

A Novel Three-stage Feature Fusion Methodology and its Application in Degradation State Identification for Hydraulic Pumps

Mochao Pei, Hongru Li, He Yu

Army Engineering University, Shijiazhuang 050003, China

Correspondence to: Hongru Li/ lihr168@sohu.com

The performance of feature is essential to the degradation state identification for hydraulic pumps. The initial feature set extracted from the vibration signal of the hydraulic pump is often high-dimensional and contains redundant information, which undermines the effectiveness of the feature set. The novel three-stage feature fusion scheme proposed in this paper aims to enhance the performance of the original features extracted from the vibration signal. First, sparse local Fisher discriminant analysis (SLFDA) performs intra-set fusion within the two original feature sets, respectively. SLFDA has a good effect on samples with intra-class multimodality, and the feature set fused by it has obvious multivariate normal distribution characteristics, which is conducive to the next fusion. Second, our modified intra-class correlation analysis (MICA) is used to fuse two feature sets in the second stage. MICA is a CCA (Canonical correlation analysis) -based method. A new class matrix is used to modify the covariance matrix between two feature sets, which allows MICA to conveniently inherit the discriminating structure while fusing features. Finally, we propose a feature selection algorithm based on kernel local Fisher discriminant analysis (KLFDA) and kernel canonical correlation analysis (KCCA) to select the desired features. This algorithm based on Max-Relevance and Min-Redundancy (mRMR) framework solves the problem that CCA cannot properly evaluate the correlation between features and the class variable, as well as accurately evaluates the correlation among features. Based on the experimental data, the proposed method is compared with several popular methods, and the feature fusion methods used in some previous studies related to the fault diagnosis of rotating machinery are compared with it as well. The results show that the fusion effectiveness of our method is better than other methods, which obtains higher recognition accuracy.

Keywords: Feature fusion, hydraulic pump, Sparse local Fisher discriminant analysis(SLFDA), Canonical correlation analysis(CCA).

1. INTRODUCTION

Hydraulic pump is one of the important components of a hydraulic system, which directly influences the entire system's security and reliability [1]. Safety accidents or even economic losses are likely to be caused by the failure of a hydraulic pump. Consequently, monitoring the performance degradation process of hydraulic pumps is essential for assuring the system's safe operation [2].

Nowadays, increasing attention has been paid to equipment condition monitoring with vibration analysis methods. The vibration signal obtained from the acceleration sensor is widely used in condition monitoring for hydraulic pumps due to its fast reactivity, and it has obvious non-stationarity and nonlinearity as the degree of degradation increases [3]. The interactions among the internal friction pairs are transmitted to the pump case in the form of vibration. Vibration-based analysis generally consists of four steps: data collection and analysis, feature extraction, model establishment and training, and pattern recognition [4]. Feature extraction is a

crucial step directly related to the accuracy and timeliness of degradation state identification, which has attracted the attention of many researchers for many years. Yonggang Xu et al. [5] proposed empirical scanning spectrum kurtosis for extracting the fault features of rolling element bearing. Hongru Li et al. [6] applied bispectral entropy to the fault diagnosis of hydraulic pump. A team composed of Yılmaz Kaya, Samet Bayram, Melih Kuncan, Kaplan Kaplan et al. has been committed to the fault diagnosis of bearing for many years, and has proposed many interesting feature extraction algorithms [7]-[11]. To meet the actual needs, equipment fault information often requires multiple or even high-dimensional features to be fully expressed. Some of the methods mentioned above do cause high-dimensional feature sets. Nevertheless, a high-dimensional feature set generally includes redundant information, which will weaken the effectiveness of fault diagnosis or degradation state identification [12], as well as cause a more time-consuming phenomenon. Thus, dimensionality reduction, redundant

information elimination, and fault information enhancement are just the tasks that need to be handled properly.

Degradation state identification for hydraulic pumps belongs to the field of pattern recognition. It is well known that information fusion is effective for improving recognition reliability and accuracy. Applications of information fusion can be found in many fields, such as image analysis, odour classification, fault diagnosis, and military decision-making. Information fusion can be classified into three levels: signal level, feature level, and decision level [13]. As a feature extraction step, feature fusion can obtain more comprehensive and polished information than signal level fusion. It utilizes the correlation of features at an early stage, and learning is performed once on the fused features [14]. Yongbing Liu et al. [15] used the kernel trick to map the original feature space into a high-dimensional space, and then applied joint approximate diagonalization of eigen-matrices to fuse the features in the new space. Hanshu Cai et al. [16] used a linear combination method to fuse electroencephalogram features of different modalities, and the genetic algorithm was used to conduct feature weighting. In Ye Tian's paper about fault diagnosis for machinery components [17], t-distributed stochastic neighbor embedding (t-SNE) was employed for feature fusion, which can enhance the separability information of features. The content described above shows that feature fusion has quite a beneficial effect on dimensionality reduction, redundant information elimination and fault information enhancement.

However, the fault data of hydraulic pump always shows non-Gaussian and non-linear characteristics, and its sample distribution often hides manifold structures. It seems probable, then, that the traditional linear fusion methods or some methods which cannot comprehensively take the local and global information of data into consideration do not work. As an important way to deal with nonlinear problems, the localization method reveals the low-dimensional manifold structure hidden in the data, so it has a remarkable application in feature fusion. Locally linear embedding (LLE) proposed by Roweis [18] assumes that any data point in the high-dimensional space can be reconstructed by local neighbors, and this reconstruction relationship remains unchanged in the low-dimensional space. Locality preserving projections (LPP) proposed by He et al. [19] keep the local relationships among the samples unchanged before and after fusion. Local Fisher discriminant analysis (LFDA) proposed by Masashi Sugiyama [20] effectively combines the ideas of LPP and Fisher discriminant analysis (FDA), which was first used for facial expression recognition. LFDA was also extended to kernel local Fisher discriminant analysis (KLFDA) by applying the kernel trick in his paper. LFDA is very suitable for samples with within-class multimodality. SLFDA proposed by Zhan Wang et al. [21] is developed based on LFDA, and it has a stronger ability to obtain discriminative information from feature sets.

Before feature fusion, multiple original feature sets may be extracted from the vibration signals of hydraulic pump, and the correlation among them needs to be considered. Canonical correlation analysis (CCA) [22] is a multivariate statistical method to study the correlation between two sets of

variables. It achieves the purpose of understanding the correlation between two groups of variables by studying the correlation between linear combinations of variables. In recent years, many variants of CCA have been proposed one after another, such as constrained CCA, nonlinear CCA, multiset CCA, and supervised local CCA [23]. In traditional CCA-based feature fusion, linear discriminant analysis (LDA) [24], principal component analysis (PCA) [25] are usually used in advance to reduce the dimensionality of each feature set, so as to assure the invertibility and non-singularity of the covariance matrix. Xiaoquan Ke et al. [26] integrated Fisher discriminant matrix, PCA and CCA in their feature fusion scheme, and the experimental results showed that it achieved good results in ship recognition. The premise of CCA is that the data follows a Gaussian distribution. In practice, non-Gaussian data may inhibit the effect of CCA. Zhiwen Chen et al. [27] proposed a generalized CCA-based fault detection approach with the randomized algorithm, which can monitor the non-Gaussian process. Kernel trick was also employed to study the non-linear correlation between feature sets, and kernel canonical correlation analysis (KCCA) was proposed [28]. To utilize class information, some variants of CCA with supervised properties have been proposed. Multiview supervised CCA (MSCCA) [29] considers the class information from between-view and within-view. Zuobin Wu et al. [30] proposed a feature fusion method that eliminates redundancy through intra-class and extra-class discriminative correlation analysis (IEDCA-IRE), which is an extended method based on CCA. To inherit the discriminative information of the feature sets, a class matrix is introduced into the CCA-based method. Feature selection based on Max-Relevance and Min-Redundancy (mRMR) [31] is improved to eliminate the irrelevant redundancy.

Although CCA has provided meaningful results in feature fusion, it still has some limitations. First, due to its unsupervised property, it cannot take the information provided by the class label into consideration. As mentioned above, some approaches have been proposed to solve this problem, but they may not be suitable for dealing with the degradation features of hydraulic pump. The non-stationarity of the vibration signal usually makes samples in a class multimodal. Consequently, when considering class information, we should investigate the structure of the within-class and between-class locally. Second, CCA maximizes the correlation of each canonical variable pair, which makes a certain amount of redundancy exist between feature sets. Third, CCA runs on the assumption that the data follow a multivariate normal distribution (MND). When it is used to evaluate the correlation between feature sets and class labels in a classification problem, inaccurate results may be obtained because the labels do not follow the normal distribution. In addition, for data that significantly does not follow the MND, CCA cannot obtain a good fusion result. Focusing on these issues, we propose a three-stage fusion methodology for the degradation features of hydraulic pumps. In the first stage, for two given feature sets, SLFDA is employed to fuse the features in each set. In this study, the data fused by SLFDA has significant MND characteristics.

The intra-class correlation analysis (ICA) proposed in [32] is modified to fuse the two feature sets in the second stage. Modified intra-class correlation analysis (MICA), a method based on CCA, can also inherit the class structure that contains rich discriminative information after the first stage of fusion. In the third stage, a feature selection algorithm based on KLFDA and KCCA (mRMR-KCLFDA) is proposed to remove redundant features. We use KLFDA to evaluate the relevance between the class variable and feature candidates, while KCCA evaluates the correlation between the selected features and the feature candidate.

The following contents are also presented: LFDA, CCA, etc. related to this study are briefly reviewed in Section 2. In Section 3, we elaborate on the theoretical basis and procedures of the proposed three-stage fusion. In Section 4, three experimental cases are conducted to illustrate the superiority and availability of the proposed approach. The last section summarizes the full paper and gives several conclusions.

2. THEORETICAL BACKGROUND

2.1. FDA and its improved versions

This subsection reviews existing FDA, LFDA, SLFDA and KLFDA theories related to this study. The basic idea of FDA is to project k classes of p dimensional data into certain directions so that the between-class data can be separated as much as possible after projection. Given a data matrix $X \in R^{n \times p}$ with n samples and p features, the sum of squares vector within classes \mathbf{S}_E and the sum of squares vector between classes \mathbf{S}_G are defined as follows:

$$\mathbf{S}_E = \mathbf{U}^T \mathbf{A} \mathbf{U} \quad (1)$$

$$\mathbf{S}_G = \mathbf{U}^T \mathbf{B} \mathbf{U} \quad (2)$$

where \mathbf{U} is the projection matrix, \mathbf{A} and \mathbf{B} represent the within-class scatter matrix and between-class scatter matrix, respectively, which can be calculated by the following two equations:

$$\mathbf{A} = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_j^{(i)} - \bar{X}^{(i)})(X_j^{(i)} - \bar{X}^{(i)})^T \quad (3)$$

$$\mathbf{B} = \sum_{i=1}^k n_i (\bar{X}^{(i)} - \bar{X})(\bar{X}^{(i)} - \bar{X})^T \quad (4)$$

where k , $\bar{X}^{(i)}$, n_i , \bar{X} and $X_j^{(i)}$ are the number of classes, mean of the i -th class, number of the i -th class, mean of all the n samples, and the j -th sample of the i -th class, respectively. If the differences among the means of the k classes samples are significant, the $\Delta(\mathbf{u})$ value calculated by the following equation should be sufficiently great.

$$\Delta(\mathbf{u}) = \frac{\mathbf{S}_G}{\mathbf{S}_E} \quad (5)$$

FDA seeks the projection matrix \mathbf{U}_{opt} that meets the following criterion:

$$\begin{cases} \mathbf{U}_{\text{opt}} = \arg \max_{\mathbf{U}} \mathbf{S}_G \\ \text{s.t. } \mathbf{S}_E = \mathbf{I} \end{cases} \quad (6)$$

After the Lagrange multiplier method is employed, the following equation can be obtained:

$$(A^{-1}B)u_f = \Delta(u)_f u_f \quad (7)$$

where $\Delta(\mathbf{u}) = [\Delta(u)_1, \Delta(u)_2, \dots, \Delta(u)_e]$ consists of the first e largest eigenvalues of $A^{-1}B$, and $\mathbf{U} = [u_1, u_2, \dots, u_e]$ consists of the corresponding eigenvectors.

LFDA is developed by combining the ideas of LPP and FDA. In LFDA, the local between-class scatter matrix $\hat{\mathbf{B}}$ and local within-class scatter matrix $\hat{\mathbf{A}}$ are defined as follows [20]:

$$\hat{\mathbf{B}} = \frac{1}{2} \sum_{g,m=1}^n \hat{W}_{g,m}^b (x_g^{cg} - x_m^{cm})(x_g^{cg} - x_m^{cm})^T \quad (8)$$

$$\hat{\mathbf{A}} = \frac{1}{2} \sum_{g,m=1}^n \hat{W}_{g,m}^a (x_g^{cg} - x_m^{cm})(x_g^{cg} - x_m^{cm})^T \quad (9)$$

where the superscripts cg and cm represent the class labels of samples x_g and x_m , respectively, and

$$\hat{W}_{g,m}^b = \begin{cases} P_{g,m} (1/n - 1/n_i) & \text{if } cg = cm = i \\ 1/n & \text{otherwise} \end{cases} \quad (10)$$

$$\hat{W}_{g,m}^a = \begin{cases} P_{g,m} / n_i & \text{if } cg = cm = i \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

the closeness between x_g and x_m is quantified by the affinity value $P_{g,m}$, which is defined as:

$$P_{g,m} = \exp(\|x_g - x_m\|^2 / \eta_g \eta_m) \quad (12)$$

where η_g is the local scaling of x_g , and η_m is the local scaling of x_m .

LFDA seeks the projection matrix $\hat{\mathbf{U}}_{\text{opt}}$ that meets the following criterion:

$$\begin{cases} \hat{\mathbf{U}}_{\text{opt}} = \arg \max_{\hat{\mathbf{U}}} \hat{\mathbf{S}}_G \\ \text{s.t. } \hat{\mathbf{S}}_E = \mathbf{I} \end{cases} \quad (13)$$

where

$$\hat{\mathbf{S}}_E = \hat{\mathbf{U}}^T \hat{\mathbf{A}} \hat{\mathbf{U}} \quad (14)$$

$$\hat{\mathbf{S}}_G = \hat{\mathbf{U}}^T \hat{\mathbf{B}} \hat{\mathbf{U}} \quad (15)$$

SLFDA is developed based on LFDA, which removes the null space of the local mixture scatter matrix of LFDA that has no discriminative information. The local mixture scatter matrix \mathbf{M}_{ls} is defined as:

$$\mathbf{M}_{ls} = \frac{1}{2} \sum_{g,m=1}^n \hat{W}_{g,m}^{ls} (\mathbf{x}_g^{cg} - \mathbf{x}_m^{cm})(\mathbf{x}_g^{cg} - \mathbf{x}_m^{cm})^T \quad (16)$$

where

$$\hat{W}_{g,m}^{ls} = \begin{cases} P_{g,m} / n & \text{if } cg = cm \\ 1 / n & \text{otherwise} \end{cases} \quad (17)$$

Suppose that the eigenvalue decomposition is applied to \mathbf{M}_{ls} as follows:

$$\mathbf{M}_{ls} = \mathbf{D}\mathbf{\Sigma}\mathbf{D}^{-1} \quad (18)$$

A matrix \mathbf{V}_{opt} is defined in advance as:

$$\begin{cases} \mathbf{V}_{opt} = \arg \max_{\mathbf{V}} (\mathbf{V}^T \bar{\mathbf{B}} \mathbf{V}) \\ \text{s.t. } \mathbf{V}^T \bar{\mathbf{A}} \mathbf{V} = \mathbf{I} \end{cases} \quad (19)$$

where $\bar{\mathbf{B}} = \mathbf{D}^T \hat{\mathbf{B}} \mathbf{D}$ and $\bar{\mathbf{A}} = \mathbf{D}^T \hat{\mathbf{A}} \mathbf{D}$.

SLFDA seeks the projection matrix $\bar{\mathbf{U}}_{opt}$ that meets the following criterion:

$$\begin{cases} \bar{\mathbf{U}}_{opt} = \arg \min_{\bar{\mathbf{U}}} \|\bar{\mathbf{U}}\|_1 \\ \text{s.t. } \mathbf{D}^T \bar{\mathbf{U}} = \mathbf{V}_{opt} \end{cases} \quad (20)$$

The linearized Bregman iteration [33] is extended to solve problem (20), and the iteration process is expressed as:

$$\begin{cases} E^{h+1} = E^h - D(D^T \bar{U}^h - V_{opt}) \\ \bar{U}^{h+1} = \xi \Gamma_{\mu}(E^{h+1}) \end{cases} \quad (21)$$

where $E^0 = \bar{U}^0 = 0$, ξ is a positive parameter, $\Gamma_{\mu}(E) = [\Gamma_{\mu}(E_{g,m})]$ and

$$\begin{cases} \Gamma_{\mu}(E) = [\Gamma_{\mu}(E(1)), \Gamma_{\mu}(E(2)), \dots, \Gamma_{\mu}(E(n))]^T \\ \Gamma_{\mu}(\alpha) = \begin{cases} 0 & \text{if } |\alpha| \leq \mu \\ \text{sign}(\alpha)(|\alpha| - \mu) & \text{otherwise} \end{cases} \end{cases} \quad (22)$$

For other unfinished details, please refer to [21].

LFDA can be extended to a non-linear variant KLFDA through the kernel trick. Similar to the form of equation (7), the solution of the following generalized eigenvalue problem is considered in the process of KLFDA [20]:

$$\mathbf{K} \hat{\mathbf{L}}^b \mathbf{K} \alpha = \lambda (\mathbf{K} \hat{\mathbf{L}}^a \mathbf{K} + r \mathbf{I}_n) \alpha \quad (23)$$

In equation (23), r is a regularization parameter with a small value. \mathbf{K} is the kernel matrix. \mathbf{I}_n is an identity matrix. In particular, the (g, m) -th element of the Gaussian kernel matrix used in this paper is given as:

$$K_{g,m} = K(x_g, x_m) = \exp\left(-\frac{\|x_g - x_m\|^2}{2\sigma^2}\right) (g, m = 1, 2, \dots, n) \quad (24)$$

where $\|x_g - x_m\|$ is equivalent to the Euclidean distance between x_g and x_m . $-1/(2\sigma^2)$ is a kernel parameter.

$\hat{\mathbf{L}}^b$ and $\hat{\mathbf{L}}^a$ are two square matrices defined by:

$$\begin{cases} \hat{\mathbf{L}}^a \equiv \hat{\mathbf{D}}^a - \hat{\mathbf{W}}^a \\ \hat{\mathbf{L}}^b \equiv \hat{\mathbf{L}}^{ls} - \hat{\mathbf{L}}^a \\ \hat{\mathbf{L}}^{ls} \equiv \hat{\mathbf{D}}^{ls} - \hat{\mathbf{W}}^{ls} \end{cases} \quad (25)$$

Where $\hat{\mathbf{D}}^a$ and $\hat{\mathbf{D}}^{ls}$ are two diagonal matrices, and their diagonal elements are given as follows:

$$\begin{cases} \hat{D}_{g,g}^a = \sum_{m=1}^n \hat{W}_{g,m}^a \\ \hat{D}_{g,g}^{ls} = \sum_{m=1}^n \hat{W}_{g,m}^{ls} \end{cases} \quad (26)$$

It is noted that FDA reduces the dimensionality of the feature set to $k - 1$ at most, where k is the number of classes, while LFDA, SLFDA and KLFDA have not this limitation.

2.2. CCA and KCCA

Suppose a data matrix \mathbf{Z} with n samples and $(p + q)$ features has the following form:

$$\mathbf{Z} = [\mathbf{X}, \mathbf{Y}] = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} & y_{11} & y_{12} & \dots & y_{1q} \\ x_{21} & x_{22} & \dots & x_{2p} & y_{21} & y_{22} & \dots & y_{2q} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} & y_{n1} & y_{n2} & \dots & y_{nq} \end{bmatrix}_{n \times (p+q)} \quad (27)$$

where each sample can be expressed as:

$$z_{(\beta)} = (x_{(\beta)}^T, y_{(\beta)}^T)^T (\beta = 1, 2, \dots, n) \quad (28)$$

CCA seeks the optimal \mathbf{W}_{xopt} and \mathbf{W}_{yopt} that meet the following criteria:

$$\begin{cases} (\mathbf{W}_{xopt}, \mathbf{W}_{yopt}) = \arg \max_{\mathbf{W}_x, \mathbf{W}_y} (\mathbf{W}_x^T \mathbf{S}_{12} \mathbf{W}_y) \\ \text{s.t. } \mathbf{W}_{xopt}^T \mathbf{S}_{11} \mathbf{W}_{xopt} = \mathbf{W}_{yopt}^T \mathbf{S}_{22} \mathbf{W}_{yopt} = \mathbf{I} \end{cases} \quad (29)$$

where

$$\begin{cases} \mathbf{S}_{11} = \frac{1}{n-1} \sum_{\beta=1}^n (x_{(\beta)} - \bar{x})(x_{(\beta)} - \bar{x})^T \\ \mathbf{S}_{22} = \frac{1}{n-1} \sum_{\beta=1}^n (y_{(\beta)} - \bar{y})(y_{(\beta)} - \bar{y})^T \\ \mathbf{S}_{12} = \frac{1}{n-1} \sum_{\beta=1}^n (x_{(\beta)} - \bar{x})(y_{(\beta)} - \bar{y})^T \\ \bar{x} = \sum_{\beta=1}^n x_{(\beta)} / n \text{ and } \bar{y} = \sum_{\beta=1}^n y_{(\beta)} / n \end{cases} \quad (30)$$

Problem (29) can be transformed into solving the orthonormal eigenvectors of the following two matrices \mathbf{M}_1 and \mathbf{M}_2 :

$$\begin{cases} \mathbf{M}_1 = \mathbf{S}_{11}^{-1} \mathbf{S}_{12} \mathbf{S}_{22}^{-1} \mathbf{S}_{12}^T \\ \mathbf{M}_2 = \mathbf{S}_{22}^{-1} \mathbf{S}_{12}^T \mathbf{S}_{11}^{-1} \mathbf{S}_{12} \end{cases} \quad (31)$$

$\mathbf{W}_{\text{xopt}} = [w_{x1}, w_{x2}, \dots, w_{xe}]$ and $\mathbf{W}_{\text{yopt}} = [w_{y1}, w_{y2}, \dots, w_{ye}]$ consist of the orthonormal eigenvectors corresponding to the first e largest eigenvalues of \mathbf{M}_1 and \mathbf{M}_2 , respectively.

KCCA is a non-linear extension of CCA, which seeks canonical variable pairs by maximizing the following correlation coefficients:

$$\begin{aligned} \rho_{\text{KCCA}} &= \frac{\tilde{\mathbf{W}}_x^T \mathbf{K}_x \mathbf{K}_y \tilde{\mathbf{W}}_y}{\sqrt{(\tilde{\mathbf{W}}_x^T \mathbf{K}_x \mathbf{K}_x \tilde{\mathbf{W}}_x + \chi \tilde{\mathbf{W}}_x^T \mathbf{K}_x \tilde{\mathbf{W}}_x)(\tilde{\mathbf{W}}_y^T \mathbf{K}_y \mathbf{K}_y \tilde{\mathbf{W}}_y + \chi \tilde{\mathbf{W}}_y^T \mathbf{K}_y \tilde{\mathbf{W}}_y)}} \end{aligned} \quad (32)$$

where \mathbf{K}_x and \mathbf{K}_y are the centralized kernel matrices. The centralized Gaussian kernel matrices are employed in this study. They are obtained by centering the matrices obtained according to equation (24), and χ is a regularization parameter. For other unfinished details about KCCA, please refer to [28].

3. PROPOSED METHOD

3.1. Modified intra-class correlation analysis (MICA)

In the paper [32], to take advantage of intra-class correlation, a class matrix was introduced into ICA. Multiplying the data matrix and the class matrix together is equivalent to summing the intra-class samples. In this paper, a new class matrix $\mathbf{L}_c \in R^{n \times k}$ is introduced (n and k are the sample size and the number of the classes, respectively). Each sample corresponds to each row of \mathbf{L}_c . For instance, if there are three classes (marked as c_1, c_2 , and c_3) and the number of intra-class samples is 10, 20, and 30, respectively, $[1/10, 0, 0]$, $[0, 1/20, 0]$, $[0, 0, 1/30]$ denote c_1, c_2 , and c_3 , respectively. Given two data matrices $\mathbf{X} \in R^{n \times p}$ and $\mathbf{Y} \in R^{n \times q}$ (where q and p are the respective numbers of features), the covariance matrix \mathbf{S}_L between \mathbf{X} and \mathbf{Y} with class correlation is defined as:

$$\mathbf{S}_L = \frac{1}{k-1} [(\mathbf{H}\mathbf{X})^T \mathbf{L}_c][(\mathbf{H}\mathbf{Y})^T \mathbf{L}_c]^T \quad (33)$$

where $\mathbf{H}\mathbf{X}$ and $\mathbf{H}\mathbf{Y}$ mean centering \mathbf{X} and \mathbf{Y} , respectively. An operation that calculates the mean values of the within-class samples is implicit in equation (33), which is more effective than the summing operation in ICA in terms of inheriting class structure. It is common to evaluate the relationship between classes with the mean values of the within-class samples, such as equation (4).

MICA seeks the optimal $\hat{\mathbf{W}}_{\text{xopt}}$ and $\hat{\mathbf{W}}_{\text{yopt}}$ that meet the following criterion:

$$\begin{cases} (\hat{\mathbf{W}}_{\text{xopt}}, \hat{\mathbf{W}}_{\text{yopt}}) = \arg \max_{\hat{\mathbf{W}}_x, \hat{\mathbf{W}}_y} (\hat{\mathbf{W}}_x^T \mathbf{S}_L \hat{\mathbf{W}}_y) \\ \text{s.t. } \hat{\mathbf{W}}_{\text{xopt}}^T \mathbf{S}_{11} \hat{\mathbf{W}}_{\text{xopt}} = \hat{\mathbf{W}}_{\text{yopt}}^T \mathbf{S}_{22} \hat{\mathbf{W}}_{\text{yopt}} = \mathbf{I} \end{cases} \quad (34)$$

where the definitions of \mathbf{S}_{11} and \mathbf{S}_{12} are shown in equation (30).

Problem (34) can be transformed into solving the orthonormal eigenvectors of the following two matrices $\hat{\mathbf{M}}_1$ and $\hat{\mathbf{M}}_2$:

$$\begin{cases} \hat{\mathbf{M}}_1 = \mathbf{S}_{11}^{-1} \mathbf{S}_L \mathbf{S}_{22}^{-1} \mathbf{S}_L^T \\ \hat{\mathbf{M}}_2 = \mathbf{S}_{22}^{-1} \mathbf{S}_L^T \mathbf{S}_{11}^{-1} \mathbf{S}_L \end{cases} \quad (35)$$

$\hat{\mathbf{W}}_{\text{xopt}} = [\hat{w}_{x1}, \hat{w}_{x2}, \dots, \hat{w}_{xe}]$ and $\hat{\mathbf{W}}_{\text{yopt}} = [\hat{w}_{y1}, \hat{w}_{y2}, \dots, \hat{w}_{ye}]$ consist of the orthonormal eigenvectors corresponding to the first e largest eigenvalues of $\hat{\mathbf{M}}_1$ and $\hat{\mathbf{M}}_2$, respectively.

3.2. Feature selection based on KLFDA and KCCA (mRMR-KCLFDA)

3.2.1. Motivation

As a popular feature selection algorithm, mRMR follows the criterion derived from the following parent criterion:

$$\max D(S, c), D = I(\{x_i, i = 1, \dots, m\}; c) \quad (36)$$

where $S = \{x_i, i = 1, \dots, m\}$ denotes a feature set S with m features, I means calculating mutual information (MI), and c denotes the class variable. The purpose of mRMR is to find D that shares the largest mutual information with c . This scheme is also called Max-Dependency. To get the value of D , we need to estimate the multivariate density $p(x_1, \dots, x_m, c)$ and $p(x_1, \dots, x_m)$, which is often hard to achieve. Consequently, a criterion which is equivalent to the Max-Dependency, mRMR, is proposed, and it has the following form:

$$\max_{x_j \in X - S_{m-1}} [I(x_j; c) - \frac{1}{m-1} \sum_{x_i \in S_{m-1}} I(x_j; x_i)] \quad (37)$$

where S_{m-1} is the existing feature subset, which consists of $m-1$ features.

MI uses an approximate strategy, i.e., the average or the sum, to evaluate the relationship between multidimensional variables and a variable [34]. As a result, it lacks consideration of complementarity between features [35]. In some papers, the measurements of CCA or KCCA were introduced into feature selection. Their roles are to measure the correlation between variables, so as to not only avoid estimating multivariate density but also take the complementarity between features into account. However, CCA contains an assumption that the data follows a multivariate Gaussian distribution, which makes it not suitable to evaluate the correlation between features and class variables when it faces a classification problem. In addition, we also consider that KLFDA and SLFDA have some similar operations, so do KCCA and MICA. Compared with other feature selection algorithms, the feature selection algorithm with similar operations is more suitable for the data after the fusion of SLFDA and MICA. Therefore, we propose a feature selection algorithm based on KLFDA and KCCA as shown in the following subsection.

3.2.2. The feature selection algorithm proposed in this paper

Inspired by equations (5)-(7), we can find that $\Delta(\mathbf{u}) = [\Delta(\mathbf{u})_1, \Delta(\mathbf{u})_2, \dots, \Delta(\mathbf{u})_e]$ reflects the discriminative ability of each feature in the reduced feature set XU . We believe that this discriminative ability is a manifestation of the correlation between a feature and the class variable. So similarly, given a feature f_k and a class variable c , equation (23) can be used to solve the correlation between them as follows:

$$\rho_{KLFDA}(f_k; c) = \text{sum}(\lambda_1, \lambda_2, \dots, \lambda_n) \quad (38)$$

where $\text{sum}(\cdot)$ is the summation over the elements in the parentheses. $\lambda_1, \lambda_2, \dots, \lambda_n$ are the n eigenvalues obtained according to equation (23). In fact, we only use one feature, $f_k \in R^{n \times 1}$, to produce the kernel matrix in equation (23). Although equation (38) does not have the form of Pearson's correlation, it does quantify the correlation between a feature and the class variable. Considering equation (32) and equation (38), mRMR-KCLFDA has the following criterion:

$$J(f_k) = \max_{f_k \in F-S} \left[\frac{\rho_{KLFDA}(f_k; c)}{\rho_{KCCA}(f_k; S)} \right] \quad (39)$$

where F denotes the original feature set, and S denotes the already selected feature set. The detailed process of the algorithm is:

Step 1: Calculate the correlation between each feature and the class variable according to equation (38).

Step 2: Select the feature with the greatest value in Step 1 as the first selected feature f_a .

Step 3: Move the selected feature from F to S .

Step 4: Select the next feature according to equation (39).

Step 5: If the number of features to be selected is greater than the size of S , go to Step 3.

3.3. Procedure of the three-stage feature fusion

Before implementing feature fusion, we first acquire the vibration signals of the hydraulic pump in different degradation states and divide them into some samples. As important sources of degradation state information, the statistical features of time domain and frequency domain are widely used in equipment fault diagnosis and state monitoring due to their simple calculation and high efficiency. In this paper, 11 time-domain features and 12 frequency-domain features are extracted from the vibration signals as shown in Table 1. and Table 2., and then they are put in two data sets for fusion. Given two data sets ($X \in R^{n \times p}, Y \in R^{n \times q}$, where p and q denote the number of features) and sample labels, the fusion methodology consists of the following three stages:

Stage 1: Utilize SLFDA to perform intra-set fusion on the two data sets X and Y , respectively. The two new data sets obtained have the following forms:

$$\begin{cases} \hat{X}_{(n \times d1)} = X_{(n \times p)} \bar{U}_{opt1(p \times d1)} \\ \hat{Y}_{(n \times d2)} = Y_{(n \times q)} \bar{U}_{opt2(q \times d2)} \end{cases} \quad (40)$$

where \bar{U}_{opt1} and \bar{U}_{opt2} are the projection matrices calculated according to equation (20). SLFDA has a good fusion effect on the data that does not follow the MND. \hat{X} and \hat{Y} have more significant MND characteristics than X and Y , which will facilitate the next stage of fusion.

Stage 2: Utilize MICA to fuse \hat{X} and \hat{Y} , and the new data set has the following form:

$$Z_{(n \times 2 \times e)} = [\hat{X}_{(n \times d1)} \hat{W}_{xopt(d1 \times e)}, \hat{Y}_{(n \times d2)} \hat{W}_{yopt(d2 \times e)}] \quad (41)$$

where \hat{W}_{xopt} and \hat{W}_{yopt} are the projection matrices calculated according to equation (34). This stage inherits the class structures of X and Y , which contain discriminative information. MICA also explores the correlation across different features within the same class.

Stage 3: Utilize mRMR-KCLFDA to select the $d3$ desired features from Z and concatenate them. Finally, X and Y are transformed into one data set through the above operations, which is more concise and efficient in classification.

4. EXPERIMENTAL VALIDATION

4.1. Experimental data acquisition

To verify the effectiveness of the proposed method, a hydraulic pump test platform shown in Fig.1. was set up for data acquisition, which consists of a cooling system, a control system, a signal monitoring, acquisition and display system, a pressure regulating system, and a drive system.

The hydraulic pump for this study is an axial piston pump with the following parameters: type: L10VS028DFR, displacement at the rated working condition: 28 ml/r, rated pressure: 22 MPa, and rated rotation speed: 1480 r/min. Two acceleration sensors are installed in two mutually orthogonal directions, respectively, as shown in Fig.2. They acquire vibration signals at a sampling frequency of 50 KHz, each sampling lasts for 1 s, and the interval between two samplings is 30 s.

The single loose boot is studied in this paper because the loose boot is one of the common fault patterns of hydraulic pump. The gap between the plunger and the boot will increase when the loose boot occurs. To acquire vibration signals close to the actual situation, the normal plungers are replaced with failed plungers obtained after equipment maintenance. Five different degrees of the loose boot, as shown in Fig.3., are considered. Vernier calliper is used to measure the maximum radial distance between the boot and the plunger under five different degrees. The five measurements considered as the loose degree are 0.12 mm, 0.18 mm, 0.3 mm, 0.42 mm, and 0.64 mm, respectively. Therefore, a total of five different degradation states are considered. For each sensor, 100 groups of vibration data in each degradation state are collected, and each group consists of 4095 data points. Some examples of vibration data are shown in Fig.4. Three experimental cases based on the vibration signals of these two channels are presented in the following 3 subsections.

Table 1. Time-domain features.

Feature mark	Feature name	Calculation formula	Feature mark	Feature name	Calculation formula
f_1	Mean	$\frac{1}{N} \sum_{n=1}^N x_n$	f_2	Square root amplitude	$\left(\frac{1}{N} \sum_{n=1}^N \sqrt{ x_n } \right)^2$
f_3	Variance	$\frac{1}{N-1} \sum_{n=1}^N (x_n - f_1)^2$	f_4	Root mean square	$\sqrt{\frac{1}{N} \sum_{n=1}^N x_n^2}$
f_5	Peak to peak	$\max(x_n) - \min(x_n)$	f_6	Skewness	$\frac{1}{N} \sum_{n=1}^N \left(\frac{x_n - f_1}{\sqrt{f_3}} \right)^3$
f_7	Kurtosis	$\frac{\sum_{n=1}^N (x_n - f_1)^4}{(N-1) f_3^2}$	f_8	Crest factor	$\frac{\max x_n }{f_4}$
f_9	Clearance factor	$\frac{\max x_n }{f_2}$	f_{10}	Form factor	$\frac{N f_4}{\sum_{n=1}^N x_n }$
f_{11}	Impulse factor	$\frac{N \max x_n }{\sum_{n=1}^N x_n }$			

where $x_n (n = 1, 2, \dots, N)$ is a signal series, N is its length.

Table 2. Frequency-domain features.

Feature mark	Calculation formula	Feature mark	Calculation formula	Feature mark	Calculation formula
f_{12}	$\frac{1}{N} \sum_{n=1}^N s_n$	f_{13}	$\frac{1}{N-1} \sum_{n=1}^N (s_n - f_{12})^2$	f_{14}	$\frac{\sum_{n=1}^N (s_n - f_{12})^3}{N(\sqrt{f_{13}})^3}$
f_{15}	$\frac{\sum_{n=1}^N (s_n - f_{12})^4}{N f_{13}^2}$	f_{16}	$\frac{\sum_{n=1}^N s_n v_n}{\sum_{n=1}^N s_n}$	f_{17}	$\sqrt{\frac{\sum_{n=1}^N (v_n - f_{16})^2 s_n}{N}}$
f_{18}	$\sqrt{\frac{\sum_{n=1}^N (v_n^2 s_n)}{\sum_{n=1}^N s_n}}$	f_{19}	$\sqrt{\frac{\sum_{n=1}^N (v_n^4 s_n)}{\sum_{n=1}^N (v_n^2 s_n)}}$	f_{20}	$\frac{\sum_{n=1}^N (v_n^2 s_n)}{\sqrt{\sum_{n=1}^N s_n \sum_{n=1}^N (v_n^4 s_n)}}$
f_{21}	$\frac{f_{17}}{f_{16}}$	f_{22}	$\frac{\sum_{n=1}^N [(v_n - f_{16})^3 s_n]}{N f_{17}^3}$	f_{23}	$\frac{\sum_{n=1}^N [(v_n - f_{16})^4 s_n]}{N f_{17}^4}$

where $s_n (n = 1, 2, \dots, N)$ is a spectrum, N is the number of spectrum lines; v_n is the frequency value of the n -th spectrum line.



Fig.1. Hydraulic pump test platform.

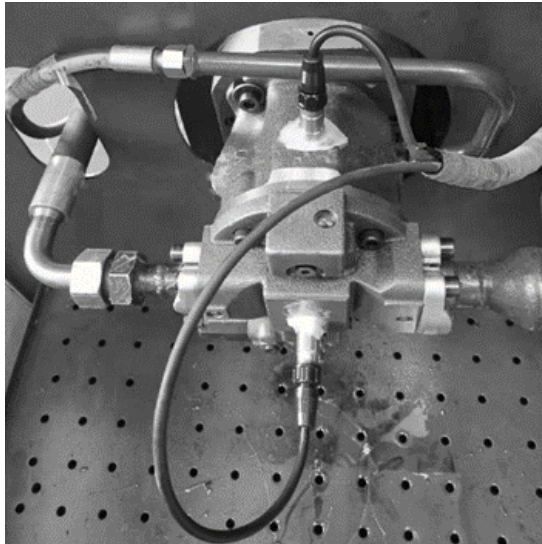


Fig.2. Layout of the two vibration sensors.

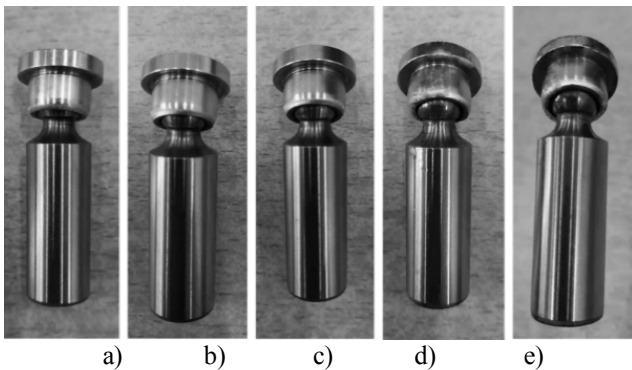


Fig.3. Five different degrees of loose boot: a) 0.12 mm; b) 0.18 mm; c) 0.3 mm; d) 0.42 mm; e) 0.64 mm.

4.2. Case 1: Validation under the condition of large sample size

As described in subsection 3.3, two original data sets $X \in R^{500 \times 23}$ and $Y \in R^{500 \times 23}$ are extracted from all the vibration data, where X and Y are derived from two different sensors (marked as channel 1 and channel 2), and each degradation state has 100 samples in each set. The 23 features in each set are those shown in Table 1. and Table 2. The method described in subsection 3.3 is used to fuse X and Y . The size of the fused feature sets at each stage are set as: $d1 = d2 = 10$, $2e = 8$, $d3 = 7$, which are determined after several simple attempts.

As shown in Fig.5., the fusion result of each stage is visualized using t-SNE [36], where the fused feature sets of Stage 1 are merged into a 20-dimensional set for visualization (Fig.5.b) to Fig.5.d)). The fusion effect of each stage is obvious. The original feature sets X and Y are also merged into a 46-dimensional set for visualization (Fig.5.a)). It can be seen that the scattered points of different colors are mixed. For a comparison, CCA is used to fuse \hat{X} and \hat{Y} , which are the two feature sets obtained from Stage 1, and the fusion result is also visualized (Fig.5.e)). It can be seen that its fusion effect is worse than that of MICA. For a comparison between mRMR and mRMR-KCLFDA, mRMR is used to select the desired features from Z , which is the feature set obtained from Stage 2. The results of these two feature selection algorithms form a data set, and the samples of each set are divided into a training set and a testing set at a ratio of 1:1. Support vector machine (SVM) [37] with Gaussian kernel is used for classification. The classification accuracy with different number of features is shown in Fig.5.f). It can be seen that mRMR-KCLFDA selected better feature combinations in most cases and found the eighth feature as a redundant one. Fig.6. is used to illustrate the significance comparison of the MND of the samples before and after the SLFDA fusion. One hundred samples from channel 1 and those from channel 2 are taken as examples, all of which belong to the loose degree 0.12 mm. The comparison results are displayed by the Q-Q diagram. It can be seen that most of the samples before fusion are farther from the line $y = x$, and most of the samples after fusion are very close to the line, which confirms that the fused samples have very significant MND characteristics.

For deeper comparisons, five existing algorithms are also used to fuse features: LLE, LPP, CCA, locality preserving CCA (LPCCA) [38], and discriminant CCA (DisCCA) [39]. Each method results in an 8-dimensional feature set, respectively. The feature selection algorithm mRMR is also used to select 7 features from each feature set fused by these methods. SVM, random forest (RF) [40] and Naïve Bayes (NB) [41] are chosen as the classifiers. The classification accuracies shown in Table 3. reinforce the following 3 conclusions: first, LLE does not perform well. Second, DisCCA performs well due to the use of class information. Third, the proposed method shows the best performance.

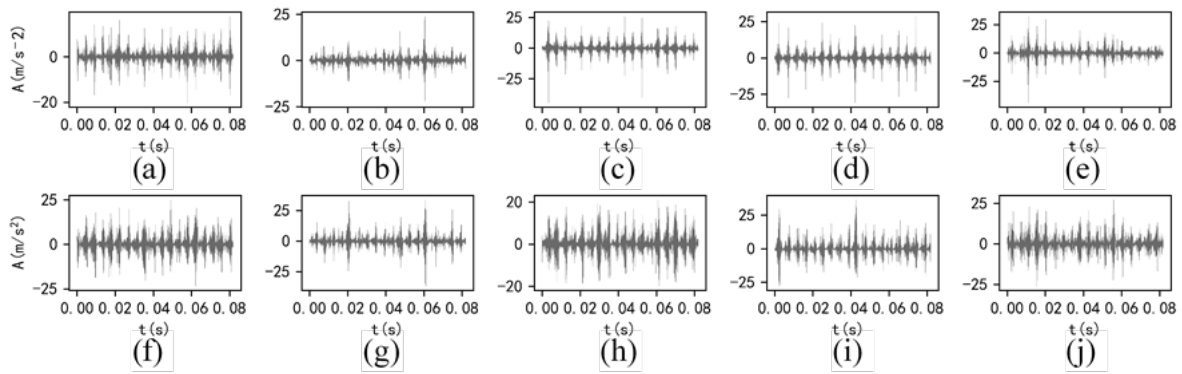


Fig.4. Time-domain waveforms of several vibration signal samples: a) - e) originate from sensor 1, and f) - j) originate from sensor 2. Each column corresponds to the same degree of loose boot. Five different degrees of loose boot are 0.12 mm, 0.18 mm, 0.3 mm, 0.42 mm, and 0.64 mm from left to right.

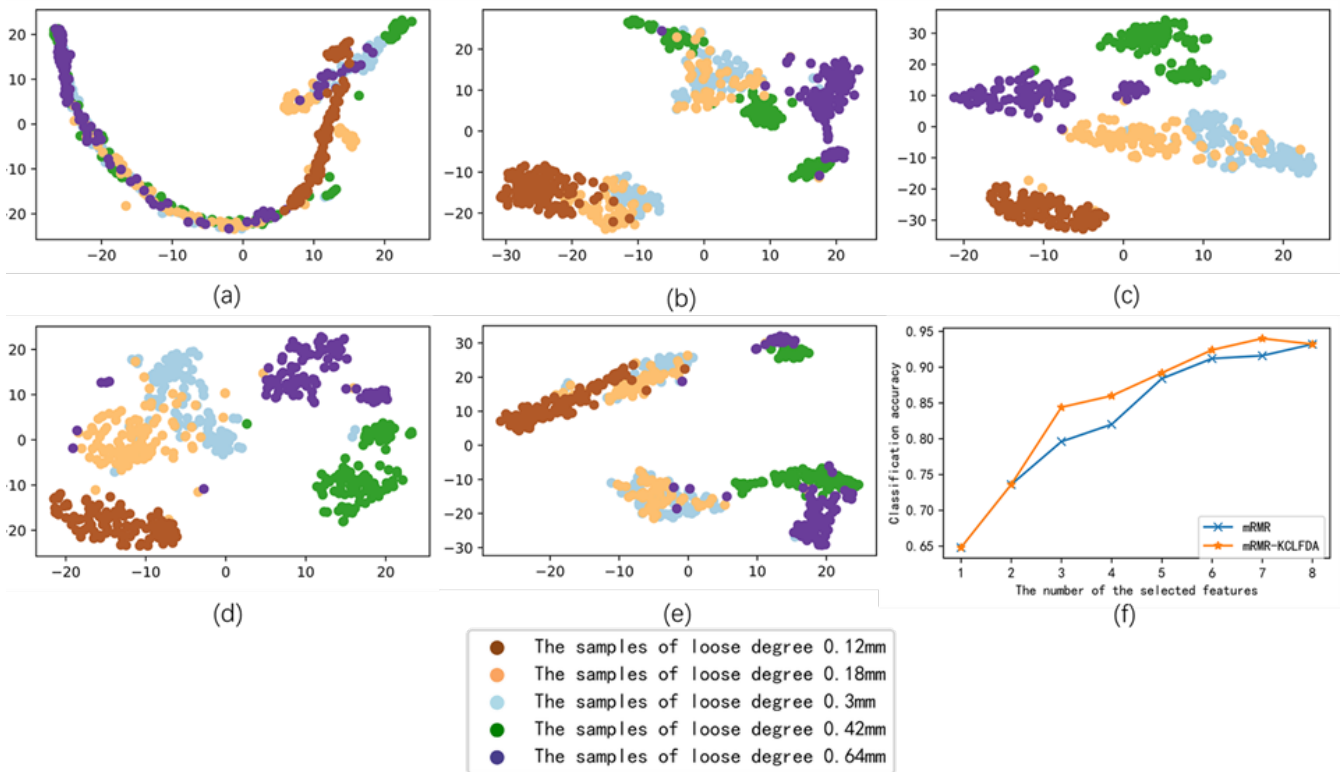


Fig.5. Some illustrations of the fusion effects in Subsection 4.2: a) t-SNE on the original features. b) - d) t-SNE on the fused features of Stage 1, Stage 2 and Stage 3, respectively. e) t-SNE on the fused features through CCA. (f) Comparison of classification accuracy between mRMR and mRMR-KCLFDA at different number of features.

Table 3. Classification accuracies for feature fusion in case 1.

	method	LLE+ mRMR	LPP+ mRMR	CCA+ mRM R	LPCCA + mRMR	DisCCA+ mRMR	The proposed method
Classification accuracy	SVM	66 %	88 %	88 %	89.8 %	91.2 %	93.2 %
	RF	55.75 %	77.74 %	78.2 %	79.78 %	86.6 %	89.38 %
	NB	58.4 %	65.6 %	70 %	75.69 %	88 %	90 %

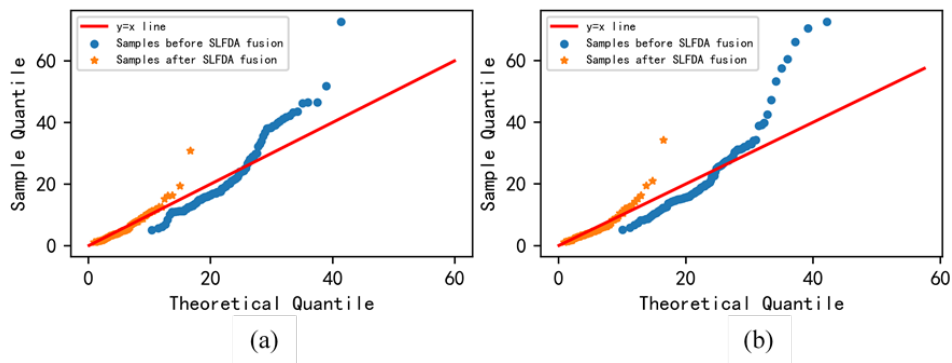


Fig.6. Significance comparisons of MND displayed by Q-Q diagram: a) channel 1; b) channel 2.

4.3. Case 2: Validation under the condition of small sample size

Two original feature sets $X \in R^{175 \times 23}$ and $Y \in R^{175 \times 23}$ are extracted from part of the vibration data, where X and Y are derived from two different channels, and each degradation state has 35 samples in each set. The method described in subsection 3.3 is used to fuse X and Y . The size of the fused feature sets at each stage are set as: $d1 = d2 = 10, 2e = 8, d3 = 7$. Similar to Subsection 4.2, all the results are visualized as shown in Fig.7. For a comparison, DisCCA is used to fuse \hat{X} and \hat{Y} , which are the two feature sets obtained from Stage 1, and the fusion result is also visualized (Fig.7.e). It can be seen that its fusion effect is worse than that of MICA. The classification accuracy obtained by the proposed method is 100 % (Fig.7.f).

For deeper comparisons, the 5 existing algorithms used in Subsection 4.2 are also used to fuse the features. Both PCA and SLFDA are used in advance to reduce the dimension of each original feature set to 10 (i.e., $d1 = d2 = 10$). The size of the fused feature sets in other cases is not specified. The values shown in Table 4. are the highest classification accuracy and the size of the corresponding feature set, where the classification accuracy is obtained by SVM. The symbol “~” in Table 4. denotes the five algorithms used. The following three conclusions can be drawn from this table and Fig.7.: first, using PCA in advance to reduce dimensionality is always unsuccessful on these original feature sets, while using SLFDA can get better results. Second, our proposed method outperforms these five existing methods in classification accuracy (100 %). Third, the fused feature set obtained by our proposed method has a lower dimension in most cases.

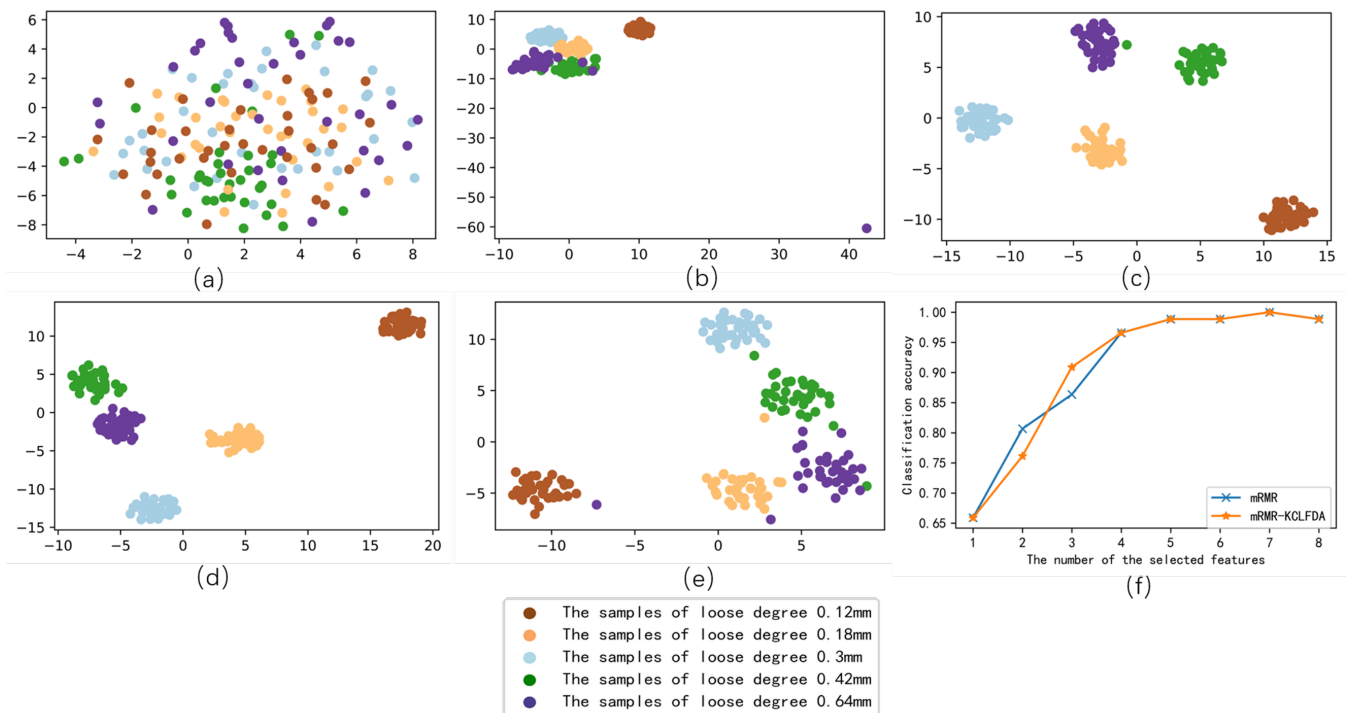


Fig.7. Some illustrations of the fusion effects in Subsection 4.3: a) t-SNE on the original features. b) - d) t-SNE on the fused features of Stage 1, Stage 2 and Stage 3, respectively. e) t-SNE on the fused features through DisCCA. f) Comparison of classification accuracy between mRMR and mRMR-KCLFDA at different number of features.

Table 4. Classification accuracies and the number of corresponding features for feature fusion in case 2 (except the proposed method).

method	LLE	LPP	CCA	LPCCA	DisCCA
PCA+~	52.43 % (15)	53.68 % (18)	57.5 % (14)	57.95 % (11)	56.35 % (10)
SLFDA+~	98.86 % (14)	97.73 % (12)	97.73 % (11)	97.73 % (13)	98.86 % (13)
SLFDA+~+mRMR	98.86 % (13)	98.86 % (12)	97.73 % (10)	98.86 % (11)	98.86 % (8)

Table 5. Comparison of some previous studies about the rotating machinery fault with the proposed method.

Authors	Feature fusion method	Model in the literature	Dataset source	Accuracy (%)
Ugochukwu Ejike Akpudo et al. [42]	LLE	Signal + MFCC + LLE + SVM	Solenoid pump testbed of authors. Our experimental hydraulic pump data.	100 60.85
Jiang Lingli et al. [43]	LPP	Signal + MPE + LPP + ELM	Gearbox test equipment of authors. Our experimental hydraulic pump data.	100 81.5
Yazhou Li et al. [44]	A method based on weighted feature fusion proposed by the authors	Signal + Original feature set + Feature fusion + SVM	CWRU bearing fault database. The experimental bearing data of Xi'an Jiaotong University. Our experimental hydraulic pump data.	99.61 99.38 89.4
Xiao Yu et al. [45]	FSASR+ SM-LFDA	Signal + EMD + Original feature set + Feature fusion + SVM	CWRU bearing fault database. Our experimental hydraulic pump data.	100 90.75
Xiaoli Zhao et al. [46]	GLMFA	Signal + multi-domain feature set + GLMFA + Improved EW-KNN	CWRU bearing fault database. The experimental bearing data of the HBERC. Our experimental hydraulic pump data.	98.7 100 92.3
Hongru Li et al. [47]	A relative entropy fusion method proposed by the authors	Signal + Several complexity features + Feature fusion + SVM	The experimental hydraulic pump data of authors. Our experimental hydraulic pump data.	93.3 85.74
Authors of this paper	A three-stage feature fusion methodology proposed by the authors	Signal + Two multi-domain feature set + Feature fusion + SVM	Experimental setup of authors.	93.2

where some acronyms and their meanings are given as follows:

MFCC	Mel Frequency Cepstral Coefficient	MPE	Multiscale Permutation Entropy
ELM	Extreme Learning Machine	CWRU	Case Western Reserve University
FSASR	Features Selection by Adjusted rand index and Standard deviation Ratio	SM-LFDA	Support Margin Local Fisher Discriminant Analysis
EMD	Empirical Mode Decomposition	GLMFA	Global-Local Margin Fisher Analysis
EW-KNN	Euclidean Weighted K-Nearest Neighbor	HBERC	Hangzhou Bearing Experimental Research Center

4.4. Case 3: Comparison of some previous studies with the proposed method

Some previous studies related to the fault diagnosis and detection of rotating machinery, as well as our study, are listed in Table 5. The rightmost column of the table shows the highest classification accuracy corresponding to each dataset. Considering that feature fusion is the focus of our study, we replace the feature fusion method in our model with that in each previous study and test each of the new models on our hydraulic pump data, which is convenient for highlighting the comparison

among different feature fusion methods. Our experimental hydraulic pump data is the same as the data used in Case 1, and all the accuracy values corresponding to it are obtained by SVM.

The accuracy value obtained by the proposed method is the greatest among all the accuracy values corresponding to our hydraulic pump data (93.2 %). All the models in the table include their original feature extraction methods, such as the MFCC in the first model, and the feature fusion method in each model needs to adapt to the characteristics of the original features. The two methods, FSASR+SM-LFDA and GLMFA,

also achieve high classification accuracy (90.75 % and 92.3 %). Not only are the original features fused by them, but also their fusion ideas are somewhat similar to our methods. Thus, the superiorities of the proposed feature fusion methodology are reinforced.

5. SUMMARY

To effectively enhance the performance of the degradation features of hydraulic pumps, a novel three-stage feature fusion is proposed. The two original feature sets are extracted from the vibration signals. SLFDA is used to perform intra-set fusion on the two original feature sets, then MICA, a CCA-based method, is used to fuse two feature sets, and finally the mRMR-KCLFDA is used to select the desired features. This paper has three major contributions as follows:

1) MICA is proposed to fuse two feature sets in the second stage, which is developed from ICA. MICA modifies the calculation method of the class matrix, the mean values of the within-class samples are used to evaluate the relationship between classes, which makes it more convenient to inherit the discriminating structure.

2) A feature selection algorithm mRMR-KCLFDA based on KLFDA and KCCA is proposed, which has a framework similar to mRMR. In this algorithm, KLFDA quantifies the correlation between each feature candidate and class variable, and KCCA quantifies the correlation between each feature candidate and the selected feature set. The algorithm avoids not only the estimation of complex multivariate density, but also the inaccuracy of CCA in estimating the correlation between feature and class variable. Additionally, since it includes some operations similar to the methods used in the first two stages, it is more suitable for our fusion framework than other feature selection algorithms.

3) SLFDA, MICA and mRMR-KCLFDA are closely linked with one another in an orderly manner, which results in the proposed three-stage feature fusion methodology. Non-linear and non-Gaussian hydraulic pump data, as well as the manifold structure and intra-class multimodal sample caused by it can be handled well with this methodology, whose effectiveness and superiority in fusing the degradation features of the hydraulic pump were demonstrated with experimental data.

Still, we have not studied how to conduct fusion on more than 2 feature sets. The inability to be easily extended to the multiple feature sets is a limitation of the proposed method, which is mainly due to the inability of MICA to handle more than 2 feature sets. In our future work, we will try addressing the limitation from two perspectives. One is to replace MICA with a method that can handle multiple feature sets, and the other is to fuse multiple data sets in advance at the signal level.

ACKNOWLEDGMENTS

This project is supported by National Natural Science Foundation of China (Grant No. 51275524).

REFERENCES

- [1] Du, J., Wang, S., Zhang, H. (2013). Layered clustering multi-fault diagnosis for hydraulic piston pump. *Mechanical Systems and Signal Processing*, 36 (2), 487-504.
- [2] Yu, H., Li, H., Li, Y. (2020). Vibration signal fusion using improved empirical wavelet transform and variance contribution rate for weak fault detection of hydraulic pumps. *ISA Transactions*, 107, 385-401.
- [3] Sun, J., Li, H., Xu, B. (2016). The morphological undecimated wavelet decomposition–discrete cosine transform composite spectrum fusion algorithm and its application on hydraulic pumps. *Measurement*, 94, 794-805.
- [4] Zhong, K., Han, M., Han, B. (2019). Data-driven based fault prognosis for industrial systems: A concise overview. *IEEE/CAA Journal of Automatica Sinica*, 7 (2), 330-345.
- [5] Xu, Y., Tian, W., Zhang, K., Ma, C. (2019). Application of an enhanced fast kurtogram based on empirical wavelet transform for bearing fault diagnosis. *Measurement Science and Technology*, 30 (3), 035001.
- [6] Li, H., Tian, Z., Yu, H., Xu, B. (2019). Fault prognosis of hydraulic pump based on bispectrum entropy and deep belief network. *Measurement Science Review*, 19 (5), 195-203.
- [7] Kaya, Y., Kuncan, M., Kaplan, K., Minaz, M.R., Ertunc, H.M. (2020). Classification of bearing vibration speeds under 1D-LBP based on eight local directional filters. *Soft Computing*, 24 (16), 12175-12186.
- [8] Kaya, Y., Kuncan, M., Kaplan, K., Minaz, M.R., Ertunc, H.M. (2021). A new feature extraction approach based on one dimensional gray level co-occurrence matrices for bearing fault classification. *Journal of Experimental & Theoretical Artificial Intelligence*, 33 (1), 161-178.
- [9] Bayram, S., Kaplan, K., Kuncan, M., Ertunc, H.M. (2014). The effect of bearings faults to coefficients obtained by using wavelet transform. In *2014 22nd Signal Processing and Communications Applications Conference (SIU)*. IEEE, 991-994. ISBN 978-1-4799-4874-1.
- [10] Kuncan, M. (2020). An intelligent approach for bearing fault diagnosis: Combination of 1D-LBP and GRA. *IEEE Access*, 8, 137517-137529.
- [11] Kaplan, K., Bayram, S., Kuncan, M., Ertunc, H.M. (2014). Feature extraction of ball bearings in time-space and estimation of fault size with method of ANN. In *Proceedings of the 16th Mechatronika 2014*, 295-300.
- [12] Hu, Q., Si, X.-S., Qin, A.-S., Lv, Y.-R., Zhang, Q.-H. (2020). Machinery fault diagnosis scheme using redefined dimensionless indicators and mRMR feature selection. *IEEE Access*, 8, 40313-40326.
- [13] Meng, T., Jing, X., Yan, Z., Pedrycz, W. (2020). A survey on machine learning for data fusion. *Information Fusion*, 57, 115-129.
- [14] Snoek, C.G., Worring, M., Smeulders, A.W. (2005). Early versus late fusion in semantic video analysis. In *Proceedings of the 13th Annual ACM International Conference on Multimedia*, 399-402.
- [15] Liu, Y., He, B., Liu, F., Lu, S., Zhao, Y. (2016). Feature fusion using kernel joint approximate diagonalization of eigen-matrices for rolling bearing fault identification. *Journal of Sound and Vibration*, 385, 389-401.
- [16] Cai, H., Qu, Z., Li, Z., Zhang, Y., Hu, X., Hu, B. (2020). Feature-level fusion approaches based on multimodal EEG data for depression recognition. *Information Fusion*, 59, 127-138.

- [17] Tian, Y., Wang, Z., Zhang, L., Lu, C., Ma, J. (2018). A subspace learning-based feature fusion and open-set fault diagnosis approach for machinery components. *Advanced Engineering Informatics*, 36, 194-206.
- [18] Roweis, S.T., Saul, L.K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290 (5500), 2323-2326.
- [19] He, X., Niyogi, P. (2004). Locality preserving projections. *Advances in Neural Information Processing Systems*, 16 (16), 153-160.
- [20] Sugiyama, M. (2007). Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *Journal of Machine Learning Research*, 8 (5), 1027-1061.
- [21] Wang, Z., Ruan, Q., An, G. (2016). Facial expression recognition using sparse local Fisher discriminant analysis. *Neurocomputing*, 174, 756-766.
- [22] Zhu, Q., Liu, Q., Qin, S.J. (2017). Concurrent quality and process monitoring with canonical correlation analysis. *Journal of Process Control*, 60, 95-103.
- [23] Zhuang, X., Yang, Z., Cordes, D. (2020). A technical review of canonical correlation analysis for neuroscience applications. *Human Brain Mapping*, 41 (13), 3807-3833.
- [24] Izenman, A.J. (2013). Linear discriminant analysis. In *Modern Multivariate Statistical Techniques*. Springer, 237-280.
- [25] Wold, S., Esbensen, K., Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2 (1-3), 37-52.
- [26] Ke, X., Yuan, F., Cheng, E. (2020). Integrated optimization of underwater acoustic ship-radiated noise recognition based on two-dimensional feature fusion. *Applied Acoustics*, 159, 107057.
- [27] Chen, Z., Ding, S.X., Peng, T., Yang, C., Gui, W. (2017). Fault detection for non-Gaussian processes using generalized canonical correlation analysis and randomized algorithms. *IEEE Transactions on Industrial Electronics*, 65 (2), 1559-1567.
- [28] Lai, P.L., Fyfe, C. (2000). Kernel and nonlinear canonical correlation analysis. *International Journal of Neural Systems*, 10 (05), 365-377.
- [29] Yuan, Y., Lu, P., Xiao, Z., Liu, J., Wu, X. (2015). A novel supervised CCA algorithm for multiview data representation and recognition. In *Chinese Conference on Biometric Recognition*. Springer, 702-709. ISBN 978-3-319-25417-3.
- [30] Wu, Z., Mao, K., Ng, G.W. (2019). Enhanced feature fusion through irrelevant redundancy elimination in intra-class and extra-class discriminative correlation analysis. *Neurocomputing*, 335, 105-118.
- [31] Peng, H., Long, F., Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27 (8), 1226-1238.
- [32] Zuobin, W., Kezhi, M., Ng, G.W. (2017). Effective feature fusion for pattern classification based on intra-class and extra-class discriminative correlation analysis. In *2017 20th International Conference on Information Fusion*. IEEE, 1-8. ISBN 978-1-5090-4582-2.
- [33] Yin, W., Osher, S., Goldfarb, D., Darbon, J. (2008). Bregman iterative algorithms for ℓ_1 -minimization with applications to compressed sensing. *SIAM Journal on Imaging Sciences*, 1 (1), 143-168.
- [34] Brown, G., Pocock, A., Zhao, M.J., Lujan, M. (2012). Conditional likelihood maximisation: A unifying framework for information theoretic feature selection. *The Journal of Machine Learning Research*, 13 (1), 27-66.
- [35] Wang, Y., Cang, S., Yu, H. (2019). Mutual information inspired feature selection using kernel canonical correlation analysis. *Expert Systems with Applications: X*, 4, 100014.
- [36] Van der Maaten, L., Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9 (11), 2579-2605.
- [37] Noble, W.S. (2006). What is a support vector machine? *Nature Biotechnology*, 24 (12), 1565-1567.
- [38] Sun, T., Chen, S. (2007). Locality preserving CCA with applications to data visualization and pose estimation. *Image and Vision Computing*, 25 (5), 531-543.
- [39] Guo, C., Wu, D. (2019). Canonical correlation analysis (CCA) based multi-view learning: An overview. *arXiv:1907.01693 [cs.LG]*.
- [40] Breiman, L. (2001). Random forest. *Machine Learning*, 45, 5-32.
- [41] Xu, S. (2018). Bayesian Naïve Bayes classifiers to text classification. *Journal of Information Science*, 44 (1), 48-59.
- [42] Akpudo, U.E., Hur, J.W. (2020). Intelligent solenoid pump fault detection based on MFCC features, LLE and SVM. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*. IEEE, 404-408. ISBN 978-1-7281-4986-8.
- [43] Jiang, L., Tan, H., Li, X., Yang, D. (2021). A novel MPE-LPP-ELM recognition method for the fault diagnosis of spiral bevel gears. *Shock and Vibration*, 2021, 5552048.
- [44] Li, Y., Dai, W., Zhang, W. (2020). Bearing fault feature selection method based on weighted multidimensional feature fusion. *IEEE Access*, 8, 19008-19025.
- [45] Yu, X., Dong, F., Ding, E., Wu, S., Fan, C. (2017). Rolling bearing fault diagnosis using modified LFDA and EMD with sensitive feature selection. *IEEE Access*, 6, 3715-3730.
- [46] Zhao, X., Jia, M. (2018). Fault diagnosis of rolling bearing based on feature reduction with global-local margin Fisher analysis. *Neurocomputing*, 315, 447-464.
- [47] Li, H., Sun, J., Ma, H., Tian, Z., Li, Y. (2019). A novel method based upon modified composite spectrum and relative entropy for degradation feature extraction of hydraulic pump. *Mechanical Systems and Signal Processing*, 114, 399-412.

Received June 30, 2021
Accepted August 31, 2021