

Article citation info:

Zhang Y, Wang J, Zhao R, Deng L, Dimensionality reduction of rotor fault dataset based on joint embedding of multi-class graphs, *Eksploracja i Niezawodność – Maintenance and Reliability* 2024; 26(1) <http://doi.org/10.17531/ein/177417>

Dimensionality reduction of rotor fault dataset based on joint embedding of multi-class graphs

Indexed by:



Yongfei Zhang^a, Jiaxuan Wang^a, Rongzhen Zhao^{a,*}, Linfeng Deng^a

^a School of Mechanical and Electrical Engineering, Lanzhou University of Technology, Lanzhou 730050, People's Republic of China

Highlights

- Constructed local and global median feature line plots that mined global information.
- Constructing hypergraphs to carve out spatially diverse relationships of features.
- Construction, and joint embedding of multi-class graphs to mine sensitive features.
- Verified the generalization and robustness using KNN, BP and SVM as classifiers.

Abstract

Traditional dimensionality reduction techniques usually rely on a single or a limited number of similar graphs for graph embedding, which limits their ability to extract more information about the internal structure of the data. To address this problem, this study proposes a rotor fault dataset dimensionality reduction algorithm based on multi-class graph joint embedding (MCGJE). The algorithm first overcomes the defect that the traditional feature space cannot take both local and global information into account by constructing local and global median feature line graphs; secondly, based on the graph embedding framework, the algorithm also constructs a hypergraph structure for inscribing complex multivariate relationships between high-dimensional data in the feature space, which in turn enables it to contain more fault information. Finally, we conducted two different rotor fault simulation experiments. The results show that the MCGJE-based algorithm has robust dimensionality reduction capability and can significantly improve the accuracy of fault identification.

Keywords

Multi-class graphs, dimensionality reduction, hypergraph structure; median feature line graph

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

1. Introduction

At present, the condition monitoring data of various machinery and equipment is becoming the basis for enterprise production to achieve high-quality operation and efficient management, and such data is also an important strategic resource indispensable for the implementation of intelligent manufacturing [25, 12, 21, 17]. Rotating machinery with a rotor system as the basic architecture has been widely used in aerospace, rail transport, wind power generation, and other important fields, and the implementation of intelligent operation and maintenance management for them is of great significance.

In engineering applications, it is usually necessary to install multiple sensors at the key cross-section of the system to collect multi-channel information and extract statistical features from the vibration signals collected from each channel, but the increase in the number of features will undoubtedly produce the problem of "dimensional catastrophe" [27, 19, 15, 14, 10]. Therefore, exploring how to reduce the relevant features in high-dimensional fault datasets and solving the problem of constructing a good knowledge representation model for the representation of equipment operating states are especially

(*) Corresponding author: Y. Zhang (ORCID: 0009-0005-6838-7648) 1312867638@qq.com, J. Wang (ORCID: 0009-0008-0450-6857) xuangesss@gmail.com, E-mail addresses: R. Zhao (ORCID: 0000-0001-9225-8700) zhaorongzhen@lut.edu.cn, L. Deng (ORCID: 0000-0001-8813-4692) denglinfeng2002@163.com

crucial for developing machine intelligence fault diagnosis and decision-making techniques for industrial big data [23, 2, 20]. In practice, the original fault features we have collected have many redundant features that greatly increase the difficulty of fault identification. Therefore, it is very necessary to carry out dimensional reduction of the original fault features. The dimensional reduction method can not only eliminate redundant features but also greatly improve the accuracy of fault identification. In recent years, a variety of dimensionality reduction methods are emerging one after another, which are mainly divided into traditional feature dimensionality reduction methods and data dimensionality reduction methods based on deep neural network technology, which greatly promote the development of dimensionality reduction methods for datasets. However, the latter requires a large amount of sample data when training the network, which makes the network training time and difficulty increase. Traditional dimensionality reduction methods can achieve the same effect as well as reduce the cost when the amount of sample data is small in a specific domain. Therefore, there is still engineering value in the study of traditional dimensionality reduction methods [7, 8, 32, 34]. In recent years, many dimensionality reduction methods have been successfully applied to the fault diagnosis of mechanical equipment, such as the widely used dimensionality reduction algorithms with principal component analysis (PCA) [24, 16], But the non-linear correlation may be lost after PCA dimensionality reduction, so in order to better deal with the nonlinear relationship between the flow of learning came into being, which is representative of the proposed domain Preserving Embedding Neighborhood Preserving Embedding (NPE) algorithm [3, 13]; Marginal Fisher Analysis Marginal Fisher Analysis (MFA) [33], Orthogonal discriminant projection Orthogonal discriminant projection (ODP) [11]. and so on. The above algorithms perform dimensionality reduction by finding the optimal projection matrix, in which the projection matrix of MFA is solved by maximizing the Fisher criterion, and ODP is solved by scatter-weighted difference. Algorithms that rely on sample point-to-point (P2P) metric learning have restricted generalization capabilities and are incapable of extracting more discriminative information for subsequent fault classification. Hence, nearest feature space embedding (NFSE) [6], weighted feature line embedding (WFLE) [1], multiple

kernel feature line embedding (MKFLE) [4], and others based on point-to-space (P2S) metrics have been introduced. These algorithms utilize the P2S metric, which not only compresses the feature space dimension but also enables the extraction of additional information for fault identification. However, there is a nearest-neighbor selection error during the computation process when employing the P2S metric. To tackle this issue, Dong [5] introduced an enhancement to the P2S projection metric, replacing it with the median metric. This modification aims to minimize the adverse effects of nearest neighbor selection errors on the algorithm. In addition, he proposed a feature space nearest neighbor selection method using samples instead of projected points, which improves the speed of nearest neighbor selection. The development of the above feature space is that the mining of local information cannot take into account global information and is a single point-to-projected point or median point relationship, which cannot portray high-dimensional data in a many-to-many relationship.

Targeting the characterization of multivariate relationships between data, hypergraph structural models for portraying multivariate relationships between data are emerging and have been fruitful in computer disciplines, not least in applications such as image classification and fault diagnosis, so hypergraphs have already made a mark in the field of fault recognition. Two hypergraph neural network models were proposed for fault classification in electromechanical coupled systems. Zhang et al. [30] proposed a network based on time translation, which achieves accurate classification. Yan et al. [26] proposed a new multiresolution hypergraph neural network to address the problem of traditional graph neural networks not being able to depict complex higher-order relationships in data. Yuan [28] et al. proposed a dimensionality approximation algorithm for semi-supervised multi-graph joint embedding in response to the problem of excessive dimensionality of fault datasets by combining graph and hypergraph models. From the above development, it can be seen that the hypergraph model has been developed rapidly, but the hypergraph, as a model describing multivariate relationships, has limited noise resistance and generalization ability. Therefore, from the perspective of multi-class graph, linking the hypergraph structural model with the feature space structural model may achieve better dimensionality reduction.

Dimensionality reduction using feature space can well improve the generalization and anti-noise properties of the algorithm. However, the median metric of the feature space is a single point-to-point pattern, which makes the algorithm itself ignores the many-to-many relationship in the high-dimensional space, while the hypergraph itself is a common method to use the hyperedge to include multiple data samples so as to better mine the multivariate fault data information. The combination of the two can improve the algorithm's noise immunity and generalization, and can also accurately describe the high-dimensional multivariate fault data information in the feature space. So for the above problems of traditional graph embedding, this paper proposes to construct local median feature line graph, global median feature line graph and hypergraph, and use the method of multi-class graph union to achieve the purpose of improving the generalization and noise resistance of graph structure and also well depicting the many-to-many relationship of high-dimensional samples of the data, and taking into account the local and global information. To this end, a new dimensionality reduction algorithm based on the joint embedding of multi-class graphs is proposed for rotor fault datasets. And it is successfully applied to rotor fault diagnosis. The main contributions of this thesis are as follows:

- The traditional feature space algorithm cannot take into account both local and global structure information. This paper aims to solve this defect by constructing local median eigenline maps and global median eigenline maps to achieve the purpose of comprehensively extracting the intrinsic structural information of data.
- In view of the fact that the composition of graph embedding algorithms in the traditional feature space is still in the form of a single point-to-point construction, which is unable to portray the many-to-many relationships of high-dimensional data, the hyperedge set, i.e., the hypergraph structure, is introduced here, which is able to portray the many-to-many relationships of the data very well, and makes up for the intrinsic limitations of the feature space.
- For the first time, the algorithm combines local median feature line graphs, hypergraphs, and global median feature line graphs to jointly extract sensitive features from multiple types of graphs, thus greatly reducing the

difficulty of subsequent fault pattern recognition.

2. Related work

2.1. Introduction to feature space algorithms

The feature space algorithm discovers the most efficient projection matrix in feature space by considering three factors, namely, category divisibility, neighborhoods structure preservation and nearest-neighbor feature space metric, in a graph embedding framework. Although the feature space algorithm has good neighborhoods structure preservation and improves the generalizations performance of the algorithm, the feature space projection metric it employs is prone to nearest neighbor selection error.

2.1.1. Median feature line

To address the shortcomings of the nearest neighbor selection error of the feature space algorithm, a graphical representation of the error is given in Fig. 1. Observe that the sample point X_1, X_2 and the point to be measured x_i are farther away from each other compared to the sample point X_3, X_4 . However, in the case of the projection metric $|X_i P_{12}| < |X_i P_{34}|$, this error results in the scattering matrix not reflecting the local topology of the sample in a real and effective way. Therefore, the literature [5] proposes to improve the projection metric to the median point metric, i.e., take the midpoint of the sample point X_1, X_2 and the sample point X_3, X_4 as the projection points and observe that $|X_i P_{m_{12}}| > |X_i P_{m_{34}}|$, and therefore this approach is more in line with the real data distribution structure, Furthermore, it can solve the problem of neighbor selection error.

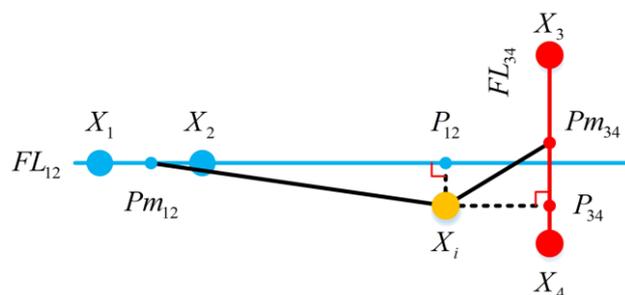


Fig. 1. Improvement of extrapolation error.

2.1.2. Nearest neighbor feature space selection guided by P2S

Taking the construction of the nearest-neighbor into account, let's use two feature line embedding graphs as an example: the

conventional feature space nearest-neighbor selection, as illustrated in Fig. 2(a). In this scenario, the sample point x_i calculates the distance to all six feature lines and then sorts them to identify the two nearest feature lines. Therefore, literature [5] proposes the nearest neighbor selection in the feature space



Fig. 2. Selection of nearest neighbor feature space.

2.2. Hypergraph Structure

As shown in Fig. 3, graph models are widely used in the field of machine learning to describe the relationships that exist between things. In fault diagnosis applications, data samples are represented by each node of the graph model, and similar relationships between the samples are represented by

guided by P2S as shown in Fig. 2(b). It determines the nearest median by computing only three points in the sample. Therefore, all the feature line algorithms proposed later use the P2S nearest neighbor selection method.

connecting lines between the nodes. The main difference between a hypergraph model and a graph model is that it has more hyperedge vertices. The edges of the graph model merely signify the connection between two points, while the edges of the hypergraph model incorporate multiple vertices. Moreover, hyperedges linked by multiple points share more vertices and thereby contain more data.

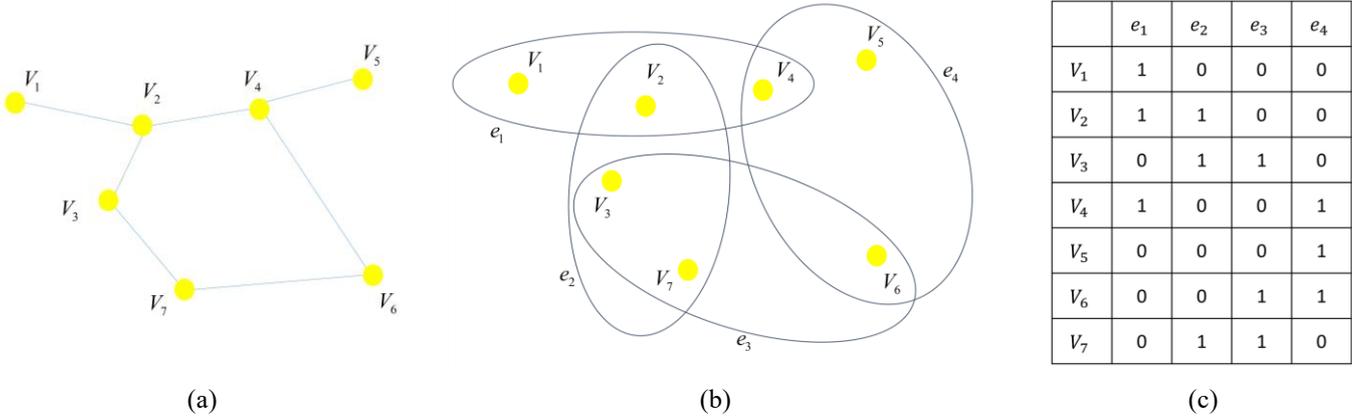


Fig. 3. Graph Structure Model: (a) simple graph. (b) hypergraph. (c) incidence matrix Q .

Although the feature space median metric, i.e., it solves the problem of nearest-neighbor selection error and improves the efficiency of nearest-neighbor selection, this algorithm and the related improved algorithms only obtain the local information of the samples, and cannot consider the global information. On the other hand, the algorithm is a single point-to-point metric, which can not portray the complex multivariate structure among high-dimensional data, and it can not mine more data

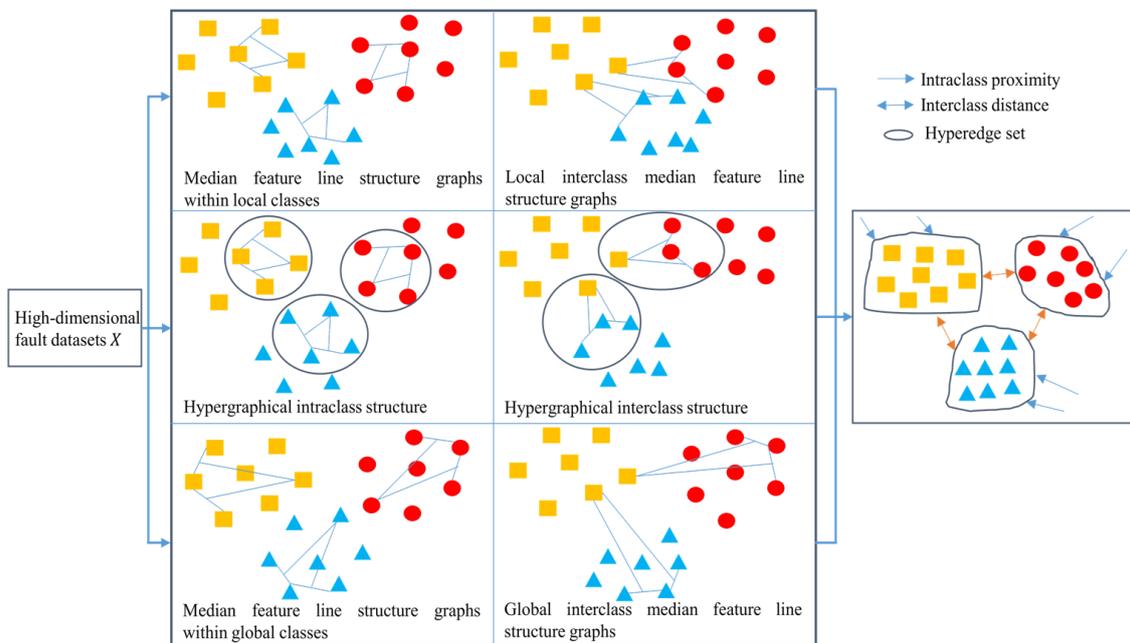
information; and the algorithm is a single graph embedding, which can not synergise with the multi-class graph structure, and the effect of dimension reduction is limited. Therefore, how to deal with the complex structure among high-dimensional data more comprehensively and more realistically becomes the focus of this study.

3. Proposed MCGJE algorithm

In the process of data dimensionality reduction, an accurate and

comprehensive representation of the structure of fault data helps to extract valuable feature information for classification. However, high-dimensional fault datasets usually have complex internal structures, and a single structure alone cannot fully describe such complex internal relationships. Therefore, in this paper, we construct a data dimensionality reduction algorithm for collaborative discrimination of multi-structured graphs (MCGJE), which combines local median feature graphs, global median feature graphs and hypergraphs. The basic principle is as follows: The original faulty high-Witte set is $X = [x_1, x_2, \dots, x_n] \in R^{D \times n}$, and the downgraded faulty low-witte set is $Y = [y_1, y_2, \dots, y_n] \in R^{d \times n}$; after the transformation of the projection matrix A of the MCGJE algorithm, $Y = A^T X$ realizes the sample features downgraded, where $(d \leq D)$. The specific

Fig. 4.



Schematic diagram of the MCGJE algorithm.

3.1. Constructed local median feature line graph

According to the schematic diagram of the MCGJE algorithm shown in Fig. 4, the inner class local median feature line graph and the inner inter-local median local median feature line graph are constructed to achieve the purpose of mining the local information and to improve the generalization and noise immunity of the reduced projection matrix.

Determine the Euclidean distance between the sample point and the median projection point.

$$\|x_i - f^{(mp)}(x_i)\|^2 \quad (1)$$

steps are as follows:

First, we capture local information by constructing local intra- and inter-class median eigenline maps, while enhancing the generalizations and noise immunity of the projection matrix. Next, we design intra- and inter-class hypergraphs to capture higher-order relationships between data samples. Subsequently, we build intra- and inter-class global median eigenline maps to reflect the global information in the set of high-dimensional fault features. Finally, in the low-dimensional embedding space, we reinforce intra-class structural relationships and suppress inter-class structural relationships. This approach can extract sensitive features and achieve dimensionality reduction of fault data.

The median projection points of sample point x_i in feature space $f^{(mp)}$, denoted as $f^{(mp)}(x_i)$, is determined using calculation $f^{(mp)}(x_i) = \frac{1}{p} \sum_{j=1}^p x_j^p$. Additionally, x_j^p represents the initial j constituent sample points. To enhance computational simplicity, this study adopts a constant value of $p = 2$.

As shown in Fig. 4, the median feature line intraclass graph $G_{nw}^{(mp)} = \{X_{nw}, W_{nw}^{(mp)}\}$ is constructed using intraclass nearest-neighbor median projection points, where X_{nw} is the intraclass nearest-neighbor sample point and $W_{nw}^{(mp)}$ is the intraclass

weight matrix of the local median feature line graph.

$$w_{nw}^{(mp)}(x_i) = \exp\left(-\frac{\|x_i - f_{nw}^{(mp)}(x_i)\|^2}{2(t_i^w)^2}\right) \quad (2)$$

where $t_i^w = \frac{1}{N} \sum_{i=1}^N \|x_i - f_{nw}^{(mp)}(x_i)\|$, $f_{nw}^{(mp)}(x_i)$ is the k_1 nearest neighbor median projection points within the class of x_i .

As shown in Fig. 4, the median feature line interclass graph $G_{nb}^{(mp)} = \{X_{nb}, W_{nb}^{(mp)}\}$ is constructed using the interclass nearest-neighbor median projection points, where X_{nb} is the interclass nearest-neighbor sample point and $W_{nb}^{(mp)}$ is the interclass weight matrix of the local median feature line graph.

$$w_{nb}^{(mp)}(x_i) = \exp\left(-\frac{\|x_i - f_{nb}^{(mp)}(x_i)\|^2}{2(t_i^b)^2}\right) \quad (3)$$

where $t_i^b = \frac{1}{N} \sum_{i=1}^N \|x_i - f_{nb}^{(mp)}(x_i)\|$, $f_{nb}^{(mp)}(x_i)$ is the k_2 nearest neighbor median projection points between classes of x_i .

Find the objective function. The intraclass spatial scatter matrix is computed using local median feature line graphing S_{nw}^{mp} , the inter-class spatial scatter matrix is computed using local median feature line graphing S_{nb}^{mp} . The expressions are respectively.

$$\begin{cases} S_{nw}^{(mp)} = \sum w_{nw}^{(mp)}(x_i)(x_i - f_{nw}^{(mp)}(x_i)) - (x_i - f_{nw}^{(mp)}(x_i))^T \\ S_{nb}^{(mp)} = \sum w_{nb}^{(mp)}(x_i)(x_i - f_{nb}^{(mp)}(x_i)) - (x_i - f_{nb}^{(mp)}(x_i))^T \end{cases} \quad (4)$$

Therefore, further simplifying the above equation, the objective function of the local Median feature line graph is defined as:

$$\begin{cases} \min_A \text{tr}(A^T X L^{nw} X^T A) \\ \max_A \text{tr}(A^T X L^{nb} X^T A) \end{cases} \quad (5)$$

where the Laplace matrix of the local Median feature line graph within and between classes can be expressed as $L^{nw} = D^{nw} - W^{nw}$, $L^{nb} = D^{nb} - W^{nb}$, D^{nw}, D^{nb} as the diagonal matrices $D_{ij}^{nw} = \sum_{j=1} w_{nw}^{(mp)}(x_i)$, $D_{ij}^{nb} = \sum_{j=1} w_{nb}^{(mp)}(x_i)$.

3.2. Constructed global median feature line graph

The method includes constructing a global intra-class median feature line graph using intra-class distances between distant neighbor points, which contains global information in each class of data. In addition, a global interclass median eigenline graph is constructed using interclass distances between distant neighbor points, thereby considering global information in each class of data, as shown in the following process.

As shown in Fig. 4, the global median feature line intraclass graph $G_{dw}^{(mp)} = \{X_{dw}, W_{dw}^{(mp)}\}$ is constructed using the intraclass

median far-neighbor projection points, where X_{dw} is the intraclass far-neighbor sample point and $W_{dw}^{(mp)}$ is the intraclass weight matrix of the global median feature line graph.

$$w_{dw}^{(mp)}(x_i) = \exp\left(-\frac{\|x_i - f_{dw}^{(mp)}(x_i)\|^2}{2(t_i^w)^2}\right) \quad (6)$$

where $t_i^w = \frac{1}{N} \sum_{i=1}^N \|x_i - f_{dw}^{(mp)}(x_i)\|$, $f_{dw}^{(mp)}(x_i)$ is the k_1 distant neighborhood median projection point within the class of x_i .

As shown in Fig. 4, the global median feature line interclass graph $G_{db}^{(mp)} = \{X_{db}, W_{db}^{(mp)}\}$ is constructed using the interclass median far-neighbor projection points, where X_{db} is the interclass far-neighbor sample points and $W_{db}^{(mp)}$ is the interclass weight matrix of the global median feature line graph.

$$w_{db}^{(mp)}(x_i) = \exp\left(-\frac{\|x_i - f_{db}^{(mp)}(x_i)\|^2}{2(t_i^b)^2}\right) \quad (7)$$

where $t_i^b = \frac{1}{N} \sum_{i=1}^N \|x_i - f_{db}^{(mp)}(x_i)\|$, $f_{db}^{(mp)}(x_i)$ is the k_2 distant neighborhood median projection point between classes of x_i .

Therefore, the intraclass spatial scatter matrix is computed using global median feature line graphing S_{dw}^{mp} , the inter-class spatial scatter matrix is computed using global median feature line graphing S_{db}^{mp} . The expressions are respectively.

$$\begin{cases} S_{dw}^{(mp)} = \sum w_{dw}^{(mp)}(x_i)(x_i - f_{dw}^{(mp)}(x_i)) - (x_i - f_{dw}^{(mp)}(x_i))^T \\ S_{db}^{(mp)} = \sum w_{db}^{(mp)}(x_i)(x_i - f_{db}^{(mp)}(x_i)) - (x_i - f_{db}^{(mp)}(x_i))^T \end{cases} \quad (8)$$

Therefore, further simplifying the above equation, the objective function of the global Median feature line graph is defined as:

$$\begin{cases} \min_A \text{tr}(A^T X L^{dw} X^T A) \\ \max_A \text{tr}(A^T X L^{db} X^T A) \end{cases} \quad (9)$$

where the Laplace matrix of the global Median feature line graphs within and between classes can be expressed as $L^{dw} = D^{dw} - W^{dw}$, $L^{db} = D^{db} - W^{db}$, $D_{ij}^{dw} = \sum_{j=1} w_{dw}^{(mp)}(x_i)$, $D_{ij}^{db} = \sum_{j=1} w_{db}^{(mp)}(x_i)$.

The constructed local median eigenline maps and global median eigenline maps not only take into account the category separability through intra-class and inter-intra-class, but also take into account the local and global information through the near-neighborhoods median eigenspace metric and far-neighborhoods median eigenspace metric factors, and well maintain the generalizations performance and noise resistance of the projection matrix. However, the projection matrix is ultimately only a single point-to-point similarity relation of the

data samples and cannot contain multiple high-dimensional information. To solve these problems, we introduce the concept of hypergraph structure.

3.3. Hypergraph structure of constructions

The median projection point is the midpoint of any two data samples, and when there are only two sample points, the hypergraph constructed using these two sample points must contain the midpoint of the Euclidean distance between these two sample points. Similarly, when extended to multiple sample points, the hypergraph constructed using those sample points must also contain the midpoint of the Euclidean distance between all those sample points. Therefore, the data sample points are used to guide the construction of the hyperedge of the hypergraph model so that it contains more sample points and projection points, and thus more information, for the purpose of portraying higher-order relationships between data samples.

Construct intraclass hypergraph $G_Q^w = \{X_Q^w, E_Q^w, W_Q^w\}$ and interclass hypergraph $G_Q^b = \{X_Q^b, E_Q^b, W_Q^b\}$ to accurately describe the multi-class corresponding to multi-sample relationship between fault data. Here, X_Q^w and X_Q^b represent intra-class and inter-class hypergraph sample sets, while E_Q^w and E_Q^b denote intra-class and inter-class hypergraph hyperedge sets, and the corresponding weight matrices are W_Q^w and W_Q^b .

In the intraclass hypergraph G_Q^w , the hyperedge weights are defined as:

$$w_i^w = w(e_i^w) = \sum_{x_j \in e_i^w} \exp\left(-\frac{\|x_i - x_j\|}{2(t_i)^2}\right) \quad (10)$$

where $t_i = \frac{1}{k_1} \sum_{x_j \in e_i^w} \|x_i - x_j\|$, x_j are the k_1 nearest neighbor points within the class of x_i .

The association matrix Q^w is defined as follows:

$$Q_{ij}^w = q(x_i, e_j^w) = \begin{cases} 1, & \text{if } x_i \in e_j^w \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

Based on the hyperedge weights W^w and the correlation matrix Q^w , we derive the intraclass sample point degrees $x_i \in X^w$ and the intraclass hyperedge degrees $e_j^w \in E^w$:

$$u(x_i, e_j^w) = \sum_{e_j \in E^w} w(e_j^w) = \sum_{j=1}^N w(e_j^w) q(x_i, e_j^w) = \sum_{j=1}^N w_j^w Q_{ij}^w \quad (12)$$

$$u_j^w = u(e_j^w) = \sum_{x_i \in X^w} q(x_i, e_j^w) = \sum_{i=1}^M Q_{ij}^w \quad (13)$$

Similarly, the hyperedge weights of the standard hypergraph G_Q^b between classes are defined as follows:

$$w_i^b = w(e_i^b) = \sum_{x_j \in e_i^b} \exp\left(-\frac{\|x_i - x_j\|}{2(t_i)^2}\right) \quad (14)$$

where $t_i = \frac{1}{k_2} \sum_{x_j \in e_i^b} \|x_i - x_j\|$, x_j are the k_2 nearest neighbor points between classes of x_i .

The corresponding association matrix Q^b is denoted as:

$$Q_{ij}^b = q(x_i, e_j^b) = \begin{cases} 1 & \text{if } x_i \in e_j^b \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

The degree of the interclass sample point $x_i \in X^b$ and the degree of the interclass hyperedge $e_j^b \in E^b$ are defined as:

$$u(x_i, e_j^b) = \sum_{e_j \in E^b} w(e_j^b) = \sum_j w(e_j^b) h(x_i, e_j^b) = \sum_j w_j^b Q_{ij}^b \quad (16)$$

$$u_j^b = u(e_j^b) = \sum_{x_i \in X^b} q(x_i, e_j^b) = \sum_{i=1}^M Q_{ij}^b \quad (17)$$

When downscaling high-dimensional data, it is necessary to make the intra-class information more compact and the inter-class information more distant in order to extract sensitive features. Therefore, the final objective function can be expressed as follows:

$$\begin{cases} \min_A \frac{1}{2} \sum_{e_k^w \in E^w} \sum_{x_i, x_j \in e_k^w} \frac{w(e_k^w) q(x_i, e_k^w) q(x_j, e_k^w)}{u(e_k^w)} \|A^T x_i - A^T x_j\|^2 \\ = \text{tr}(A^T X L_Q^w X^T A) \\ \max_A \frac{1}{2} \sum_{e_k^b \in E^b} \sum_{x_i, x_j \in e_k^b} \frac{w(e_k^b) q(x_i, e_k^b) q(x_j, e_k^b)}{u(e_k^b)} \|A^T x_i - A^T x_j\|^2 \\ = \text{tr}(A^T X L_Q^b X^T A) \end{cases} \quad (18)$$

Equation (18) can be used to derive the optimization objective function for hypergraphs:

$$\begin{cases} \min_A \text{tr}(A^T X L_Q^w X^T A) \\ \max_A \text{tr}(A^T X L_Q^b X^T A) \end{cases} \quad (19)$$

3.4. Designed MCGJE objective function

In order to cluster a subset of intra-class samples during low-dimensional embedding and as far away from inter-class samples as possible, MCGJE needs to maximize the inter-class scatter matrix while minimizing the intra-class scatter matrix.

Therefore, the objective function of the MCGJE algorithm is:

$$\begin{aligned} & \underset{A}{\text{argmax}} \frac{\text{tr}\{\alpha[(A^T X L^{nb} X^T A) + (A^T X L^{db} X^T A)] + (1 - \alpha)(A^T X L_Q^b X^T A)\}}{\text{tr}\{\beta[(A^T X L^{nw} X^T A) + (A^T X L^{dw} X^T A)] + (1 - \beta)(A^T X L_Q^w X^T A)\}} \\ & = \underset{A}{\text{argmax}} \frac{\text{tr}\{A^T [\alpha(X L^{nb} X^T + X L^{db} X^T) + (1 - \alpha)X L_Q^b X^T] A\}}{\text{tr}\{A^T [\beta(X L^{nw} X^T + X L^{dw} X^T) + (1 - \beta)X L_Q^w X^T] A\}} \end{aligned} \quad (20)$$

where the parameter $\alpha, \beta \in [0, 1]$, is used to regulate the contribution of different scatter matrices to the feature extraction.

Due to the small sample issue, Fisher's criterion uses results in

matrix singularity, which lowers the algorithm's discrimination performance. The scatter-weighted difference is therefore maximized as the objective function. Finally, the objective function is built as shown in equation (21) to improve the algorithm's suitability for fault classification.

$$\begin{cases} J = \operatorname{argmax} A^T [(\alpha N^{mb} + (1 - \alpha)N^b)] - [(\beta N^{mw} + (1 - \beta)N^w)]A \\ \text{st. } A^T A = I \end{cases} \quad (21)$$

where $N^{mw} = XL^{dw}X^T + XL^{nw}X^T$, $N^{mb} = XL^{db}X^T + XL^{nb}X^T$, $N^b = XL_Q^bX^T$, $N^w = XL_Q^wX^T$.

Equation (21) corresponds to the generalized eigenvalue problem in the Lagrange multiplier approach.

$$[(\alpha N^{mb} + (1 - \alpha)N^b)] = \lambda [(\beta N^{mw} + (1 - \beta)N^w)] \quad (22)$$

Solve eq. (22) for eigenvalues and eigenvectors. Thus, the steps of the proposed MCGJE dimensionality reduction algorithm are as follows:

Input the number of nearest neighbor points required by MCGJE k_1, k_2 , the high dimensional dataset $X = \{x_i | i = 1, 2, \dots, n\} \in R^{d \times n}$ and the balance parameter α, β . Then output the low dimensional dataset $Y = \{y_i | i = 1, 2, \dots, n\} \in R^{d \times n}$ and the projection matrix $A = [a_1, a_2, \dots, a_r] \in R^{d \times r}$. The steps are as follows.

1. Create a local intraclass median feature line graph $G_{nw}^{(mp)}$. Compute the Laplace matrix L^{nw} of $G_{nw}^{(mp)}$;

2. Create a local interclass median feature line graph $G_{nb}^{(mp)}$. Compute the Laplace matrix L^{nb} of $G_{nb}^{(mp)}$;

3. Construct the global intraclass median feature line graph $G_{dw}^{(mp)}$. Compute the Laplace matrix L^{dw} of $G_{dw}^{(mp)}$;

4. Construct the global interclass median feature line graph $G_{db}^{(mp)}$. Compute the Laplace matrix L^{db} of $G_{db}^{(mp)}$;

5. Construct the intraclass hypergraph structure G_Q^w . Compute the hyper-Laplacian matrix L_Q^w of G_Q^w ;

6. Construct the interclass hypergraph structure G_Q^b . Compute the hyper-Laplacian matrix L_Q^b of G_Q^b ;

7. The eigenvalues in equation (22) are sorted in descending order. The eigenvectors associated with the first d eigenvalues are selected to construct the projection matrix $A = [\lambda_1, \lambda_2, \dots, \lambda_d] \in R^{D \times d}$, This matrix is then processed through $Y = A^T X$ to obtain a subset of low-dimensional embedded features.

4. Rotor fault diagnosis process for the developed MCGJE algorithm

The rotor fault diagnosis process is illustrated in Fig. 5 using the MCGJE algorithm that was proposed in this study.

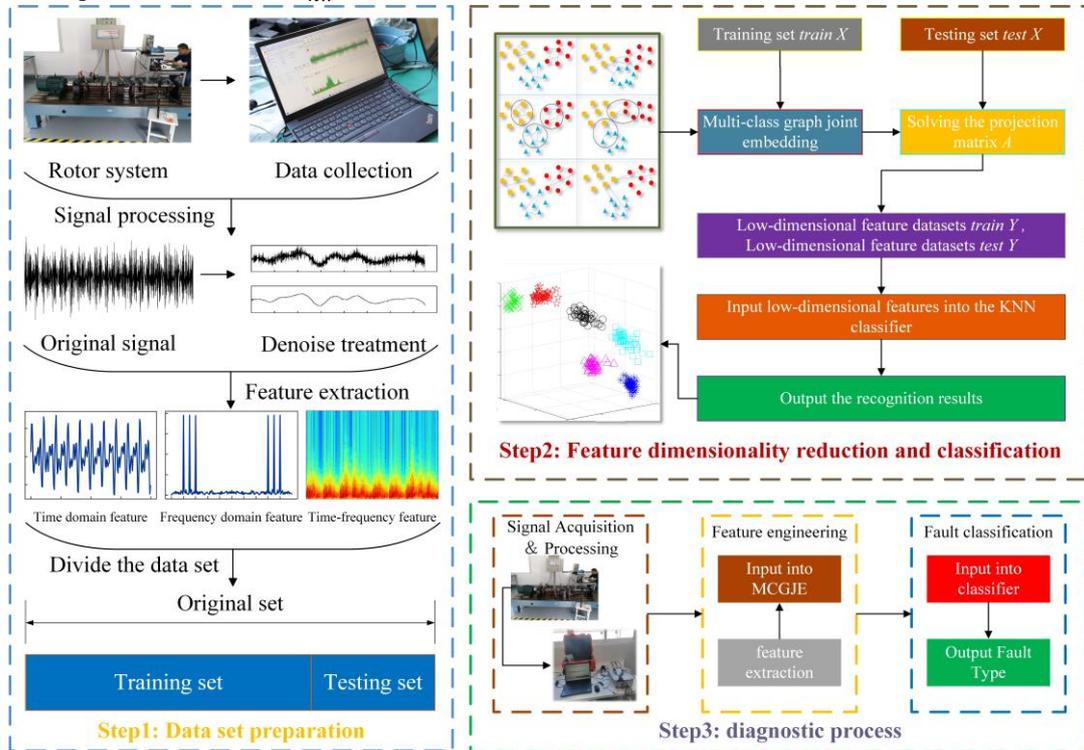


Fig. 5. Procedure of fault diagnosis.

The specific troubleshooting steps are as follows:

1. Firstly, wavelet noise was removed from the original

vibration signal. Then, the time domain, frequency domain and time-frequency domain feature parameters are collected. Then,

the high-voltage feature X is created, and then X is normalized. Finally, the training set and test set are divided into $trainX$ and $testX$.

2. The low-dimensional fault feature set $trainY$ and the projection matrix A are obtained by dimensional simplification of $trainX$ by MCGJE, and the low-dimensional fault feature set $testY$ is obtained by dimensional reduction of $testX$ by the projection matrix A as well.

3. Input $trainY$ and $testY$ into k-nearest neighbor (KNN) [29], back propagation (BP) [22], support vector machine (SVM) [9] for fault pattern recognition to get fault diagnosis results.

5. Experimental analysis

This study used two different types of experimental platforms, Experiment 1 and Experiment 2, to evaluate the performance of the MCGJE algorithm in various aspects.

5.1. Experiment 1

Unbalance failure is a common type of failure in rotor systems. In order to study the dynamic change of the state of this fault, the HZXT-DS-001 type experimental bench of Wuxi Houde Automation Instrumentation Co. Ltd. shown in Fig. 6, is selected to carry out the rotor unbalance fault simulation experiment. Four accelerometers were positioned on the test bench to capture two radial (X, Y) and one axial (Z) vibration signals from each bearing housing. Fig. 6(a) is a physical drawing of the rotor test bench, and Fig. 6(b) shows the exact location of the bearing housing and the mounted sensors. Among them, I~IV are the four bearing seats, and ch1~12 are the signal acquisition channels of the four acceleration sensors, respectively.

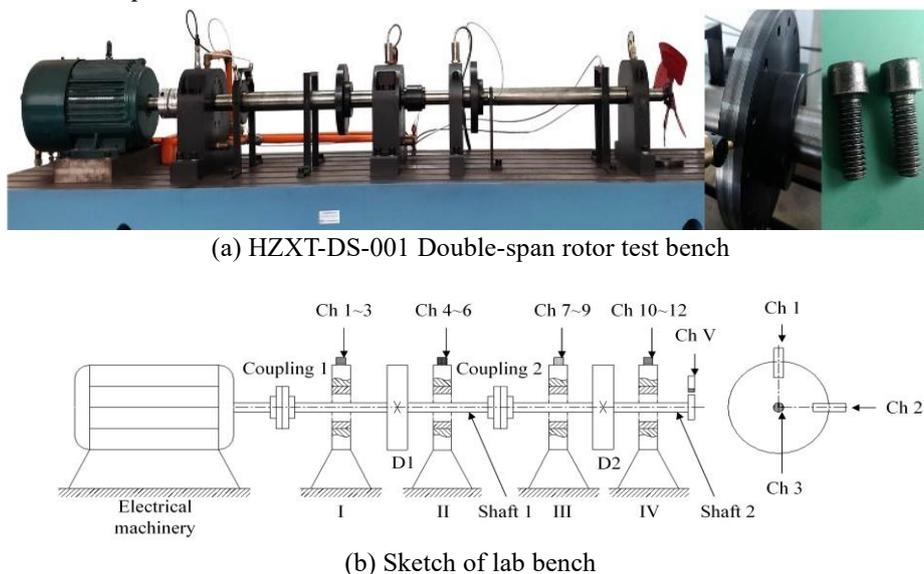


Fig. 6. Model HZXT-DS-001 double span rotor test bench and its sketch.

The configurations of mass disks with unbalanced faults in one of the experiments are shown in the following table. For ease of Table 1. Configuration of the mass disk.

description, the six faults are labeled RF 1, RF 2, RF 3, RF 4, RF 5, and RF 6.

Fault type	Number of unbalanced mass blocks	
	Mass disk 1	Mass disk 2
RF 1	2	0
RF 2	2	3
RF 3	3	0
RF 4	3	3
RF 5	0	2
RF 6	0	3

Each mass imbalance state is simulated by allocating different numbers of mass blocks on the mass disc 1 and disc 2, and the number of mass blocks on each mass disc has three groups of 0, 2, and 3, which ensures that six mass imbalance states can be simulated by adding a different number of mass blocks on the two mass discs. At a sampling frequency of 20 KHz and a

rotational speed of 3400 r/min, 100 sets were collected for each fault condition. Among them, 50 groups are used for training and 50 groups are used for testing. According to Table 2, a total of 38 statistical feature parameters in time domain, frequency domain and time-frequency domain are extracted, thus forming a high-dimensional set, in which $12 \times 38 = 456$.

Table 2. Characteristic parameters.

Serial number	Features	Serial number	Features	Serial number	Features
1	Maximum	9	Root-mean-square amplitude	17	Average frequency
2	Minimum	10	Root mean square	18	Spectral second-order moments
3	Peak-to-peak ratio	11	Absolute average	19	Standard deviation rate
4	variance	12	Shape factor	20	Kurtosis frequency
5	Averages	13	Crest factor	21	Root-mean-square frequency
6	Skewness	14	Impulse factor	22	center frequency
7	Kurtosis	15	Clearance factor	23~38	Four-layer wavelet packet decomposition of band energy features
8	Mean square	16	Kurtosis value		

Note: 1~16 are time-domain features, 17~22 are frequency-domain features, 23~38 are time-frequency-domain features

5.1.1. Parameter settings

In this study, the parameters to be configured for the MCGJE algorithm are the target dimension d , the nearest neighbor counts k_1 and k_2 , and the special equilibrium coefficients α and β . According to the literature [18], it is usually chosen as the number of categories minus 1, i.e., $d = 5$.

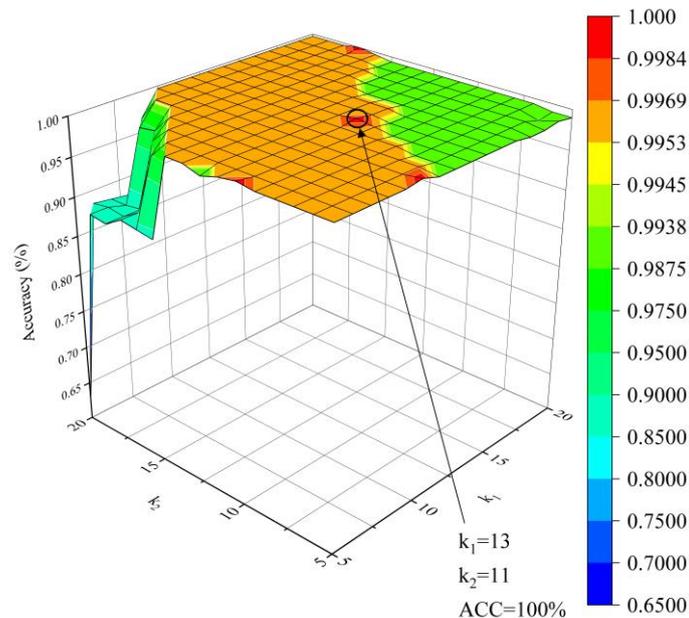


Fig. 7. Selection of parameters k_1 , k_2 .

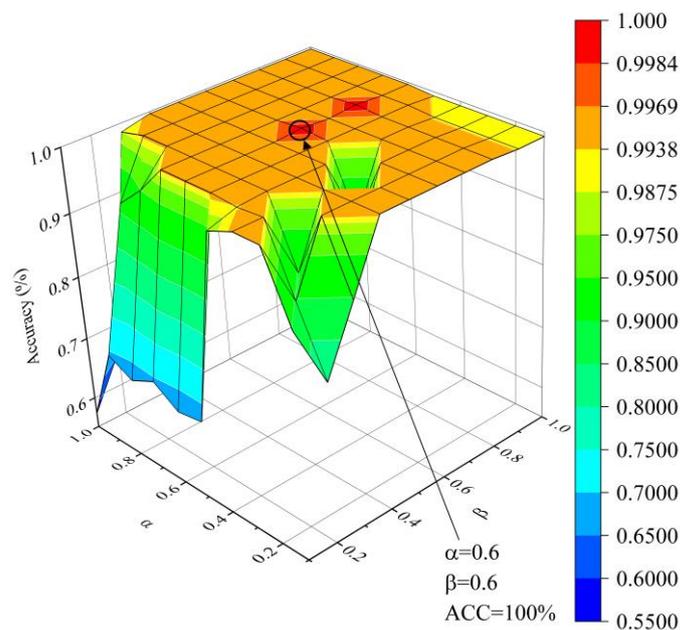


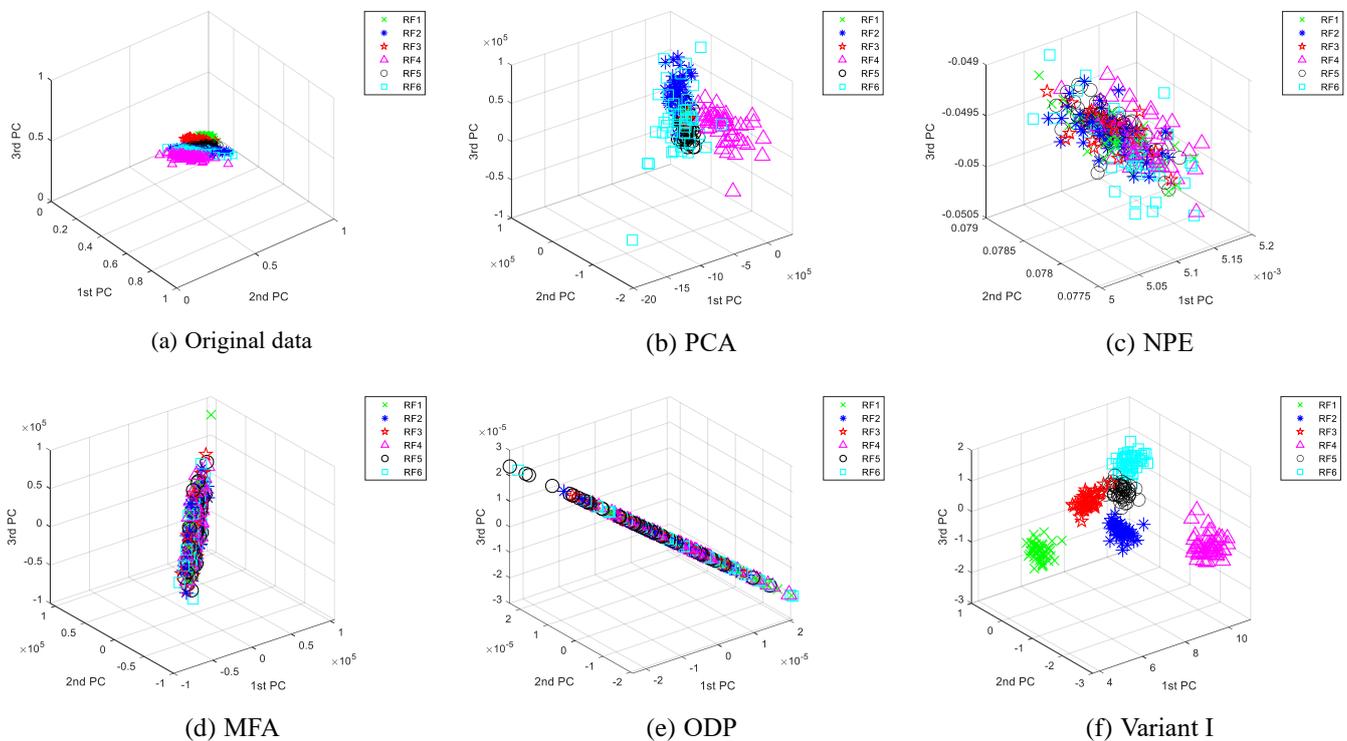
Fig. 8. Selection of parameter α , β .

The nearest neighbor points usually need to be larger than the embedding dimension and smaller than the number of samples. Furthermore, as elucidated in the scholarly works by authors [31], it has been observed that when the value of the number of nearest neighbor points is too small, the graph structure of MCGJE may not adequately reflect the inherent data information. On the contrary, when the value of the number of

nearest neighbor points is too large, it not only increases the computational cost significantly, but also ignores the local information of multi-class graphs. To solve the above problem, the number of nearest neighbor points is determined using the grid search method. Grid search is a model hyperparameter optimization technique commonly used to optimize three or fewer numbers of hyperparameters, and is essentially an exhaustive method. For each hyperparameter, the user chooses a smaller finite set to explore, and then the Cartesian product of these hyperparameters yields several sets of hyperparameters. Grid search uses each set of hyperparameters to train the model, picking the hyperparameter with the smallest validation set error as the best hyperparameter. Therefore, in this study, the range of values of k_1 and k_2 is set to $5 < k_1 < 20$, $5 < k_2 < 20$; the intervals of α, β are both set to $\{0, 0.1, \dots, 1\}$. The grid search method is employed to identify their optimal values, $k_1 = 13, k_2 = 11$ and $\alpha = 0.6, \beta = 0.6$ can be deduced from Fig. 7 and Fig. 8, guided by the principles of nearest neighbor point selection.

5.1.2. Visualization of the outcomes of dimensionality reduction

An attempt has been made to conduct a comparative analysis between the MCGJE algorithm and various existing methods in the field of dimensionality reduction, including the traditional PCA method, graph-based traffic learning algorithms such as NPE, MFA, and ODP, as well as the recently introduced feature space algorithms exemplified by MFSCDP [5]. Additionally, two variations of the MCGJE algorithm were also included in the comparison. The two variants of MCGJE take the form of variant I: local feature line graph and global feature line graph, and variant II: hypermap. For the other five algorithms, a cross-validation method was used to obtain the optimal state parameter values. Where NPE corresponds to the nearest neighbor value of $k = 7$, ODP corresponds to the nearest neighbor value of $k = 9$, MFA corresponds to the eigenmap nearest neighbor value of $k = 8$, and the penalty map nearest neighbor value of $k = 10$, the corresponding parameter values of variants I and II are the same as those of the MCGJE algorithm.



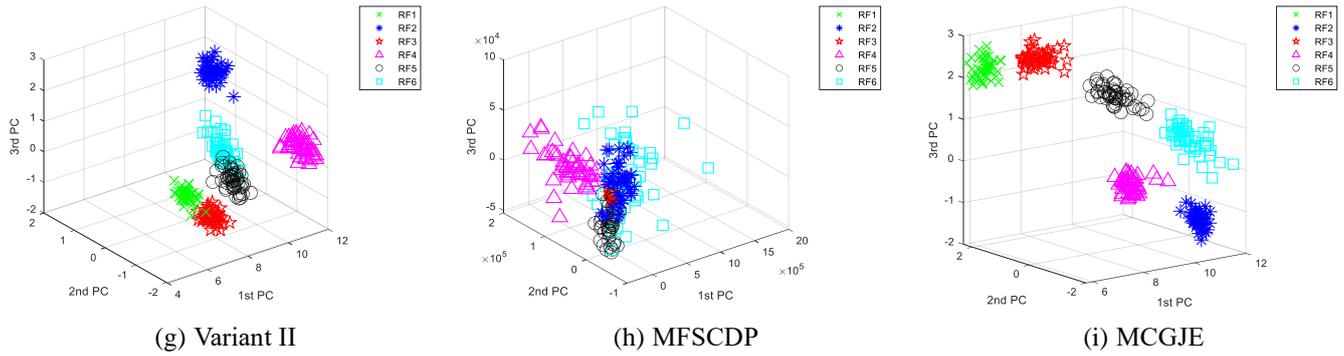


Fig. 9. Different algorithms for 3D visualization.

The MCGJE algorithm demonstrates a superior low-dimensional visualization effect compared to the other seven algorithms, as depicted in Fig. 9, following the process of dimensionality reduction. As can be seen from the figure, according to the algorithm proposed in this paper, samples of the same kind are more clustered into clusters, and dissimilar samples are more far away. Hence, it can be observed that the dimensionality reduction capability of MCGJE surpasses that of the remaining seven algorithms.

5.1.3. The assessment criteria for fault classification and dimension reduction outcomes

The $\delta = \frac{S_B}{S_W}$ [18] evaluation metric is employed to evaluate the classification and clustering effectiveness of the test sample set. Its purpose is to visualize the distinguishability between fault classes within the low-dimensional test set of the MCGJE algorithm. The parameters of the metric are outlined as follows.

$$Z_B = \frac{1}{t} \sum_{i=1}^t (s_i - s)(s_i - s)^T \quad (23)$$

$$Z_{Wi} = \frac{1}{c} \sum_{i=1}^c (y_n^{(i)} - s_i)(y_n^{(i)} - s_i)^T \quad (24)$$

$$\eta = \frac{Z_B}{\sum_{i=1}^c Z_{Wi} p_i} \quad (25)$$

where t is the number of fault categories, s_i is the class i mean, s represents the overall mean, c is the number of samples in each class, p_i represents the prior probability of class i and $p_i = \frac{n_i}{n}$, $y_n^{(i)}$ represents the samples contained in class i , and S_{Wi} represents the scatter of class i . Larger η indicates more compact aggregation of similar classes and more dispersal of dissimilar classes, indicating better dimensionality reduction of the algorithm.

Calculate the differentiability index for each method of dimensionality reduction according to Eq. (25), and the results are shown in Fig. 10. From the figure, it is obvious that the MCGJE algorithm has the largest divisibility index.

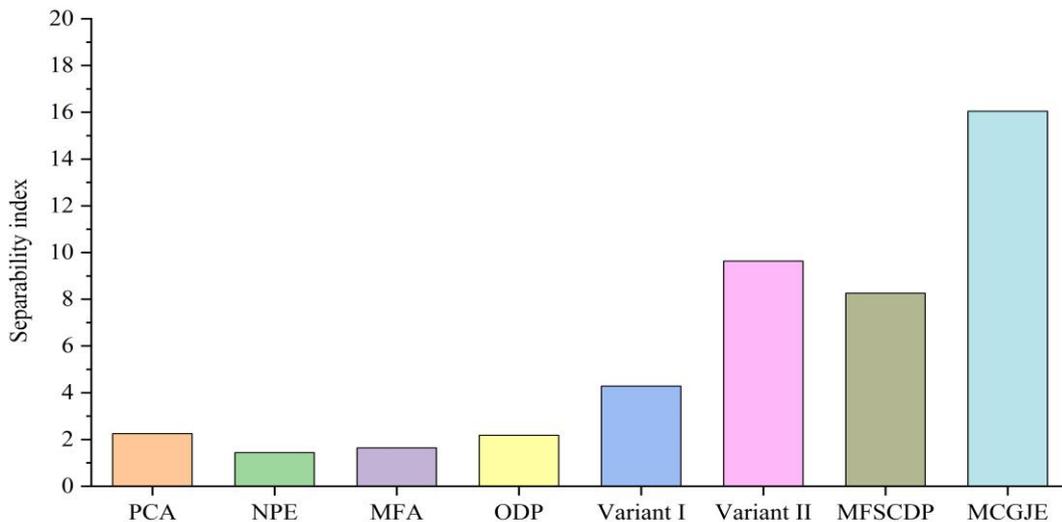


Fig. 10. Algorithm separability index.

In order to better quantify the effect of the established dimensionality reduction, this study inputs the low-dimensional features obtained from the test samples after dimensionality

reduction by each algorithm into the KNN for fault identification, and the identification results are shown in Table 3.

Table 3. Recognition rate of different states.

Dimensionality reduction method	Diagnostic accuracy by status category (%)						Average recognition accuracy (%)
	RF1	RF2	RF3	RF4	RF5	RF6	
PCA	100.0	86.0	100.0	98.0	90.0	82.0	92.7
NPE	50.0	0	2.0	46.0	0	0	16.3
MFA	70.0	0	0	72.0	0	0	23.6
ODP	64.0	76.0	82.0	98.0	74.0	74.0	78.0
Variant I	98.0	91.0	93.0	96.0	93.0	91.0	93.6
Variant II	96.0	92.0	94.0	96.0	96.0	98.0	95.3
MFSCDP	98.0	84.0	90.0	96.0	96.0	84.0	91.3
MCGJE	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Through a comparative and analytical examination of Fig. 9 and Fig. 10, along with the data presented in Table 3, several conclusions can be derived.

1. Both PCA and NPE methods show low performance in terms of 3D visualization and separability indices. This is due to the fact that the NPE method only takes into account the nonlinear structure of the data but ignores the topological structure of the data, while PCA fails to capture the nonlinear correlation information during the dimensionality reduction process.

2. In order to better consider the intrinsic structure of the data, the stream learning algorithms MFA and ODP are widely used, but the dimensionality reduction of the MFA and ODP algorithms is not ideal, and the fundamental reason for this is that they are all embedded in a single graph structure, which is not able to characterize more information about the data.

3. The recently proposed feature space algorithm MFSCDP has improved in terms of separability index and discriminative accuracy. However, it ignores the intricate many-to-many relationships that exist in high-dimensional data and the global information contained in the data.

4. The Variant I algorithm considers local and global information in the median eigenspace but does not consider multivariate structural relationships in high-dimensional data, while the Variant II algorithm only considers multivariate structural relationships in high-dimensional data.

5. Compared with the other seven methods, the MCGJE algorithm considers the inherent defects of the other algorithms. Firstly, it takes into account the local and global structural information of the data by constructing the local median eigenline map and the global median eigenline map; secondly, it takes into account the feature space multivariate structural information by the hypergraph structure on this basis. Therefore, the MCGJE algorithm has the highest separability index and the highest discriminative accuracy.

5.1.4. Experiments with different sample proportions

In order to verify the stability of MCGJE algorithms, the study also used different scales to divide the sample sets. Then, the low-dimensional test sample set after dimensionality reduction of each algorithm was input into the KNN nearest neighbor classifier, and the results are shown in Fig. 11.

Observation shows that the accuracy of NPE and MFA algorithms is generally low; ODP, MFSCDP and variant I algorithms tend to be smooth; PCA and variant II algorithms obviously reach a relatively high recognition accuracy, but not 100% recognition accuracy. The MCGJE algorithm proposed in this paper, on the other hand, has reached 100% recognition accuracy at training set 50 and has been maintaining this recognition accuracy very smoothly; the main reason for this is the use of multi-structural graph co-discrimination.

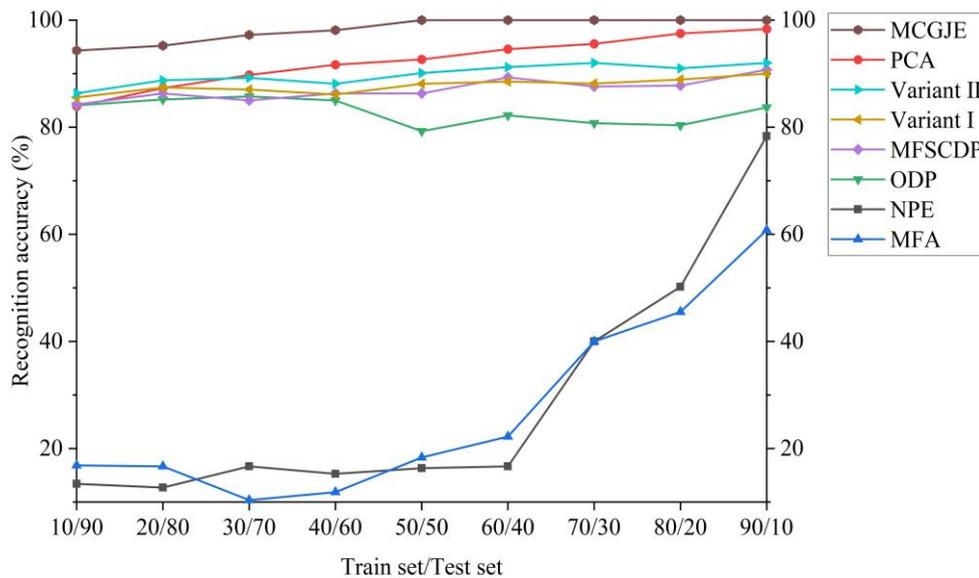


Fig. 11. Experiments with different scales.

5.1.5. Anti-noise experiment

In the actual production of equipment, environmental noise is prevalent, so this subsection is for each algorithm to test its anti-noise performance. Different noises are used to interfere with

the vibration signal, and their signal-to-noise ratios are set to -8dB, -6dB, -4dB, -2dB, 0dB, 2dB, 4dB, 6dB, and 8dB, and then the low-dimensional features are inputted into the KNN to obtain the fault identification results as shown in Fig. 12.

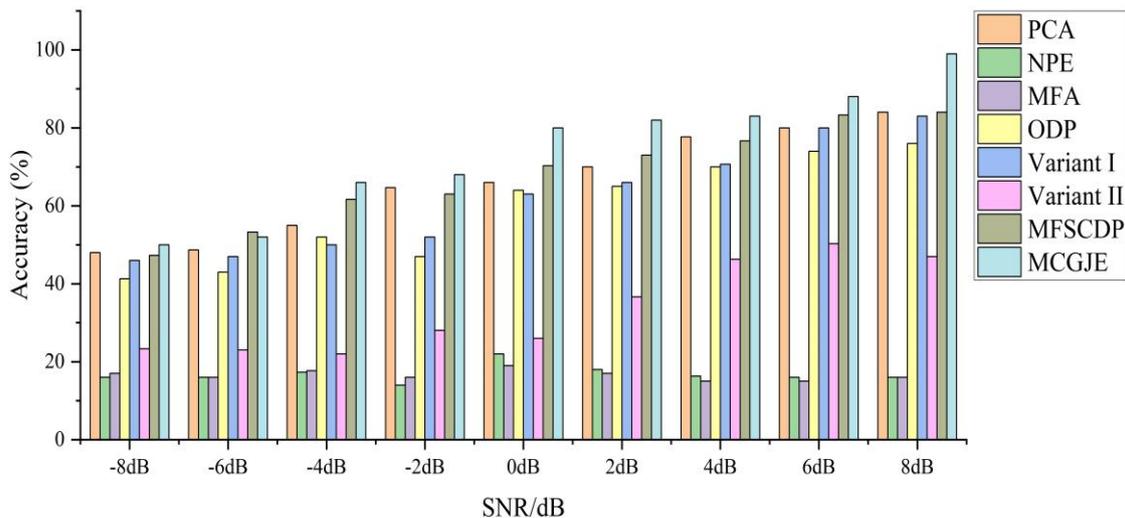


Fig. 12. Anti-noise experiment.

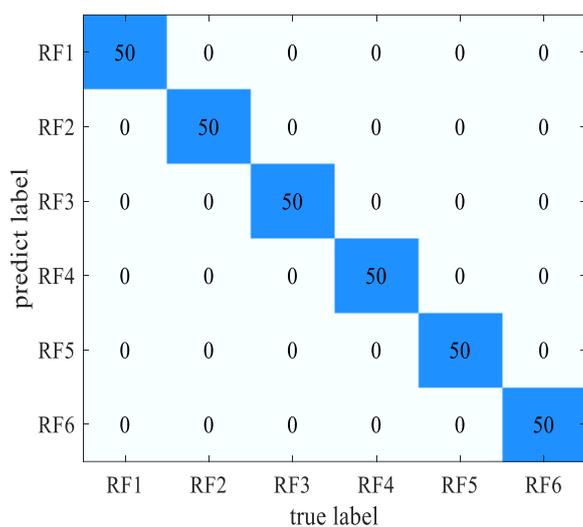
Observation shows that NPE, MFA, and variant II have been at a relatively low level; MFSCDP and variant I are at a higher position, which is due to the fact that they have adopted a feature space structure that enhances the noise immunity of the algorithm very well. However, only the MCGJE algorithm continues to increase in recognition accuracy at a relatively high state as the signal-to-noise ratio increases. This experiment shows that the method proposed in this paper is less sensitive and more robust to noise.

5.1.6. Comparative experiments with multiple classifiers

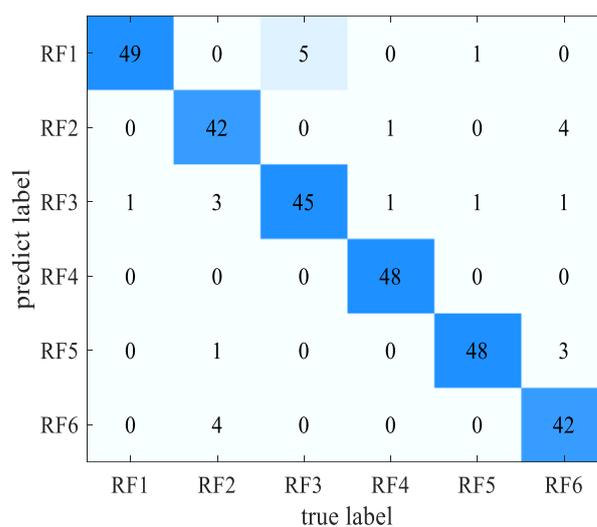
In order to comprehensively analyze the performance of the MCGJE algorithm, the feature sets obtained during the dimensionality reduction process of the MCGJE algorithm are input into three different classifiers: KNN, SVM and BP, and the recognition results of the test samples are visually represented by the confusion matrix, as shown in Fig. 13. In addition, Table 3 provides the recognition accuracies for each classifier.

Table 4. Recognition accuracy of MCGJE with different classifiers.

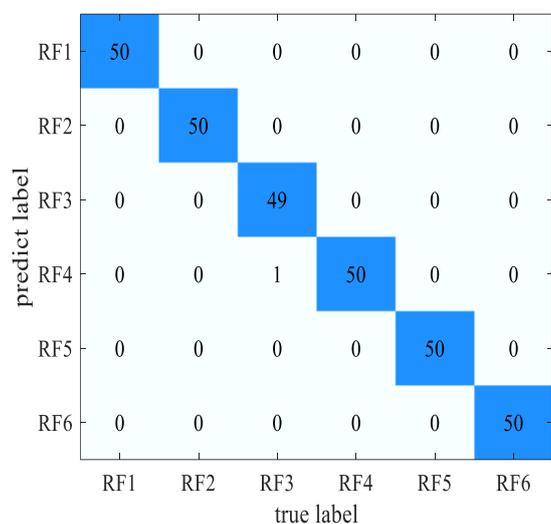
Dimensionality reduction method	Diagnostic accuracy by status category (%)						Average recognition accuracy (%)
	RF1	RF2	RF3	RF4	RF5	RF6	
KNN	100.0	100.0	100.0	100.0	100.0	100.0	100.0
SVM	100.0	100.0	98.00	100.0	100.0	100.0	99.6
BP	98.0	84.0	90.0	96.0	96.0	84.0	91.3



a) KNN



c) BP



b) SVM

Fig. 13. (a)(b)(c) shows the confusion matrix plot for different classifiers.

Combined with Figure 13 and Table 3, it can be seen that MCGJE exhibits strong stability in recognition accuracy in the face of different classifiers.

5.2. Experiment 2

The double-span rotor experimental setup shown in Fig. 14 was used to investigate the generality of the MCGJE algorithm [5]. Twelve eddy current sensors were arranged at each of the six critical sections of this experimental bench. Simulation experiments were carried out on five different states, such as normal, loose, unbalance, misalignment, and touch grinding. The eddy current sensor model is a JX20XL type sensor with a frequency range of 0~10KHz. In the rotational speed 2800r/min, the sampling frequency is 5000Hz, every 1024 sampling points as a sample, the sampling time is set to 20.5s, 100 samples are sampled for each state, according to the 50:50 construction of the training and testing data set, and then according to Table 3 for each channel vibration signal samples to extract 38 feature parameters, a total of $38 \times 12 = 456$ features.

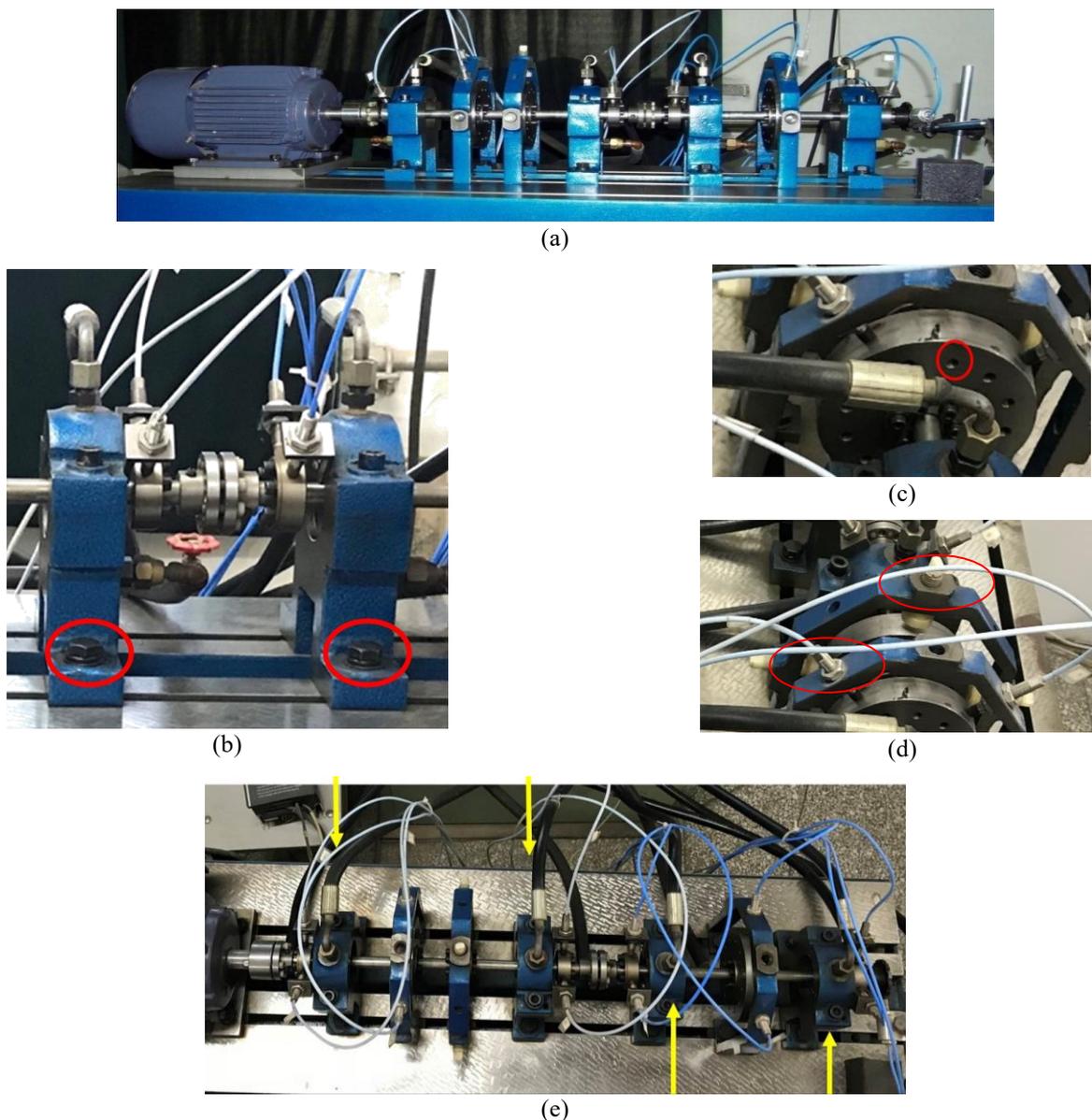


Fig. 14. Double span rotor experimental platform. Where Figure (a) for the normal state, (b) for Bolt looseness, (c) for Mass imbalance, (d) for Rubbing, (e) for Rotor misalignment.

Table 5. Key parameters of the test rig.

Parameters	Value
Shaft 1 length (mm)	415
Shaft 2 length (mm)	350
Diameter of the disc (mm)	120
Diameter of the shaft 1 and shaft 2 (mm)	15
Rated power of the motor (kW)	1.1
Mass of the disc (kg)	2.1
Maximum allowable torque of the coupling (N·m)	20
Sensitivity of the eddy current sensor (V/mm)	10
Speed range of the motor (rpm)	0-12000

To make the article more concise, the rotor fault types and normal states are noted as, Rotor misalignment fault(RMF),

Mass imbalance fault(MIF), Rubbing fault(RF), Bolt looseness fault(BLF), normal state(NS), respectively.

5.2.1. Parameter settings

This experiment employed identical parameters to those utilized in experiment 1, thereby ensuring the reliability and uniformity of the study. Hence, the parameters of the MCGJE algorithm can be defined as follows: $d = 4$ represents the target dimension, while $k_1 = 13, k_2 = 11$ denotes the count of nearest neighbor points within and between classes. The

equilibrium coefficient α, β are set to $\alpha = 0.6, \beta = 0.6$ respectively

5.2.2. Visualization of the outcomes of dimensionality reduction

The experimental platform shown in Fig. 14 was used to conduct comparative experiments to verify the reliability of the algorithms discussed in this thesis.

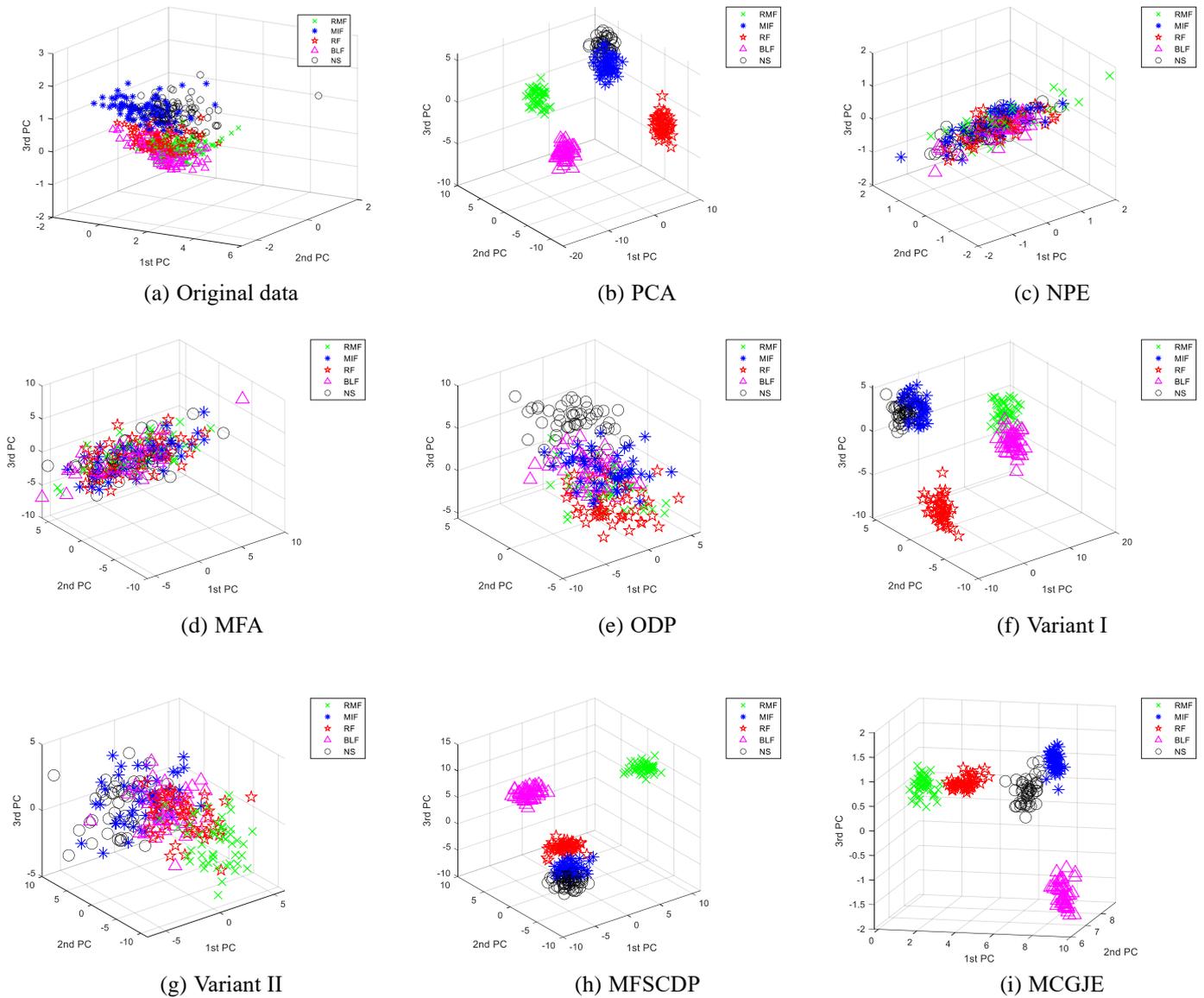


Fig. 15. Different algorithms for 3D visualization.

Observation of Fig. 15 shows that the CDEMSP algorithm has intra-class aggregation and well-defined inter-class intervals, while the other seven algorithms have varying degrees of confusions.

5.2.3. The assessment criteria for fault classification and dimension reduction outcomes

The quantification of the downscaling effect of each algorithm was performed using Eq. (25) and assessed on the experimental rig for the double-span rotor, as depicted in Fig. 14. The separability index is depicted in Fig. 16.

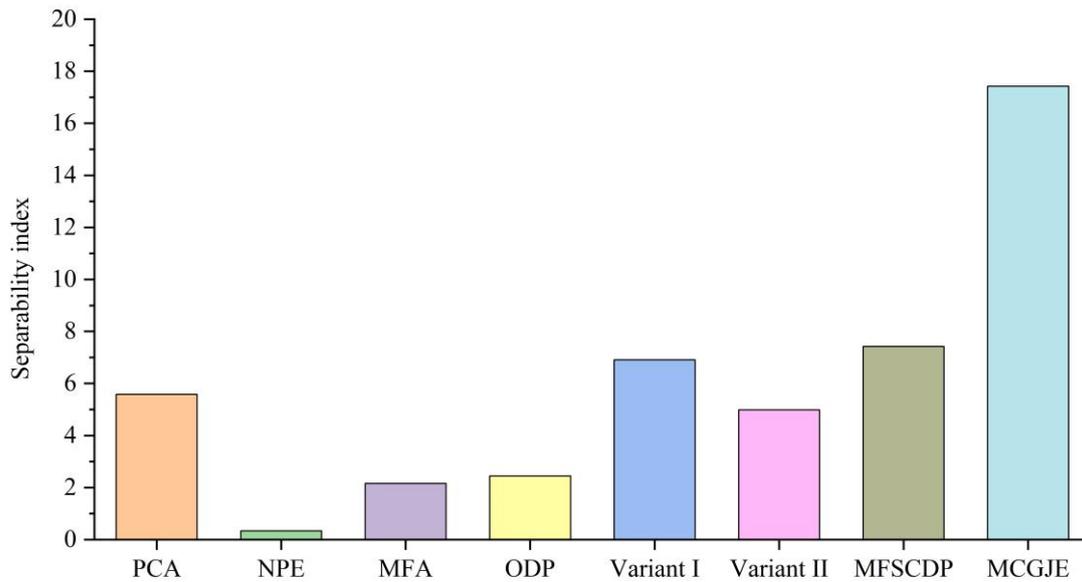


Fig. 16. Algorithm separability index.

To provide a comprehensive description of the algorithm's performance, the KNN classifier is utilized for fault pattern

recognition by inputting low-dimensional samples. The findings are presented in Table 6.

Table 6. Recognition rate of different states.

Dimensionality reduction method	Diagnostic accuracy by status category (%)					Average recognition accuracy (%)
	RMF	MIF	RF	BLF	NS	
PCA	100.0	96.00	100.0	100.0	100.0	99.2
NPE	54.0	2.0	4.0	0.0	60.0	24.0
MFA	2.0	0.0	98.0	0.0	0.0	20.0
ODP	44.0	74.0	30.0	12.0	98.0	51.6
Variant I	100.0	98.0	100.0	100.0	96.0	98.8
Variant II	100.0	88.0	100.0	100.0	94.0	96.4
MFSCDP	100.0	96.0	100.0	100.0	98.0	98.8
MCGJE	100.0	100.0	100.0	100.0	100.0	100.0

6. Conclusion

In order to solve the problem that traditional graph embedding algorithms cannot accurately describe the intrinsic real relationship between data, thus reducing the accuracy of fault classification. In this thesis, the MCGJE algorithm is proposed from the perspective of multi-class graph joint embedding. The algorithm not only takes into account the information of local and global feature space, but also makes full use of the hypergraph structure to portray the multivariate relationships of high-dimensional data in feature space, so as to extract more sensitive low-dimensional features. In addition, the MCGJE

algorithm has better generalizations ability and noise resistance, which is inseparable from the use of feature space structure. By using the rotor fault dataset, we performed experiments on anti-noise, variable scale, effect of different classifiers, three-dimensional visualization, and multiple method separability metrics. The same experiments were conducted using another testbed to validate the implementation and the reliability of the proposed algorithms. Based on the results of empirical analyses on two rotor test benches, it can be concluded that the algorithm significantly improves the accuracy of recognition. This leads to more reliable diagnostic results. However, it is worth noting that the following aspects of this study deserve further

investigation:

1. Currently, we use feature lines ($P = 2$) to make connections to the hypergraph structure. In subsequent work, how to combine eigenfaces and special holes with hypergraph structures will be the next focus of attention.

2. In real engineering environments, where most of the

data collected is unlabeled, the question of how to use limited labelled data to predict unlabeled data is one that should be given more consideration.

3. In addition to the application of the algorithm to rotor faults, how to apply it to more areas, such as bearing and gearbox faults, has become a key concern.

Acknowledgements

This study was jointly funded by the Special Project of the National Natural Science Foundation of China (No. 62241308) and the National Natural Science Foundation of China (No. 51675253).

References

1. Alshingiti Z, Alaqel R, Al-Muhtadi J, Haq QEU, Saleem K, Faheem MH. A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM-CNN. *Electronics*. 2023; 12(1): 232. <https://doi.org/10.3390/electronics12010232>.
2. Banach R, Jeske C. Simple feature engineering via neat default retrenchments. *Journal of Logic and Algebraic Programming*. 2011; 80(8): 453-80. <https://doi.org/10.1016/j.jlap.2010.12.001>.
3. Chen F, Cheng M, Tang B, Chen B, Xiao W. Pattern recognition of a sensitive feature set based on the orthogonal neighborhood preserving embedding and adaboost_SVM algorithm for rolling bearing early fault diagnosis. *Measurement Science and Technology*. 2020; 31(10). <https://doi.org/10.1088/1361-6501/ab8c11>.
4. Chen Y-N. Multiple Kernel Feature Line Embedding for Hyperspectral Image Classification. *Remote Sensing*. 2019; 11(24). <https://doi.org/10.3390/rs11242892>.
5. Dong X, Zhao R, Yuan J, Chen P, He T, Wei K. Dimensionality reduction method based on multiple feature-space collaborative discriminative projection for rotor fault diagnosis. *Measurement Science and Technology*. 2023; 34(5). <https://doi.org/10.1088/1361-6501/acb454>.
6. Du H, Wang S, Li T. A nearest feature space embedding method based on the combination of nonlinear distance metric and included angle. *Computer Engineering and Science*. 2018; 40(5): 888-97. <http://dx.doi.org/10.3969/j.issn.1007-130X.2018.05.018>.
7. Gao D, Huang K, Zhu Y, Zhu L, Yan K, Ren Z and Soares C G. Semi-supervised small sample fault diagnosis under a wide range of speed variation conditions based on uncertainty analysis. *Reliability Engineering & System Safety*. 2024; 242. <https://doi.org/10.1016/j.ress.2023.109746>.
8. Gao D, Zhu Y, Yan K, Fu H, Ren Z, Kang W and Soares C G. Joint learning system based on semi-pseudo-label reliability assessment for weak-fault diagnosis with few labels. *Mechanical Systems and Signal Processing*. 2023; 189. <https://doi.org/10.1016/j.ymsp.2022.110089>.
9. Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Molecular Diversity*. 2021; 25(3): 1315-60. <https://doi.org/10.1007/s11030-021-10217-3>.
10. Kang Y, Chen G, Wang H, Pan W, Wei X. A New Dual-Input Deep Anomaly Detection Method for Early Faults Warning of Rolling Bearings. *Sensors (Basel, Switzerland)*. 2023; 23(18). <https://doi.org/10.3390/s23188013>.
11. Li B, Wang C, Huang D-S. Supervised feature extraction based on orthogonal discriminant projection. *Neurocomputing*. 2009; 73(1-3): 191-6. <https://doi.org/10.1016/j.neucom.2008.09.030>.
12. Li X, Su K, He Q, Wang X, Xie Z. Research on Fault Diagnosis of Highway Bi-LSTM Based on Attention Mechanism. *Eksploatacja i Niezawodność-Maintenance and Reliability*. 2023; 25(2). <https://doi.org/10.17531/ein/162937>.
13. Li Y, Feng C. A nonlinear method for monitoring industrial process. *Transactions of the Institute of Measurement and Control*. 2021; 43(2): 400-11. <https://doi.org/10.1177/0142331220959232>.
14. Lyu Y, Zhang Q, Chen A, Wen Z. Interval Prediction of Remaining Useful Life based on Convolutional Auto-Encode and Lower Upper Bound Estimation. *Eksploatacja i Niezawodność-Maintenance and Reliability*. 2023; 25(2). <https://doi.org/10.17531/ein/165811>.
15. Peng B, Wan S, Bi Y, Xue B, Zhang M. Automatic Feature Extraction and Construction Using Genetic Programming for Rotating Machinery Fault Diagnosis. *Ieee Transactions on Cybernetics*. 2021; 51(10): 4909-23. <https://doi.org/10.1109/tcyb.2020.3032945>.

16. Segovia Ramirez I, Mohammadi-Ivatloo B, Garcia Marquez FP. Alarms management by supervisory control and data acquisition system for wind turbines. *Eksploatacja I Niezawodnosc-Maintenance and Reliability*. 2021; 23(1): 110-6. <https://doi.org/10.17531/ein.2021.1.12>.
17. Shao H, Xia M, Han G, Zhang Y, Wan J. Intelligent Fault Diagnosis of Rotor-Bearing System Under Varying Working Conditions With Modified Transfer Convolutional Neural Network and Thermal Images. *Ieee Transactions on Industrial Informatics*. 2021; 17(5): 3488-96. <https://doi.org/10.1109/tii.2020.3005965>.
18. Shi M, Zhao R. A Method of Mechanical Fault Diagnosis Based on Locality Margin Discriminant Projection. *Journal of Vibration, Measurement and Diagnosis*. 2021; 41(1): 126-32. <http://dx.doi.org/10.16450/j.cnki.issn.1004-6801.2021.01.018>.
19. Shi X, He S, Xie X, Sun Y. Review on Feature Extraction of Lubrication and Wear Fault Diagnosis in Tribology System. *Tribology*. 2023; 43(3): 241-55. <http://dx.doi.org/10.16078/j.tribology.2021066>.
20. Shu X, Zhang S, Li Y, Chen M. An anomaly detection method based on random convolutional kernel and isolation forest for equipment state monitoring. *Eksploatacja I Niezawodnosc-Maintenance and Reliability*. 2022; 24(4): 758-70. <https://doi.org/10.17531/ein.2022.4.16>.
21. Sun S, Przystupa K, Wei M, Yu H, Ye Z, Kochan O. Fast bearing fault diagnosis of rolling element using Levy Moth-Flame optimization algorithm and Naive Bayes. *Eksploatacja I Niezawodnosc-Maintenance and Reliability*. 2020; 22(4): 730-40. <https://doi.org/10.17531/ein.2020.4.17>.
22. Wang D, Zhang D, Tang M, Zhang H, Sun T, Yang C, et al. Ethylene chlorotrifluoroethylene/hydrogel-based liquid-solid triboelectric nanogenerator driven self-powered MXene-based sensor system for marine environmental monitoring. *Nano Energy*. 2022;100. <https://doi.org/10.1016/j.nanoen.2022.107509>.
23. Wang G, Shang G, Pu P, Li X, Peng H. Fake Review Identification Methods Based on Multidimensional Feature Engineering. *Mobile Information Systems*. 2022; 2022. <https://doi.org/10.1155/2022/5229277>.
24. Wang J, Gao D, Zhu Y, Ren Z, Zhao R, Lin T, et al. Match-reinforcement learning with time frequency selection for bearing fault diagnosis. *Measurement Science and Technology*. 2023; 34(12). <https://doi.org/10.1088/1361-6501/ace644>.
25. Wang J, Ye L, Gao RX, Li C, Zhang L. Digital Twin for rotating machinery fault diagnosis in smart manufacturing. *International Journal of Production Research*. 2019; 57(12): 3920-34. <https://doi.org/10.1080/00207543.2018.1552032>.
26. Yan X, Liu Y, Zhang C-a. Multiresolution Hypergraph Neural Network for Intelligent Fault Diagnosis. *Ieee Transactions on Instrumentation and Measurement*. 2022; 71. <https://doi.org/10.1109/tim.2022.3212532>.
27. Yang C, Ma S, Han Q. Robust discriminant latent variable manifold learning for rotating machinery fault diagnosis. *Engineering Applications of Artificial Intelligence*. 2023; 126. <https://doi.org/10.1016/j.engappai.2023.106996>.
28. Yuan J, Zhao R, He T, Chen P, Wei K, Xing Z. Fault diagnosis of rotor based on Semi-supervised Multi-Graph Joint Embedding. *Isa Transactions*. 2022; 131: 516-32. <https://doi.org/10.1016/j.isatra.2022.05.006>.
29. Zhang J, Yin Z, Chen P, Nichele S. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. *Information Fusion*. 2020; 59: 103-26. <https://doi.org/10.1016/j.inffus.2020.01.011>.
30. Zhang K, Li H, Cao S, Yang C, Sun F, Wang Z. Motor current signal analysis using hypergraph neural networks for fault diagnosis of electromechanical system. *Measurement*. 2022; 201. <https://doi.org/10.1016/j.measurement.2022.111697>.
31. Zhao X, Jia M. Fault diagnosis of rolling bearing based on feature reduction with global-local margin Fisher analysis. *Neurocomputing*. 2018; 315: 447-64. <https://doi.org/10.1016/j.neucom.2018.07.038>.
32. Zhao X, Yao J, Deng W, Jia M and Liu Z. Normalized Conditional Variational Auto-Encoder with adaptive Focal loss for imbalanced fault diagnosis of Bearing-Rotor system. *Mechanical Systems and Signal Processing*. 2022; 170. <https://doi.org/10.1016/j.ymsp.2022.108826>.
33. Zhong G, Zhang K, Wei H, Zheng Y, Dong J. Marginal Deep Architecture: Stacking Feature Learning Modules to Build Deep Learning Models. *Ieee Access*. 2019; 7: 30220-33. <https://doi.org/10.1109/access.2019.2902631>.
34. Zhu X, Zhao X, Yao J, Deng W, Shao H and Liu Z. Adaptive Multiscale Convolution Manifold Embedding Networks for Intelligent Fault Diagnosis of Servo Motor-Cylindrical Rolling Bearing Under Variable Working Conditions. *Ieee-Asme Transactions on Mechatronics*. 2023. <https://doi.org/10.1109/tmech.2023.3314215>.