is completely consistent with the true labels.

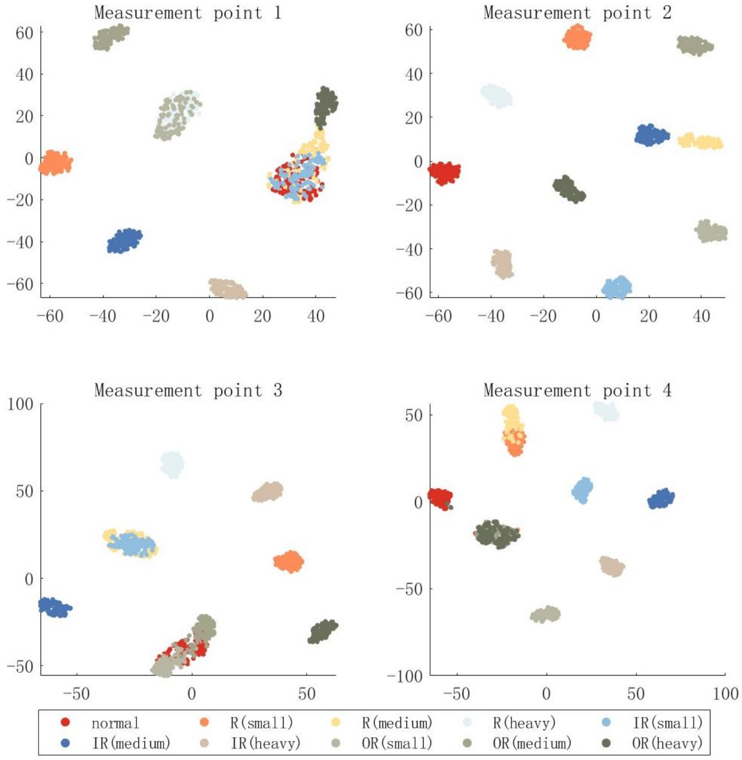

Fig. 7 VAE feature extraction t-SNE

Fig. 7 shows the t-SNE visualization of features extracted by VAE at each measurement point of the fixed-axis gearbox. It can be observed that some categories at measurement points 1, 3, and 4 are partially overlapped, while measurement point 2 shows a classification result that is close to the true distribution. This indicates that the proposed method may be influenced by the high-quality data from measurement point 2, resulting in extremely high accuracy.

To further verify the effectiveness of the proposed method in fusing multi-source data, measurement point 2 is removed and the experiment is conducted again. In this experiment, only the ACC metric is retained as the quantitative indicator. The results are shown in Tab. 4.

Tab. 4 Experimental results

Dataset	Fixed-axis gearbox	
Method	Optimal clusters	ACC
Measurement point 1	5	0.500
Measurement point 3	7	0.700
Measurement point 4	6	0.591
Proposed method	**10**	**1.000**
LTBPL	7	0.871
t-SVD-MSC	8	0.794
LTMSC	8	0.748

From Tab. 4, it can be observed that after removing measurement point 2, the proposed method still achieves 100 percent accuracy, while the comparison methods exhibit a decrease in performance. This demonstrates the advantage of the proposed method in feature utilization and in mining the potential complementary information among multi-source signals. The proposed method does not rely on the superior performance of a single data source. Instead, it can effectively extract key information from multi-source data. Even when some measurement points are missing or their quality deteriorates, the method can still maintain stable diagnostic performance.

5. Conclusion

This paper proposes a rolling bearing fault diagnosis technique based on MVC. The method first employs a VAE to extract latent features from multi-source signals. Then, based on the extracted features, a joint matrix and a high-confidence matrix are constructed. Complementary information among views is mined through low-rank approximation. Finally, secondary clustering is performed to achieve accurate identification of fault types. Experimental results demonstrate that the proposed method is validated on public datasets. In the fixed-axis gearbox dataset, the proposed method can accurately identify the true number of categories and sample labels. It is noteworthy that after removing the well-performing measurement point 2, the proposed method still maintains 100 percent diagnostic accuracy, while the comparison methods exhibit performance degradation. This reflects that the proposed method does not rely on information from a single data source. Instead, it effectively integrates all data sources and achieves comprehensive mining of multi-source data. As a result, the robustness and reliability of diagnostic results are significantly improved.

Finally, sincere gratitude is extended to Professor Guangming Dong from Shanghai Jiao Tong University for providing the valuable Dingzhou gearbox dataset for this study, which establishes a solid foundation for the validation and experimental analysis of the proposed method.

References

[1] Lei Y. G., Yang B., Du Z. J., et al. Deep transfer diagnosis method for mechanical equipment faults under big data [J]. Journal of Mechanical Engineering, 2019, 55(07): 1-8.
[2] Shao H. D., Xiao Y. M., Yan S. Simulation data driven improved unsupervised domain adaptation for bearing fault diagnosis [J]. Journal of Mechanical Engineering, 2023, 59(03): 76-85.
[3] Li Y., Zhao M., Xu M. Y., et al. Review of multi-source information fusion technology [J]. Intelligent Computer and Applications, 2019, 9(05): 186-189.
[4] Fu J. Y., Xu J. H. Shaft system fault diagnosis for hydropower units based on dual-channel data main vibration feature extraction [J]. Noise and Vibration Control, 2024, 44(05): 172-178.
[5] Zhang L., Zhen C. Z., Yi J. Y., et al. Gearbox fault diagnosis based on dual-channel feature fusion CNN-GRU [J]. Journal of Vibration and Shock, 2021, 40(19): 239-245+94.
[6] Liu T. T., Wang Z. M., Yu W. Y., et al. Gearbox fault diagnosis based on decision fusion method and transfer learning [J]. Journal of University of Jinan(Science and Technology), 2025, 39(03): 379-388.
[7] HUSSAIN S F, MUSHTAQ M, HALIM Z. Multi-view document clustering via ensemble method [J]. J Intell Inf Syst, 2014, 43(1): 81-99.
[8] LU M, XU Y, CHU W, et al. Contrastive multi-view clustering based on multi-head attention mechanisms and three-way decision [J]. Knowledge and Information Systems, 2025, 67(7): 1-24.
[9] YANG Y, ZHU C. Deep multi-view clustering based on global hybrid alignment with cross-contrastive learning [J]. The Visual Computer, 2024, 41(7): 1-13.
[10] Zhang Y. H., Zhang Z. Y., Zhao X. P., et al. Bearing fault diagnosis method for imbalanced samples based on VAE-GAN and FLCNN [J]. Journal of Vibration and Shock, 2022, 41(09): 199-209.
[11] Deng M. Y., Li C. Z., Yang H. Bearing fault diagnosis based on frequency domain feature variational autoencoder [J]. Computer Measurement & Control, 2023, 31(04): 70-75+148.
[12] Yin A. J., Wang Y., Dai Z. X., et al. Bearing health condition assessment based on variational autoencoder [J]. Journal of Vibration, Measurement & Diagnosis, 2020, 40(05): 1011-1016+30.
[13] Wang Z. L., Lu T. F., Wang Y. B. Contact network defect detection method based on variational autoencoder [J]. Journal of Harbin Institute of Technology: 1-13.
[14] She B., Tian F. Q., Liang W. G. Fault diagnosis method based on deep convolutional variational autoencoder network [J]. Chinese Journal of Scientific Instrument, 2018, 39(10): 27-35.
[15] Zhai Z. L., Liang Z. M., Zhou W., et al. Review of variational autoencoder models [J]. Computer Engineering and Applications, 2019, 55(03): 1-9.
[16] YU Z, DONG Z, YU C, et al. A review on multi-view learning [J]. Frontiers of Computer Science, 2024, 19(7): 197334.
[17] XIAO Y, HUI L, YUXIU L, et al. Auto-weighted sample-level fusion with anchors for incomplete multi-view clustering [J]. Pattern Recognition, 2022, 130.