MDPI

Article

Euler Representation-Based Structural Balance Discriminant Projection for Machinery Fault Diagnosis

Maoyan Zhang [1], Yanmin Zhu [2,*], Shuzhi Su [1,3], Xianjin Fang [1,3] and Ting Wang [1]

[1] School of Computer Science and Engineering, Anhui University of Science & Technology, Huainan 232001, China
[2] School of Mechanical Engineering, Anhui University of Science & Technology, Huainan 232001, China
[3] Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230031, China
* Correspondence: ymzhu@aust.edu.cn

Abstract: Fault diagnosis methods are usually sensitive to outliers and it is difficult to obtain and balance global and local discriminant information, which leads to poor separation between classes of low-dimensional discriminant features. For this problem, we propose an Euler representation-based structural balance discriminant projection (ESBDP) algorithm for rotating machine fault diagnosis. First, the method maps the high-dimensional fault features into the Euler representation space through the cosine metric to expand the differences between heterogeneous fault samples while reducing the impact on outliers. Then, four objective functions with different structure and class information are constructed in this space. On the basis of fully mining the geometric structure information of fault data, the local intra-class aggregation and global inter-class separability of the low-dimensional discriminative features are further improved. Finally, we provide an adaptive balance strategy for constructing a unified optimization model of ESBDP, which achieves the elastic balance between global and local features in the projection subspace. The diagnosis performance of the ESBDP algorithm is explored by two machinery fault cases of bearing and gearbox. Encouraging experimental results show that the algorithm can capture effective fault discriminative features and can improve the accuracy of fault diagnosis.

Keywords: fault diagnosis; Euler representation; adaptive balance strategy; dimensional reduction

1. Introduction

In modern industrial production, rotating machinery and equipment has been widely used. With the continuous improvement of automation and intelligence, the internal structure of rotating machinery has become more complex. Moreover, such equipment works in harsh environment for a long time, and its core components such as gears and bearings are prone to damage, and failure to diagnose the fault in time may lead to the failure of normal operation of the equipment and even cause major accidents. Thus, the effective fault diagnosis and condition monitoring of key components of rotating machinery is of great significance [1,2]. Usually, the vibration signals collected from machinery equipment contain plenty of important information reflecting its operational status [3,4], and the machinery data and information collected are increasing in order to improve information integrity, which unavoidably leads to problems such as data redundancy and high dimensionality, making the performance of fault diagnosis systems adversely affected. Therefore, extracting valuable discriminative information from high-dimensional fault data and improving the diagnostic accuracy is a pressing challenge [5].

Currently, data mining and machine learning have been proven to be effective methods for data analysis and processing [6–10]. In the field of fault diagnosis based on data mining and machine learning, the process of processing and identifying high-dimensional fault feature datasets usually includes the following three aspects: acquisition and processing of fault data, extraction of fault features, and identification and classification of fault features.

Among them, extracting effective fault features is the crucial of diagnosis. As an effective method, dimensional reduction [11,12] aims to extract low-dimensional discriminative features that reflect intrinsic information from high-dimensional data to achieve better fault diagnosis performance. According to the retained data structure, dimensional reduction algorithms are classified into two main types, i.e., algorithms based on local structure and algorithms based on global structure. The most widely used global algorithms include principal components analysis (PCA) [13] and kernel principal components analysis (KPCA) [14], etc., which are concerned with obtaining global structural information of machinery fault data for diagnosis. Among them, PCA obtains the projection direction via maximizing the covariance among samples. However, PCA is a linear diagnosis algorithm, which is difficult to employ for handling complex nonlinear fault data. For this reason, KPCA introduces kernel tricks [15] based on PCA, which can further obtain the nonlinear structural information of fault data. However, the global algorithm ignores the information of the sub-manifold structure, which has been shown to better reflect the underlying structure of the data.

Manifold learning algorithms can effectively explore the intrinsic structure of fault data and have been widely used in fault diagnosis. Commonly used manifold algorithms include local linear embedding (LLE) [16], Marginal fisher analysis (MFA) [17], neighborhood preserving embedding (NPE) [18], locality preserving projection (LPP) [19], etc. Among them, LPP constructs a nearest neighbor matrix to reflect the local neighborhood relationship of the data and expects to preserve this relationship after projection. Unlike LPP, LLE treats the data located on a smaller local structure as linear and uses local nearest neighbor points to linearly represent any point in the neighborhood. However, both LPP and LLE do not utilize the class information of the samples. MFA belongs to a type of supervised algorithm which expects homogeneous samples in the neighborhood to be closer in the subspace and heterogeneous samples to be further away. In view of the effectiveness of manifold algorithms, scholars have introduced many new manifold algorithms for fault diagnosis in recent years [20–22]. However, conventional dimensional reduction algorithms are unable to combine local and global information of data during fault diagnosis.

The literature [23,24] shows that both local and global information of fault data can provide an active role in fault diagnosis. For fault data, global information describes its overall characteristics, while local information reflects its internal structure. Therefore, in order to obtain more comprehensive fault discrimination information, Luo et al. [25] proposed a nonlocal and local structure preserving projection (NLSPP). It not only uncovers the non-local structure of the data by calculating the variance between samples, but also maintains the neighbor relationship between samples, and the extracted features are more powerful compared to PCA and LPP. However, NLSPP does not utilize supervised information, which lacks advantages in the classification of fault data. Tang et al. [26] offered a Fisher discriminative global local preserving projection (FDGLPP) based on GLPP by combining the ideas of FDA to reach a superior recognition performance. Yang et al. [27] integrated the GLPP and multiple marginal fisher analysis [28] methods to propose a global-local marginal discriminative preserving projection (GLMDPP) method, which can extract both intrinsic and discriminative features of the data. To overcome the drawback of the MFA algorithm being challenged in obtaining both local and global features from bearing fault data, Zhao et al. [29] proposed a global local margin Fisher analysis (GLMFA) algorithm by introducing two regularization terms. In addition, since the orthogonality criterion can reduce the impact of data noise on the classification performance, Su et al. [30] proposed an orthogonal locality and globality preserving projection (OLGPP). Under the orthogonality constraint, the algorithm not only balances the manifold structure and the Euclidean structure, but also reduces the influence generated by noise in the data. Since the orthogonal discriminant projection [31] cannot balance the neighborhood relationship of rotor fault data, Shi et al. [32] proposed a local global balanced orthogonal discriminant projection (LGBODP) that improves the separability of heterogeneous samples by considering the neighborhood and non-neighborhood structure of the samples in an integrated manner.

However, among these existing methods, the Euclidean distance-based similarity matrix is widely used to reflect the relationship between fault samples to obtain relevant features, but the Euclidean distance is very sensitive to outliers in the fault data [33], which can affect the effective extraction of discriminative information and lead to decreased fault classification performance. In addition, in order to take into account both global and local features, these algorithms need to set an additional balancing parameter when constructing the model. For the value of this parameter, the range is usually first set manually and then determined using a grid search method [34]. This approach makes it difficult to effectively balance the relative importance between global and local structures, which largely reduces the discriminative power of the obtained fault features.

To address the above problems, an Euler representation-based structural balance discriminant projection (ESBDP) algorithm is proposed in this paper for rotating machinery fault diagnosis. First, the algorithm represents the high-dimensional fault features by mapping them into the Euler space via the cosine metric [35]. The Euler representation approximates an ideal robust kernel that can suppress the effect of outliers in the fault data to a certain extent. Meanwhile, heterogeneous fault samples have greater differences in this space, which improves inter-class separability to a certain extent and facilitates fault classification. Secondly, ESBDP constructs four objective functions with different structure and class information in Euler space. While fully exploiting the global and local structural information of fault data, it effectively improves the local intra-class aggregation and global inter-class separation of low-dimensional discriminative features. Finally, we offer an adaptive balance strategy to construct a unified optimization model of ESBDP, which achieves adaptive balance between global and local features in the projection subspace and further enhances the discriminative power of fault features while improving the adaptiveness of the algorithm. To verify the effectiveness of our method, we have conducted experimental analysis on two rotating machinery fault cases of bearing and gearbox, and the encouraging experimental results show that our method has superior fault diagnosis performance.

2. Review of Locality Preserving Projection

LPP is one of the classical manifold dimension reduction algorithms based on the nearest neighbor criterion [36], and many existing methods are extension studies of LPP, so we provide a brief review of LPP. Given a fault sample set $X \in [x_1, x_2, \cdots, x_n] \in R^{M \times n}$, LPP aims to detect an optimal projection matrix P that maps the high-dimensional fault data to the low-dimensional space and retains the local adjacency information or structure of the original dataset. Define $y_i \in R^d (d << M)$ as a low-dimensional representation of x_i. Then, y_i should satisfy

$$y_i = P^T x_i (i = 1, 2, \cdots, n). \tag{1}$$

First, LPP describes the local adjacency structure of the data by constructing a nearest neighbor graph $G_{lpp} = (X, W_{lpp})$, where X denotes the set of nodes and W_{lpp} is the weight matrix. Based on the k nearest neighbor criterion and the Euclidean distance, the local nearest neighbor matrix is constructed as

$$W_{lpp}^{ij} = \begin{cases} \exp(-\|x_i - x_j\|_2^2 / 2t^2), & if \ x_j \in N_k(x_i) \\ 0, & otherwise \end{cases}, \tag{2}$$

where t is a kernel parameter, $N_k(x_i)$ denotes the set of k nearest samples of x_i. For two samples that are neighbors of each other in the original space, LPP expects them to still maintain this neighbor relationship in the projection subspace. Therefore, LPP defines the minimization objective function as

$$\min_P \sum_{i,j} \|y_i - y_j\|_2^2 W_{lpp}^{ij}. \tag{3}$$

According to Equation (1), Equation (3) can be further expressed as

$$
\begin{aligned}
\min_{P} \sum_{i,j} \left\| P^T x_i - P^T x_j \right\|_2^2 W_{\text{lpp}}^{ij} & \\
= P^T X D X^T P - P^T X W X^T P, & \\
= P^T X L X^T P &
\end{aligned}
\tag{4}
$$

where D is a diagonal matrix whose diagonal elements correspond to the sum of the row vectors of W_{lpp}^{ij}. Moreover, to avoid trivial solutions, the constraint is set to $P^T X D X^T P = I$. Finally, the required mapping matrix P is obtained by solving the optimal objective model.

3. Proposed Method

Similar to LPP, existing fault diagnosis methods usually use Euclidean distance to calculate the similarity weights among fault samples when performing feature extraction, which makes the methods sensitive to outliers in the data and thus reduces the diagnosis performance. In addition, both global and local information can play different important roles in improving the performance of fault diagnosis; however, the relative importance between them is difficult to be balanced effectively, which is not conducive to obtaining more effective fault features. In order to capture and balance more discriminative global and local fault information while avoiding the outlier sensitivity problem caused by Euclidean distance, we propose the ESBDP algorithm. Its main process is as follows.

First, the fault sample set is converted into Euler representation data. The Euler representation is not only robust to outliers, but also capable of relatively expanding the sample differences between classes. Second, the ESBDP algorithm constructs a unified objective model in the Euler space. The model combines four functions with different structural and class information, which can take into account more global and local geometric similarity information of fault data. Finally, the unified model is further optimized. In the optimized model, weights corresponding to local and global information can be automatically learned through the adaptive balancing strategy, which effectively improves the discriminative power of extracted features. The detailed construction process of this algorithm is described as follows.

3.1. Euler Representation

The Euler representation based on the cosine metric has an important role in the ESBDP algorithm, by which the Euler representation can relatively expand the distance between dissimilar samples, which is beneficial for fault classification. At the same time, the Eulerian representation can approximate the ideal robust kernel, thus suppressing the influence of outliers in fault data. First, we introduce the definition of the cosine metric as follows:

Definition 1. *Given two arbitrary vectors* v_i *and* $v_j \in R^Q$ *, the cosine distance between them is*

$$
\psi(r) = \begin{cases} \sin(\pi r), & \text{if } -1 \leq r \leq 1 \\ 0, & \text{otherwise} \end{cases},
\tag{5}
$$

where $\eta \in R^+$, $v_i(c)$ *denote the cth components of* v_i.

Equation (5) can be regarded as Fourier cosine series [37] in form. It is pointed out by Fourier theory that any continuous function can be described by a series of finite number of sinusoids, and therefore such sinusoids can be approximated as a desired kernel [38]. It is worth noting that this cosine distance function can be viewed as Andrews' M-estimate [39]

when $v_i, v_j \in [0, 1]$, and its influence function, too, is equivalent to Andrews' influence function. The influence function is shown in Equation (6):

$$d(v_i, v_j) = \sum_{c=1}^{Q} \left\{ 1 - \cos\left(\eta \pi \left(v_i(c) - v_j(c)\right)\right) \right\}. \tag{6}$$

For M-estimation, the effect of the outliers on the target function is bounded if the influence function is bounded, and on this basis, the estimates of M-estimation are more robust if the target function is also bounded [40], and it can be observed that Andrews' M-estimation is bounded and robust. Since the cosine metric function is formally identical to this estimate, the cosine distance function is also robust to outliers, which can be inferred from the robustness of the M-estimate.

By algebraic transformation, Equation (6) can be derived as follows:

$$\begin{aligned} d(v_i, v_j) &= \sum_{c=1}^{Q} \left\{ 1 - \cos\left(\gamma \pi \left(v_i(c) - v_j(c)\right)\right) \right\} \\ &= \left\| \frac{1}{\sqrt{2}} \left(e^{i\gamma \pi v_i} - e^{i\gamma \pi v_j} \right) \right\|^2 \\ &= \left\| \delta_i - \delta_j \right\|^2 \end{aligned}, \tag{7}$$

where

$$\delta_i = \frac{1}{\sqrt{2}} \begin{bmatrix} e^{i\gamma \pi v_i(1)} \\ \vdots \\ e^{i\gamma \pi v_i(Q)} \end{bmatrix} = \frac{1}{\sqrt{2}} e^{i\gamma \pi v_i}, \tag{8}$$

where δ_i is the Euler representation of the vectors v_i. From Equation (7), the cosine distance between v_i and v_j is equivalent to the Euclidean distance between δ_i and δ_j. For the sake of description, we refer to the sample distance in Euler space as the Euler distance.

Compared to the Euclidean metric, the Euler metric enlarges the distance of the inter-class samples, which is useful for improving the separability of fault data. We conducted a set of distance comparison experiments to verify the conclusion. Ten samples were randomly selected from each type of data on the bearing dataset from Case Western Reserve University [41] and the gear dataset from the University of Connecticut [42]. We used Euler distance and Euclidean distance to calculate the distance between homogeneous samples and the distance between heterogeneous samples, respectively, and the results are shown in Table 1, where the bearing data are used for the drive end data at 1797 r/min speed. As can be seen from Table 1, the Euler distance simultaneously expands the distance between the samples and, importantly, expands the distance between heterogeneous samples by a greater multiple compared to samples of the same class.

Table 1. Comparison of different distance metrics.

Metric Method	Bearing Dataset		Gear Dataset	
	Same	**Different**	**Same**	**Different**
Euclidean distance	0.14	2.06	1.11	2.28
Euler distance	1.99	29.63	11.92	31.68

Based on the above analysis, the Euler representation not only can approximate the ideal robust kernel with insensitivity to outliers, but also the machinery fault data can relatively expand the differences between heterogeneous samples in the Euler space, which is helpful to increase the separability of fault samples. Therefore, we first normalize the high-dimensional fault data $X = [x_1, x_2, \cdots, x_n] \in R^{M \times n}$ to $[0, 1]$ and then map it to the Euler

space via Equation (8) to obtain the Euler representation data $\hat{X} = [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_n] \in R^{M \times n}$, where

$$\hat{x}_i = \frac{1}{\sqrt{2}} \begin{bmatrix} e^{i\gamma\pi x_i(1)} \\ \vdots \\ e^{i\gamma\pi x_i(m)} \end{bmatrix} = \frac{1}{\sqrt{2}} e^{i\gamma\pi x_i}. \tag{9}$$

3.2. Construction of Local Objective Function

The local information reflects the neighborhood structure relationship between samples, and extracting effective neighborhood discriminant features is important to improve the accuracy of fault classification. For better classification, ESBDP expects samples belonging to the same class in the neighborhood to be more aggregated and samples in different classes to be separated from each other. Therefore, we combine the label information to construct local intra-class graph G^+_{local} and local inter-class graph G^-_{local}. An example of graph construction is given in Figure 1.
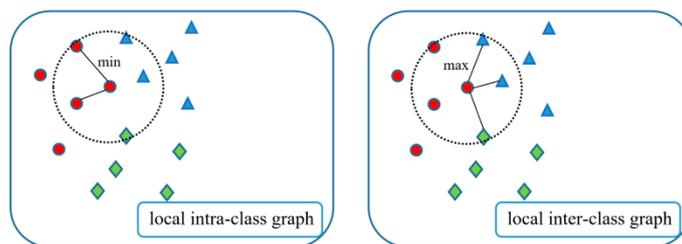

Figure 1. Example of local graph construction, where k = 5 and different shapes represent different classes of samples.

$G^+_{local} = (X, W^+_{local})$ describes the structural relationships among the intra-class samples in the neighborhood, where X denotes the set of samples and W^+_{local} is the similarity matrix. W^+_{local} is defined as

$$W^+_{local} = \begin{cases} 1 + \exp(-\frac{\|\hat{x}_i - \hat{x}_j\|_2^2}{2t^2}), & if\ \hat{x}_j \in N_k(\hat{x}_i)\ and\ c(\hat{x}_i) = c(\hat{x}_j) \\ 0, & \text{otherwise} \end{cases}, \tag{10}$$

where $\hat{x}_i$ denotes the Euler representation of x_i, t is a kernel parameter and $c(\hat{x}_i)$ is the class of $\hat{x}_i$.

According to Equation (10), the closer two sample points in the Euler space are, the greater the similarity between them. To maintain the structural relationship between the near-neighbor samples within the class after projection and obtain more compact intra-class local features, the local intra-class objective is defined as

$$\begin{aligned} G^+_{lcoal} &= \min_P \sum_{i,j=1}^n W^+_{local} \|P^T \hat{x}_i - P^T \hat{x}_j\|^2 \\ &= \min_P P^T \hat{X} D^+_{lcoal} \hat{X}^T P - P^T \hat{X} W^+_{local} \hat{X}^T P, \\ &= \min_P P^T \hat{X} L^+_{local} \hat{X}^T P \end{aligned} \tag{11}$$

where D^+_{lcoal} is a diagonal matrix with diagonal elements equal to the column sum of W^+_{local} and $L^+_{lcoal} = D^+_{lcoal} - W^+_{local}$ is the Laplace matrix.

The local inter-class graph $G^-_{local} = (X, W^-_{local})$ is used to reflect the interclass structure of the data in the neighborhood, and its similarity matrix W^-_{local} is constructed as

$$W^-_{local} = \begin{cases} 1 - \exp(-\frac{\|\hat{x}_i - \hat{x}_j\|_2^2}{2t^2}), & if\ \hat{x}_j \in N_k(\hat{x}_i)\ and\ c(\hat{x}_i) \neq c(\hat{x}_j) \\ 0, & \text{otherwise} \end{cases}. \tag{12}$$

To improve the separability of local inter-class samples, the maximization objective function is defined as

$$
\begin{aligned}
G_{local}^{-} &= \max_{P} \sum_{i,j=1}^{n} W_{local}^{-} \left\| P^T \hat{x}_i - P^T \hat{x}_j \right\|^2 \\
&= \max_{P} P^T \hat{X} D_{local} \hat{X}^T P - P^T \hat{X} W_{local} \hat{X}^T P, \\
&= \max_{P} P^T \hat{X} L_{local}^{-} \hat{X}^T P
\end{aligned}
\tag{13}
$$

where D_{lcoal}^{-} is a diagonal matrix, $(D_{lcoal}^{-})_{ii} = \sum_j (W_{local}^{-})_{ij}$ and $L_{lcoal}^{-} = D_{lcoal}^{-} - W_{local}^{-}$ is the Laplace matrix. According to Equation (12), the smaller the distance between two near-neighbor sample points belonging to different classes in Euler space, the smaller similarity is assigned to them. By maximizing the objective function, the local inter-class features with more separation can be obtained.

3.3. Construction of Global Objective Function

The global information can reflect the overall characteristics of the fault data, which is equally important for improving the fault diagnosis accuracy. To comprehensively explore the global information, we constructed a global intra-class graph $G_{global}^{+} = (X, W_{global}^{+})$ and a global inter-class graph $G_{glocal}^{-} = (X, W_{glocal}^{-})$. An example of global graph construction is given in Figure 2.
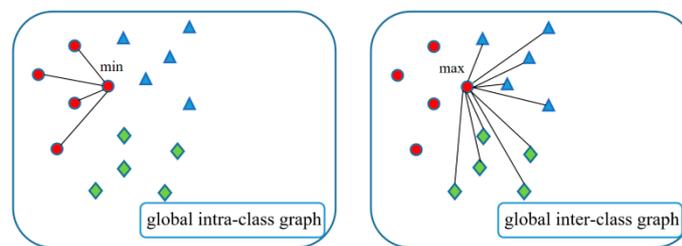

Figure 2. Example of global graph construction.

G_{global}^{+} reflects the geometric relationship between any two samples in the same class, and its similarity matrix W_{global}^{+} is defined as

$$
W_{global}^{+} = \begin{cases} \left\| \hat{x}_i - \hat{x}_j \right\|_2^2 (1 + \exp(-\frac{\left\| \hat{x}_i - \hat{x}_j \right\|_2^2}{2t^2})), & \text{if } i \neq j \text{ and } c(\hat{x}_i) = c(\hat{x}_i) \\ 0, & \text{otherwise} \end{cases}.
\tag{14}
$$

To make the fault data of the same class more aggregated after projection, the global intra-class objective is constructed as

$$
\begin{aligned}
G_{global}^{+} &= \min_{P} \sum_{i,j=1}^{n} W_{global}^{ij} \left\| P^T \hat{x}_i - P^T \hat{x}_j \right\|^2 \\
&= \min_{P} P^T \hat{X} D_{global} \hat{X}^T P - P^T \hat{X} W_{global} \hat{X}^T P, \\
&= \min_{P} P^T \hat{X} L_{global}^{+} \hat{X}^T P
\end{aligned}
\tag{15}
$$

where D_{global}^{+} is a diagonal matrix, $(D_{global}^{+})_{ii} = \sum_j (W_{global}^{+})_{ij}$ and $L_{global}^{+} = D_{global}^{+} - W_{global}^{+}$ is the Laplace matrix.

G_{glocal}^{-} reflects the geometric relationship between any two samples of different categories, and its similarity matrix W_{global}^{-} is defined as

$$
W_{global}^{-} = \begin{cases} \left\| \hat{x}_i - \hat{x}_j \right\|_2^2 (1 - \exp(-\frac{\left\| \hat{x}_i - \hat{x}_j \right\|_2^2}{2t^2})), & \text{if } i \neq j \text{ and } c(\hat{x}_i) \neq c(\hat{x}_i) \\ 0, & \text{otherwise} \end{cases},
\tag{16}
$$

The maximization global inter-class objective is defined as follows:

$$
\begin{aligned}
G_{global}^- &= \max_P \sum_{i,j=1}^n W_{global}^- \left\| P^T \hat{x}_i - P^T \hat{x}_j \right\|^2 \\
&= \max_P P^T \hat{X} D_{global} \hat{X}^T P - P^T \hat{X} W_{global} \hat{X}^T P, \\
&= \max_P P^T \hat{X} L_{global}^- \hat{X}^T P
\end{aligned}
\tag{17}
$$

where D_{global}^- is a diagonal matrix, $(D_{global}^-)_{ii} = \sum_j (W_{global}^-)_{ij}$ and $L_{global}^- = D_{global}^- - W_{global}^-$ is the Laplace matrix. In Equation (16), two samples belonging to different classes in Eulerian space are approximately close to each other; then, a lower similarity is imposed between them, and under the incentive of maximization objective, the two samples in low-dimensional space move away from each other, thus obtaining global inter-class features with well separability.

3.4. The Uniform Object Function of ESBDP

Based on the above analysis, ESBDP expects to capture both local and global discriminative information of fault data in Euler space and make the data have better inter-class separability and intra-class aggregation after projection. Therefore, the objective function of ESBDP is formulated as four optimization problems:

$$
\begin{cases}
\min_P P^T \hat{X} L_{local}^+ \hat{X}^T P \\
\max_P P^T \hat{X} L_{local}^- \hat{X}^T P \\
\min_P P^T \hat{X} L_{global}^+ \hat{X}^T P \\
\max_P P^T \hat{X} L_{global}^- \hat{X}^T P
\end{cases}
\tag{18}
$$

To better unify the four objective functions into a single objective model, we transform the intra-class minimization problem into the equivalent maximization problem.

$$
\begin{aligned}
\min_P P^T \hat{X} L_{local}^+ \hat{X}^T P &= P^T \hat{X} (D_{local}^+ - W_{local}^+) \hat{X}^T P \\
&\Rightarrow \max_P P^T \hat{X} W_{local}^+ \hat{X}^T P
\end{aligned}
\tag{19}
$$

$$
\begin{aligned}
\min_P P^T \hat{X} L_{global}^+ \hat{X}^T P &= P^T \hat{X} (D_{global}^+ - W_{global}^+) \hat{X}^T P \\
&\Rightarrow \max_P P^T \hat{X} W_{global}^+ \hat{X}^T P
\end{aligned}
\tag{20}
$$

Combining the above four optimization problems, the objective model of ESBDP is defined as

$$
\begin{aligned}
\max_P P^T \hat{X} \Big((1-\delta)(L_{local}^- + W_{local}^+) + \delta(L_{global}^- + W_{global}^+) \Big) \hat{X}^T P \\
s.t.\ P^T \hat{X} (D_{local} - D_{global}) \hat{X}^T P = 1.
\end{aligned}
\tag{21}
$$

where $D_{local} = D_{local}^+ + D_{local}^-$, $D_{local} = D_{global}^+ + D_{global}^-$ and $\delta \in [0,1]$ is a balance parameter to regulate the elasticity between global and local information.

In Equation (21), the balance parameter determines the relative importance of the global and local targets, which plays an essential role in obtaining effective fault features. However, there are some problems that need to be solved here, such as the determination of the appropriate parameter values. Usually, the parameter needs to be set artificially in the range first, and the final parameter value is obtained by methods such as grid search. It is a challenge to manually select a suitable parameter in different fault applications, and the discriminative power of the extracted relevant features may reduce if the parameter is not selected properly. Therefore, we offer an adaptive balancing strategy to further optimize the uniform objective function of ESBDP.

3.5. Optimization of the Uniform Objective Function

In this section, we offer an adaptive balance strategy to construct an optimized model for ESBDP. Compared with setting the parameters manually, we expect the ESBDP algorithm to adaptively select the best balancing parameters to achieve an effective balance of global and local features, thus further enhancing the discriminative power of the extracted fault features. Therefore, Equation (21) is optimized as follows:

$$\arg\max_{P,\alpha_l,\alpha_g} \alpha_l{}^2 P^T \hat{X} H_{local} \hat{X}^T P + \alpha_g{}^2 P^T \hat{X} H_{global} \hat{X}^T P$$
$$s.t. \ P^T \hat{X}(D_{local} - D_{global})\hat{X}^T P = 1. \quad\quad , \quad\quad (22)$$
$$\alpha_l + \alpha_g = 1.$$

where $H_{local} = L_{local}^- + W_{local}^+$, $H_{global} = L_{global}^- + W_{global}^+$, α_l corresponds to the weight parameter of the local target and α_g is the weight parameter of the global target. While learning the local and global structural features of the fault data, the model of Equation (22) obtains intra-class minimization and inter-class maximization, and in addition, the model achieves adaptive balance between global and local objectives.

In the optimization model of ESBDP, the projection direction and the balance parameters are the objectives to be solved. Since it is difficult to solve both objectives at the same time, we use a way of iterative update, i.e., fix one solution objective, update the other one, and stop the iterative process when the solution converges. The following are the solution steps.

(1) Fixing the balance parameters α_l and α_g, we update the projection matrix P.

Fixing the balance parameters as constants, Equation (22) can be expressed as

$$\arg\max_{P} \alpha_l{}^2 P^T X H_{local} X^T P + \alpha_g{}^2 P^T X H_{global} X^T P$$
$$s.t. \ P^T X(D_{local} - D_{global})X^T P = 1. \quad\quad (23)$$

The Lagrange multiplier method was used to solve for Equation (23).

$$L(P) = \alpha_l^2 P^T X H_{lcoal} X^T P + P^T X \alpha_g^2 H_{global} X^T P$$
$$-\lambda(P^T X(D_{local} - D_{global})X^T P - 1) \quad\quad (24)$$

where λ is the Lagrangian multiplier. The value of the partial derivative of Equation (24) with respect to P is zero.

$$\frac{\partial L(P)}{\partial P} = \alpha_l^2 X H_{local} X^T P + \alpha_g^2 X H_{global} X^T P - \lambda\left(X(D_{local} - D_{global})X^T P - 1\right) = 0. \quad (25)$$

Then, Equation (25) can be transformed into

$$\alpha_l^2 X H_{lcoal} X^T P + \alpha_g^2 X H_{global} X^T P = \lambda X(D_{local} - D_{global})X^T P. \quad\quad (26)$$

Solving Equation (26), we can calculate that the eigenvector $\lambda = [\lambda_1, \lambda_2, \cdots, \lambda_d]$ corresponds to the first d largest eigenvalues, which is the desired projection direction.

(2) Fixing P, we update α_l and α_g.

Fixing the projection matrix, Equation (18) can be expressed as

$$\arg\max_{\alpha_l,\alpha_g} \alpha_l{}^2 P^T X H_{local} X^T P + \alpha_g{}^2 P^T X H_{global} X^T P$$
$$s.t. \ \alpha_l + \alpha_g = 1. \quad\quad (27)$$

Using the Lagrange multiplier method to solve Equation (27).

$$L(\alpha_l, \alpha_g) = \alpha_l^2 tr(P^T X H_{local} X^T P) + \alpha_g^2 tr(P^T X H_{global} X^T P) - \lambda(\alpha_l + \alpha_g - 1). \quad (28)$$

We let the value of the partial derivative of Equation (28) with respect to the balance parameter be zero.

$$
\begin{cases}
\frac{\partial L(\alpha_l, \alpha_g)}{\partial \alpha_l} = 2\alpha_l tr(P^T X H_{local} X^T P) - \lambda = 0 \\
\frac{\partial L(\alpha_l, \alpha_g)}{\partial \alpha_g} = 2\alpha_g tr(P^T X H_{global} X^T P) - \lambda = 0 \\
\frac{\partial L(\alpha_l, \alpha_g)}{\partial \lambda} = \alpha_l + \alpha_g - 1 = 0
\end{cases}
\tag{29}
$$

By mathematical transformation, the balance parameters can be expressed as

$$
\alpha_l = \frac{tr(P^T X H_{global} X^T P)}{tr(P^T X H_{local} X^T P) + tr(P^T X H_{global} X^T P)},
\tag{30}
$$

$$
\alpha_g = \frac{tr(P^T X H_{local} X^T P)}{tr(P^T X H_{local} X^T P) + tr(P^T X H_{global} X^T P)}.
\tag{31}
$$

Finally, the fault data are projected into the subspace by $Y = P^T X$ to obtain the low-dimensional representation features. Algorithm 1 describes the steps of the ESBDP algorithm.

Algorithm 1 Euler Representation Based Structural Balance Discriminant Projection

---

Input: Training fault sample set $X \in [x_1, x_2, \cdots, x_n] \in R^{m \times n}$
Output: projection matrix P
(1) Conversion of the original feature data X into Euler representation data $\hat{X}$ by Equation (9).
(2) Construction the weight matrix W_{local}^+, W_{local}^+, W_{global}^+ and W_{global}^-.
(3) Fixing the balance parameters α_l and α_g, update P.
(4) Fixing the projection matrices P, update α_l and α_g.
(5) Repeat (3) and (4) until convergence.

---

4. Fault Diagnosis Process Based on ESBDP Algorithm

In this section, the fault diagnosis process based on the ESBDP algorithm is described, and the diagnosis flowchart is provided in Figure 3, where n denotes the number of classes of fault data. The following is a specific description of the diagnosis steps:

Step 1. The vibration data of machinery equipment during operation are collected, and the original data are transformed into time–frequency domain features to construct a high-dimensional fault feature set. The high-dimensional feature set can reduce the influence of the noise and the nonlinear features in the original data [43]. The relevant characteristic parameters are provided in Table 2, where $v(i)$ denotes the time series of the signal and $s(k)$ is the frequency spectrum of $v(i)$. In Table 3, p1~p17 are the time domain features, which are mean value, square root amplitude, peak value, absolute mean, maximum value, root mean square value, variance, minimum value, standard deviation, impulse factor, kurtosis, skewness, waveform factor, peak factor, peak-to-peak value, margin factor and gap factor. p18~p21 are the frequency domain features of the signal, which are the mean frequency, center of gravity frequency, root mean square frequency, and standard deviation of frequency. p22~p29 are the time–frequency domain features of the signal. More specific definitions of the feature parameters can be referred to [9,10,28,44]. Then, the feature set is divided into a training set X_{train} and a testing set X_{test} in a certain proportion by means of a random selection.
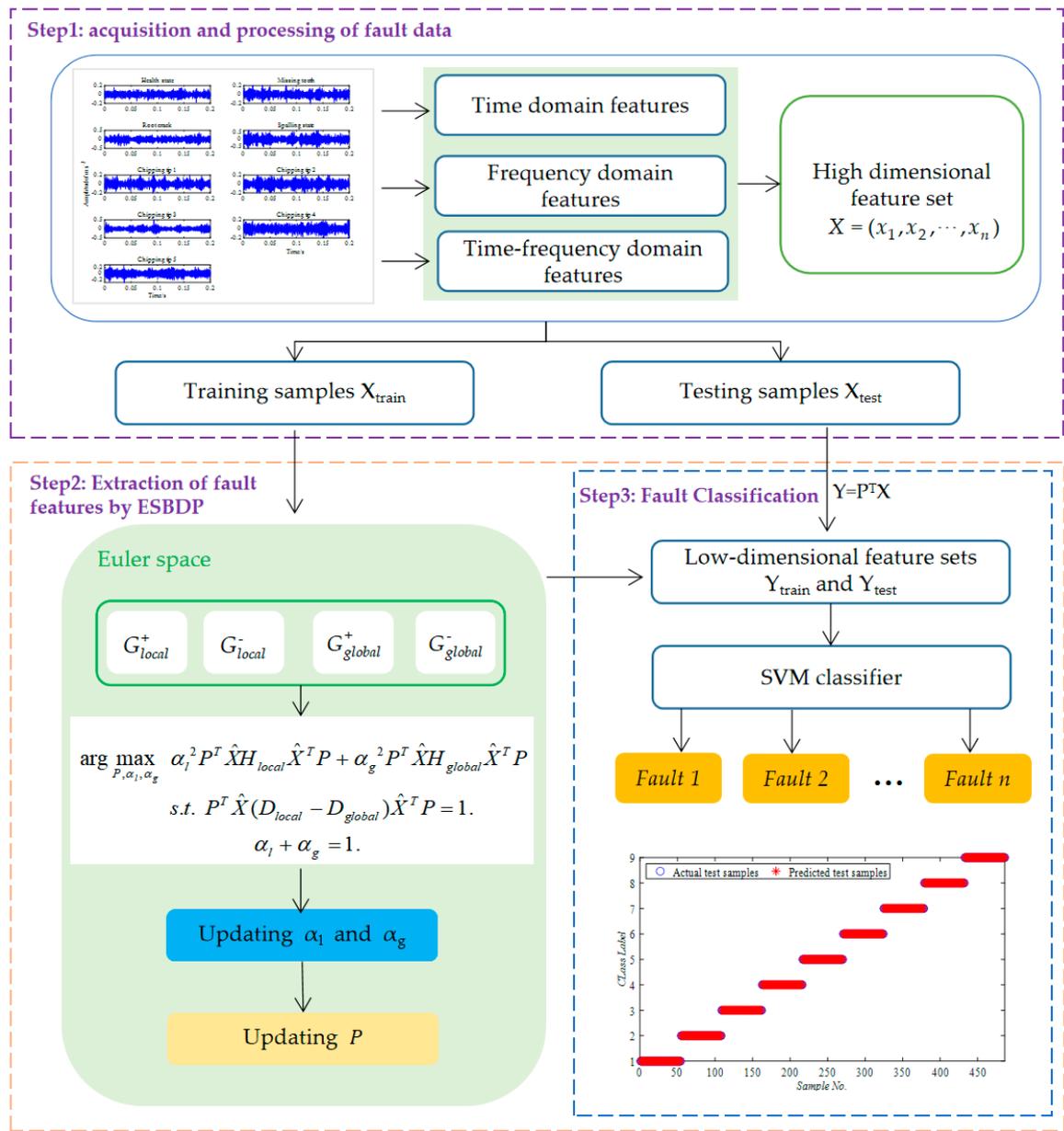
Figure 3. Procedure of fault diagnosis based on ESBDP.

Table 2. Time–frequency characteristic parameters.

No.	Parameters	No.	Parameters	No.	Parameters
1	$p_1 = \sum_{i=1}^{n} v(i)/n$	9	$p_9 = \sqrt{\sum_{i=1}^{n} (v(i) - p_4)^2 / (n-1)}$	17	$p_{17} = p_3 / p_6^2$
2	$p_2 = (\sum_{i=1}^{n} \sqrt{|v(i)|}/n)^2$	10	$p_{10} = p_3 / p_4$	18	$p_{18} = \sum_{k=1}^{k} f_k s(k) / \sum_{k=1}^{k} s(k)$
3	$p_3 = \max|v(i)|$	11	$p_{11} = \sum_{i=1}^{n} (v(i) - p_1)^4 / (n-1) p_9^4$	19	$p_{19} = \sum_{k=1}^{k} s(k)/k$
4	$p_4 = \sum_{i=1}^{n} |v(i)|/n$	12	$p_{12} = \sum_{i=1}^{n} (v(i) - p_1)^2 / (n-1) p_9^3$	20	$p_{20} = \sqrt{\sum_{k=1}^{k} f_k^2 s(k) / \sum_{k=1}^{k} s(k)}$
5	$p_5 = \max(v(i))$	13	$p_{13} = p_2 / p_4$	21	$p_{21} = \sqrt{\sum_{k=1}^{k} (f_k - T_{18})^2 s(k) / \sum_{k=1}^{k} s(k)}$
6	$p_6 = \sqrt{\sum_{i=1}^{n} (v(i))^2 / n}$	14	$p_{14} = p_3 / p_6$	22~29	Three-layer wavelet packet decomposition band energy characteristic.
7	$p_7 = \sum_{i=1}^{n} (v(i) - p_4)^2 / n$	15	$p_{15} = p_5 - p_8$		
8	$p_8 = \min(v(i))$	16	$p_{16} = p_3 / p_2$		

Table 3. Fault recognition rate of bearing dataset processed by different method.

Fault Type	Recognition Accuracy (%)					
	LLE	**LPP**	**KPCA**	**OLGPP**	**LSPD**	**ESBDP**
Normal	100	100	100	100	99.83	100
Inner race	93.33	83.67	96.17	98.17	100	100
Ball	88.83	83.83	82.33	96.67	100	100
Outer race	91.0	99.67	100	96.67	100	100
Average recognition rate	93.29	91.79	94.63	98.63	99.96	100
Standard deviation	0.95	6.55	0.72	0.94	0.13	0.00
Processing time (s)	0.80	0.32	0.33	0.71	0.50	0.34

Step 2. The training set X_{train} is mapped to the Euler space, and the ESBDP optimization model based on four objectives is constructed in this space. Solving this model, we can obtain the projection matrix P. Then, the low-dimensional representation features Y_{train} and Y_{test} are obtained by projecting X_{train} and X_{test} into the subspace using P.

Step 3. The low-dimensional training sample Y_{train} is input to the SVM classifier for learning, and finally, the fault type of the test sample is predicted based on the completed training classifier.

5. Experimental Results and Analysis

5.1. Experiments on the Bearing Dataset

In this experiment, the bearing data were used from the Case Western Reserve University (CWRU) rolling bearing dataset. As shown in Figure 4, the data acquisition platform is divided into three main parts, from left to right, the motor, the torque sensor, and the dynamometer. The drive-end bearing vibration data at 12 kHz sampling frequency and 1772 rpm were used in our diagnosis experiment, containing four different types of data, namely normal data, inner race defect data, outer race defect data and ball defect data. Figure 5 shows the one-dimensional signals of the four states. With a sampling length of 1024, 100 samples were obtained for each type of signal data. Then, the high-dimensional feature set is obtained by extracting the time–domain, frequency–domain and time–frequency features of the original fault samples according to the process described in Section 4. The training set is composed of 40 samples randomly selected from each class, and the rest are used as test samples. To avoid the chance of experimental results, the random experiment was repeated ten times by us. Finally, the fault recognition accuracy was taken as the average of the ten-time results.
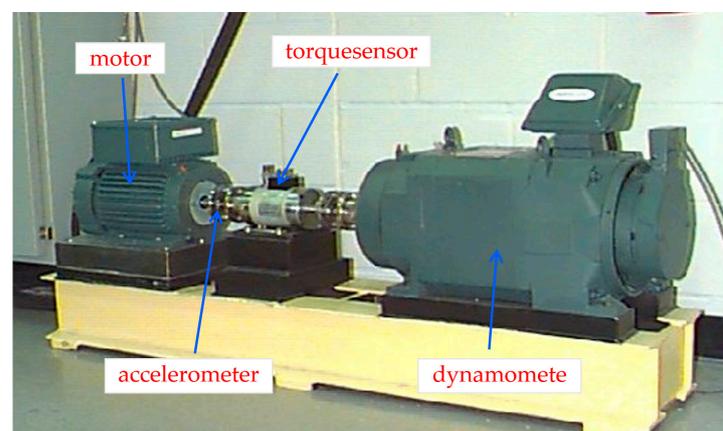

Figure 4. Bearing test platform.
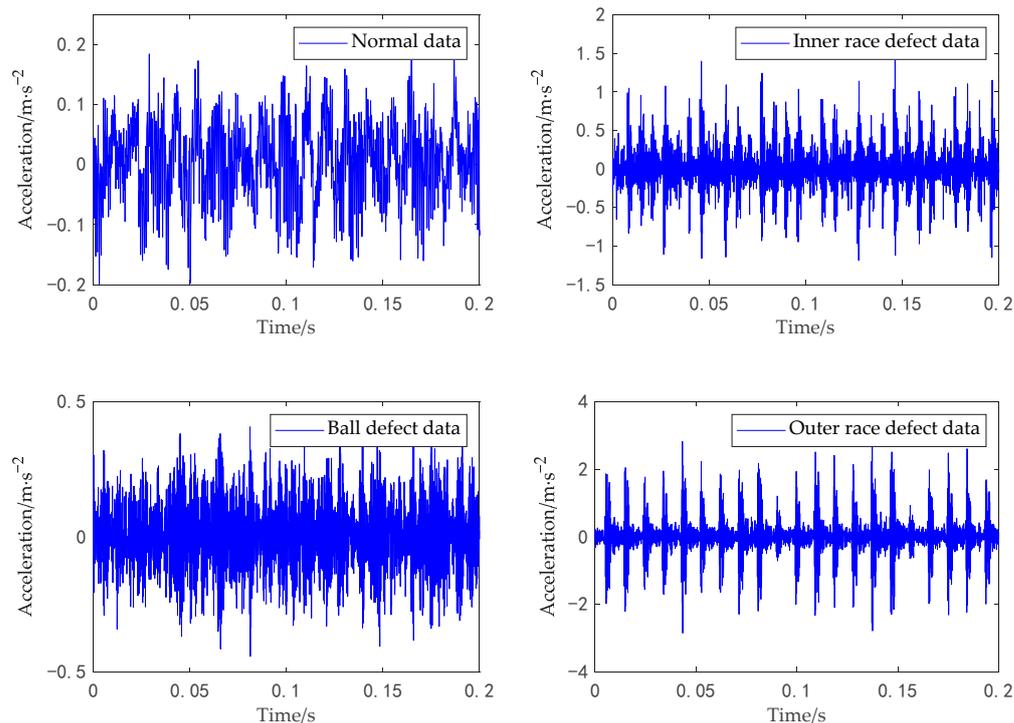
Figure 5. Raw vibration signals of four bearing data.

To obtain the appropriate parameters for the ESBDP algorithm, we conducted several experiments using the grid search method, and through comparison, we finally set the neighborhood parameter value $k = 9$, the adjustment parameter $\gamma = 1$, and the kernel parameter $t = 0.01$. In addition, for comparison with our algorithm, five related algorithms, KPCA, LPP, LLE, OLGPP, and LSPD were used for the same experiments, and the parameters of each algorithm were selected using the grid search method. Among them, the kernel parameter of the global nonlinear algorithm KPCA is $t_{kpca} = 2$. The neighborhood parameters of the LPP, LLE, OLGPP, and LSPD [45] algorithms are set to $k_{lpp} = 12$, $k_{lle} = 7$, $k_{\mathrm{olgpp}} = 17$, and $k_{\mathrm{lspd}} = 10$, respectively. The kernel parameters of LPP, KPCA and OLGPP are set to $t_{lpp} = 1$ and $t_{\mathrm{olgpp}} = 0.1$, respectively.

To visualize the classification performance of ESBDP, we show the distribution of the sample features in the three-dimensional space after projection in Figure 6 and compare it with the LLE, LPP, KPCA, OLGPP and LSPD algorithms, where the horizontal axis denotes the first dimension of Y_{test}, the vertical axis denotes the second dimension of Y_{test}, and the other axis denotes the third dimension of Y_{test}. It can be seen that the visualization distribution results of LLE, LPP and KPCA are relatively poor, showing that the same kind of features are scattered and there is a serious overlap between different types of data without a clear demarcation. The reason for this situation is that all three algorithms only singularly consider the neighborhood information or global structure of the fault data and extract incomplete information; in addition, the supervision information is not used by these algorithms. Compared with the above three algorithms, the visualization results of OLGPP and LSPD are relatively good, but there are some cases of scattered data within the class and the heterogeneous samples are relatively close to each other. The visualization of the three-dimensional feature distribution based on the ESBDP method is optimal, with high aggregation between the same class of samples, high separation between the heterogeneous samples, and obvious demarcation between classes, which provides a favorable basis for the subsequent fault classification.
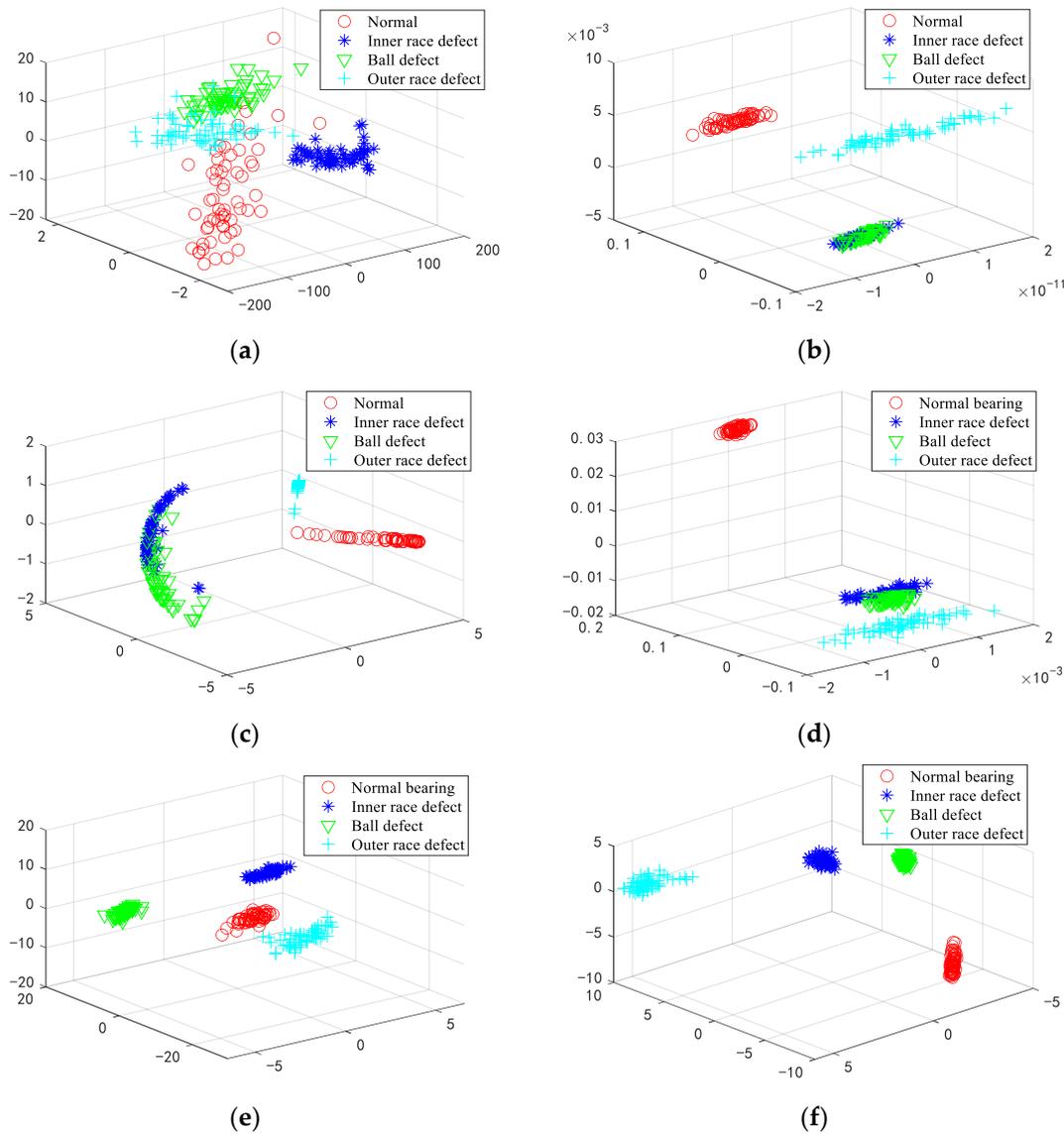
Figure 6. Three-dimensional distribution of test bearing samples: (**a**) LLE; (**b**) LPP; (**c**) KPCA; (**d**) OLGPP; (**e**) LSPD; (**f**) ESBDP.

To further explain the low-dimensional feature separability of ESBDP, we adopted the ratio of inter-class distance and intra-class distance as the separability index. The inter-class distance S_b is used to reflect the separability between classes, while the intraclass distance S_w reflects the aggregation of samples within classes, and they can be calculated by Equations (32) and (33), where c denotes the number of classes of samples, $\overline{y}_i = \sum_{i=1}^{n} y_i / n$ denotes the class center, and y_{ji} denotes the ith sample of the jth class. For low-dimensional features, the larger the separability parameter metric, the better its relative separability. Figure 7 provides the divisibility metrics based on the six algorithms, and it can be seen that the divisibility parameters of EGBDP are higher than those of the other algorithms. Combining Figures 6 and 7, we can conclude that ESBDP has better dimensionality reduction performance and can provide a favorable basis for subsequent fault classification.

$$S_b = \frac{1}{c(c-1)} \sum_{i,j}^{c} \left\| \overline{y}_i - \overline{y}_j \right\|, \tag{32}$$

$$S_w = \frac{1}{c} \sum_{j=1}^{c} \frac{1}{n} \sum_{i}^{n} \left\| y_{ji} - \bar{y}_j \right\| \tag{33}$$
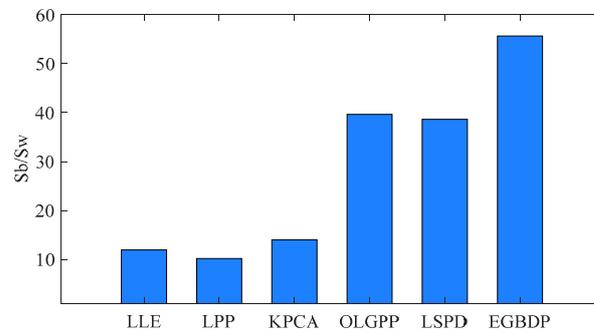
Figure 7. Low-dimensional feature separability index based on different algorithms.

Table 3 provides the average fault recognition accuracy and its standard deviation and processing time for ten random experiments based on the six methods of LLE, LPP, KPCA, OLGPP, LSPD and ESBDP. The standard deviation is used to reflect the fluctuation of the recognition rate such as the smaller standard deviation indicating the smoother performance of the algorithm. From Table 3, it can be observed that the recognition results of the three algorithms, LLE, LPP and KPCA, which only consider the global or local structure, are lower for nonlinear unstable machinery fault data. Among them, LPP has a recognition rate of 91.79% by discovering the local discriminative features on the sample manifold, and the large standard deviation indicates its poor stability. LLE is a nonlinear algorithm in the manifold algorithm that can respond to the sample globally with local linearity, and its low-dimensional samples can keep the original topology, and the fault diagnosis accuracy of this algorithm is 93.29%. With the kernel function, KPCA can capture the nonlinear global structure information of the fault data, and the recognition rate is 94.63%. Since OLGPP captures both global and local information of the data, its diagnosis accuracy is higher than the above three methods, but OLGPP does not utilize the supervised information of the fault data, so it is difficult to achieve the correct classification of all the data. While using supervised information, LSPD measures and preserves the similarity between fault samples by constructing a similarity function, which improves the accuracy to some extent, but the algorithm focuses more on the local structure of the samples and the global information is not effectively used.

ESBDP achieves the highest recognition accuracy because the algorithm maps the fault data into the Euler space through the cosine metric and integrates the structural relationships of local intra-class, local inter-class, global intra-class and global inter-class samples in this space. On the basis of expanding the differences between heterogeneous fault samples, the inter-class separation and intra-class aggregation between samples are effectively improved. In addition, ESBDP achieves an adaptive balance between global and local features, which further enhances the discriminative power of the extracted fault features. Remarkably, our method has the smallest standard deviation and the processing time is only 0.34 s, which means that the ESBDP algorithm has a low computational burden and is more stable.

The specific classification results are provided in Figure 8. It can be observed that the classification errors exist in all algorithms except our algorithm. Among them, in the prediction results of the LPP and KPCA algorithms, the classification errors are mainly concentrated on the second and third fault types of data. Classification errors are detected in LLE and OLGPP for all other three types of faults except normal data. The classification results of the LSPD algorithm are better, but there is also a normal sample that is misclassified. Our method can accurately identify samples of each fault type, which shows the superiority of ESBDP in clustering and classification of all types of fault data.
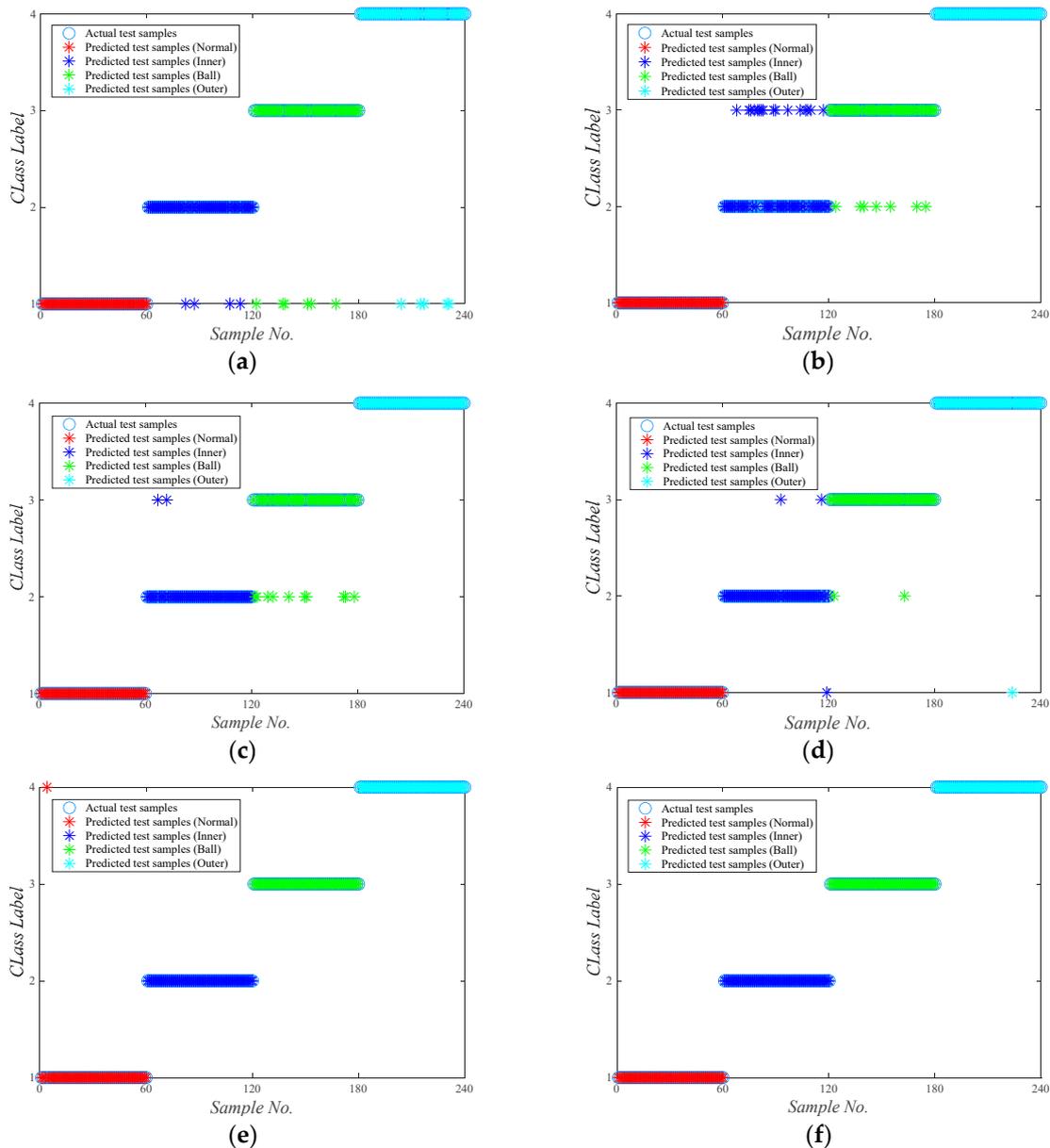
Figure 8. The classification results of the bearing data: (**a**) LLE; (**b**) LPP; (**c**) KPCA; (**d**) OLGPP; (**e**) LSPD; (**f**) ESBDP.

To further verify the ability of the ESBDP algorithm to capture fault information, we used different numbers of samples to train each algorithm and then observed the change in fault diagnosis results. Figure 9 shows the corresponding experimental results. Overall, as the number of training samples increases, the recognition rate of all six algorithms increases, because the training samples contain the discriminative information required for fault classification, and the more training samples, the more discriminative features can be learned. Among them, LLE, LPP and KPCA algorithms have lower recognition rates. In addition, the accuracy of LPP, LLE and KPCA can hardly be improved even if the training samples are increased due to the performance of the algorithms. ESBDP outperforms the other five algorithms in terms of recognition rate and stability. It is worth noting that ESBDP can achieve a 100% fault recognition rate at 20 training samples per class, which means that our algorithm has a strong ability to capture fault discriminative information and can better perform fault diagnosis tasks.
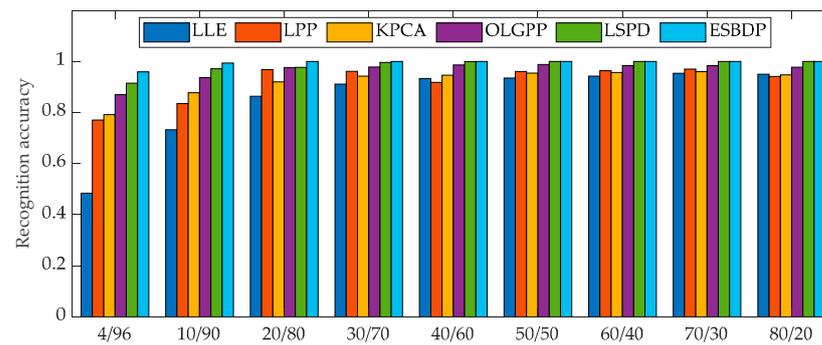
Figure 9. Recognition accuracy under sample variation.

To further investigate the performance of the ESBDP algorithm, we experimented with bearing fault data for four operating conditions at 1797 r/min, 1772 r/min, 1750 r/min, and 1730 r/min, respectively. Following the process in Chapter 3, the identification results of the six algorithms are shown in Figure 10. It can be concluded that for the bearing data under different working conditions, the recognition rates of the first three algorithms are generally low and unstable, and the accuracy rates of OLGPP and LSPD are relatively good, but they cannot classify the four data correctly at the same time. The ESBDP algorithm is the most stable and achieves the highest accuracy rate for the diagnosis of all four data. Experiments show that compared with other methods, ESBDP has stronger adaptability and can also mine effective fault discrimination features for data under different working conditions to achieve accurate classification.
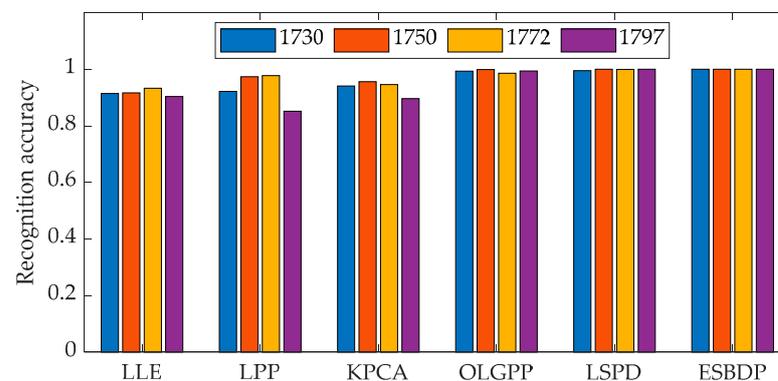
Figure 10. Recognition accuracy at different speeds.

5.2. Experiments on the Gear Dataset

The experimental data were obtained from a gear fault dataset published by the University of Connecticut, and the data collection platform is shown in Figure 11. It is a two-stage gearbox containing a motor for controlling the gear speed, a tachometer for measuring the speed, an electromagnetic brake for providing torque, and two-stage input shaft equipped with gears. The gear operation data are obtained from the pinion of the first stage via an accelerometer with a sampling frequency of 20 KHZ and contain nine types of gear states for spalling, root crack, missing tooth, five different severity levels of chipping tip, and health state. To facilitate the description, we denote these nine types of data by {F1, F2, F3, F4, F5, F6, F7, F8, F9}. The vibration signals of the nine gear states are provided in Figure 12. A total of 104 samples were collected for each type of data, totaling 936 samples with a dimension of 3600. Following the procedure in Section 4, the statistical features in the time–frequency domain were computed for each sample to obtain the high-dimensional feature set. In this experiment, we randomly chose 50 samples from each of the nine data types as training samples and the rest of the samples as prediction samples. In addition, the

parameters of each algorithm in the gear experiment are the same as those of the bearing diagnosis experiment.
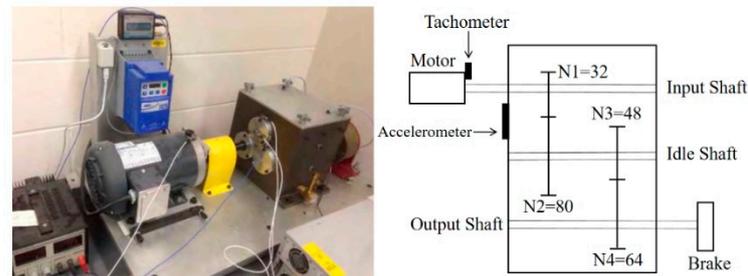

Figure 11. Gearbox experimental platform.
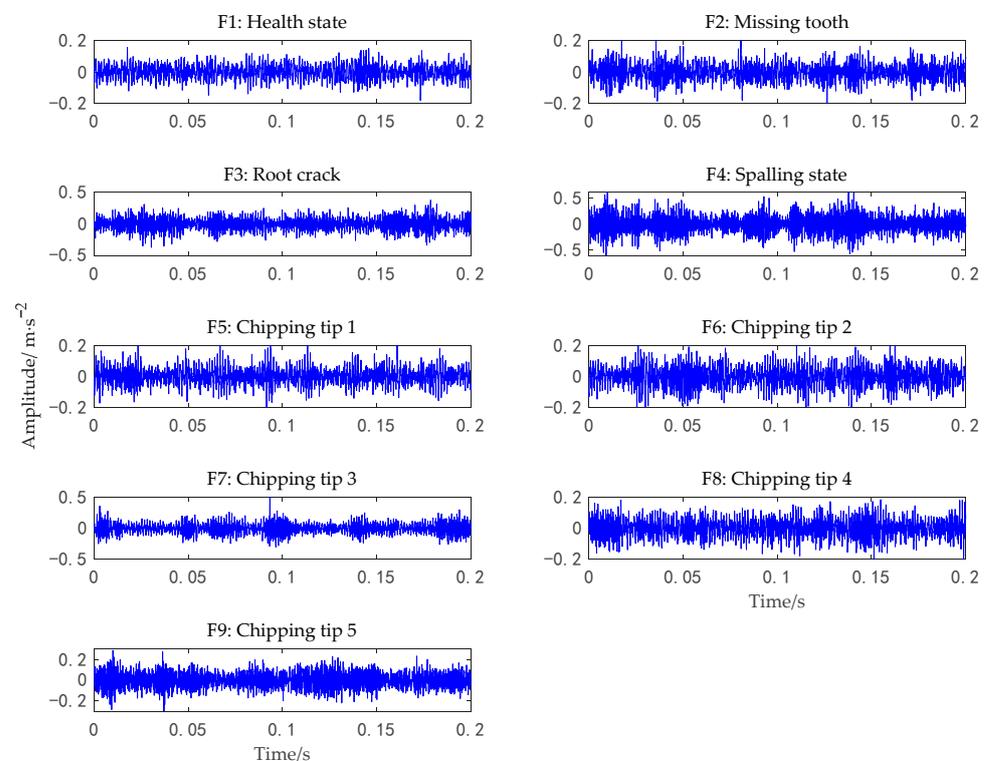
Figure 12. Raw vibration signals of nine gear data.

Figures 13 and 14 show the three-dimensional distribution and separability indexes of the low-dimensional features after projection based on the six algorithms, respectively. Due to the increase in categories, the distribution of features is more complex. Among them, the separable indexes of LLE, LPP and KPCA are relatively low and the visualization effect is relatively poor. There is a serious overlap between multi-category features without clear demarcation, which makes it difficult to identify effectively. The main reason is that these three methods only consider the neighborhood structure or global information of the fault data. OLGPP is also more confounded between classes because it does not utilize the label information between samples. The comparison shows that for multi-class gear fault data, the ESBDP method has the highest separability index and its three-dimensional feature distribution visualization is also optimal, where homogeneous features are aggregated with each other and heterogeneous features have relatively obvious boundaries, which shows that our algorithm can effectively improve the aggregation between samples of the same class and the separability between heterogeneous samples.
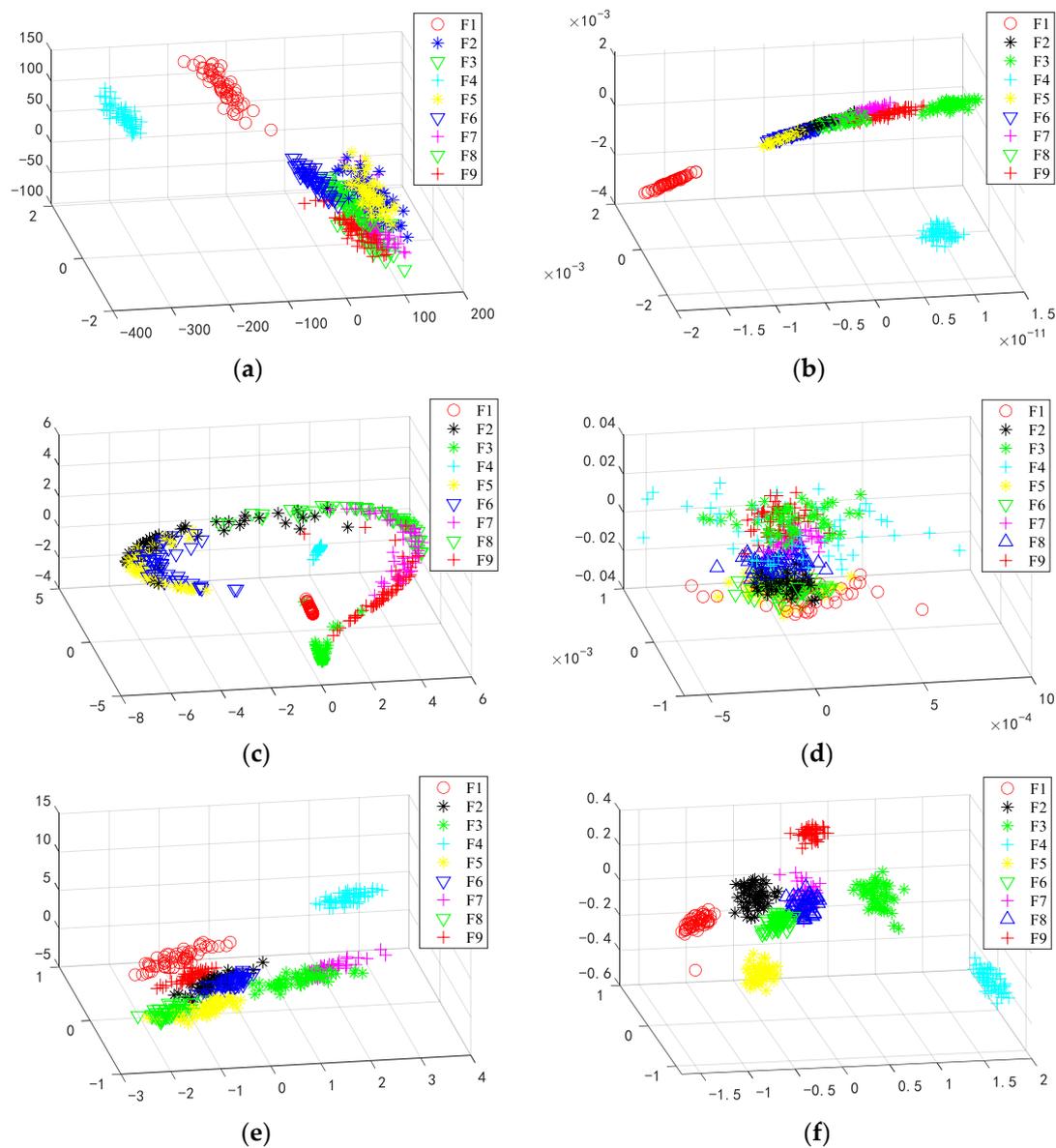
Figure 13. Three-dimensional distribution of test gear samples: (**a**) LLE; (**b**) LPP; (**c**) KPCA; (**d**) OLGPP; (**e**) LSPD; (**f**) ESBDP.
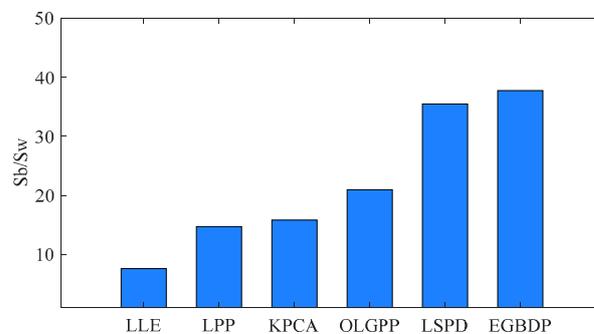

Figure 14. Low-dimensional feature separability index based on different algorithms.

The gear diagnosis accuracy is provided in Table 4, and its corresponding classification details are provided in Figure 15. For gear data with multiple classes, the recognition accuracy of LLE, LPP and KPCA algorithms, which fail to combine both local and global

discriminative features of fault data, decreases more. The accuracy of the OLGPP algorithm also decreases because the supervised information is not utilized by the algorithm, which makes it not advantageous in the diagnosis of multi-category fault data. In addition, there are more parameters in the OLGPP algorithm, which makes it hard to effectively diagnose different fault data without changing the parameters. It is worth noting that our algorithm achieves accurate classification for all types of gear data. On the basis of expanding the differences between heterogeneous samples, ESBDP combines label information to fully consider the geometric structure relationship of fault data to achieve the acquisition and balance of local and global features, which enables effective fault discriminative features to be extracted. The results of gear diagnosis further demonstrate the effectiveness of our algorithm.

Table 4. Fault recognition accuracy of gear dataset processed by different method.

| Fault Type	Recognition Accuracy (%)					
	LLE	LPP	KPCA	OLGPP	LSPD	ESBDP
Health state	92.59	100	100	100	100	100
Missing tooth	57.41	66.67	74.07	77.78	100	100
Root crack	74.07	98.15	94.44	96.30	100	100
Spalling	88.89	100	100	100	100	100
Chipping tip 1	64.81	88.89	81.48	81.48	96.30	100
Chipping tip 2	92.59	96.30	100	100	100	100
Chipping tip 3	79.63	87.04	83.33	92.59	100	100
Chipping tip 4	87.04	87.04	88.89	98.15	100	100
Chipping tip 5	62.96	77.78	96.30	100	98.15	100
Average recognition accuracy	77.78	89.09	90.95	93.83	99.38	100
Processing time (s)	0.57	0.39	0.24	0.51	0.33	0.40
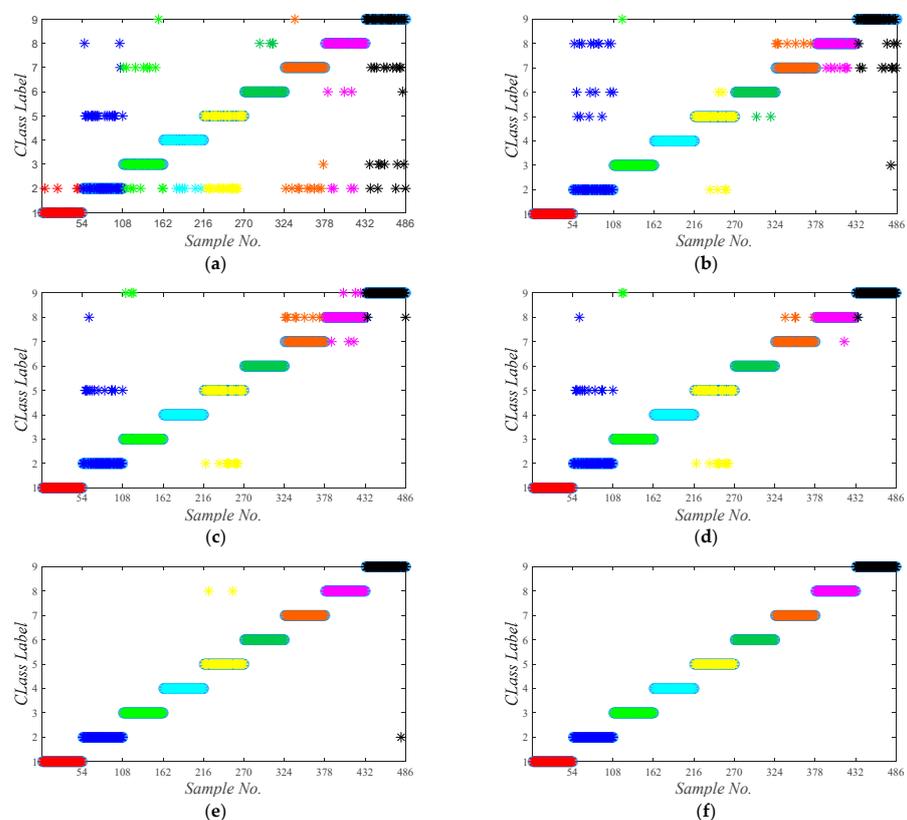
Figure 15. Three-dimensional distribution of test gear samples, where the different colors indicate different classes of samples. (**a**) LLE; (**b**) LPP; (**c**) KPCA; (**d**) OLGPP; (**e**) LSPD; (**f**) ESBDP.

To further verify the ability of ESBDP to capture fault features, the number of samples used for training and the number of samples used for prediction were set to 2/102, 5/99, 10/94, 20/84, 30/74, 40/64, 50/54, 60/44, and 70/34 in the gear diagnosis experiments, respectively, to observe the variation of the accuracy rate of each algorithm. The results are provided in Figure 16. For different numbers of training samples, the fault diagnosis accuracy of ESBDP is consistently higher than that of the other five algorithms. Notably, the recognition accuracy of ESBDP reaches 99.53% at 10 training samples per class, which indicates that the algorithm has a great ability to capture the discriminative features hidden in the fault data.
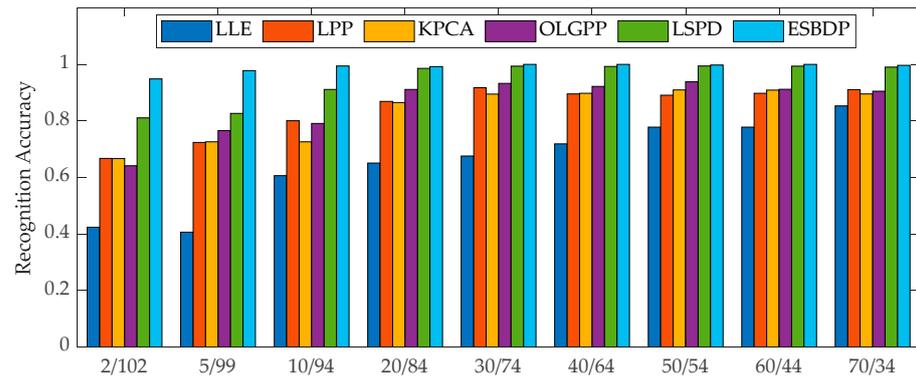
Figure 16. Recognition accuracy under sample variation.

6. Conclusions

To address the problem of fault diagnosis methods sensitive to outliers and the difficulty to obtain and balance global and local discriminative information simultaneously, this paper proposes the ESBDP algorithm for rotating machinery fault diagnosis.

(1) The algorithm maps the high-dimensional fault features to the Euler representation space through the cosine metric, and the Euler representation approximates an ideal robust kernel that can suppress the influence of outliers in fault data. Meanwhile, the comparative analysis of Euler distance and Euclidean distance concludes that the heterogeneous fault samples have greater differences in the Euler space, which improves the inter-class separability to a certain extent.

(2) Through the adaptive balance strategy, the algorithm fully considers the geometric structure relationship of the fault data and constructs an optimization model based on four objective functions with different structure and class information in the Euler space, which can balance the local and global structures of the fault data and further im-prove the local intra-class aggregation and global inter-class separability of the low-dimensional discriminant features.

(3) In two rotating machinery fault experiments of bearings and gearboxes, the effectiveness and stability of the algorithm are analyzed from the perspectives of three-dimensional feature distribution of test samples, confusion matrix and fault recognition rate under multiple working conditions. The experimental results show that the algorithm proposed in this paper can effectively improve the accuracy of fault classification and has superior fault diagnosis performance.

Although ESBDP has many advantages, there are still some potential problems. For example, some parameter tuning is needed in the ESBDP algorithm, as well as a way to select parameter values adaptively. The selection of time and frequency domain feature parameters also affects the results of fault diagnosis and ways to obtain the optimal combination of time and frequency statistical features. All these issues will be further investigated in the future.

Abbreviations

PCA	principal components analysis
KPCA	kernel principal components analysis
LLE	local linear embedding
MFA	marginal fisher analysis
NPE	neighborhood preserving embedding
LPP	locality preserving projection
NLSPP	nonlocal and local structure preserving projection
FDGLPP	Fisher discriminative global local preserving projection
GLMDPP	global–local marginal discriminative preserving projection
GLMFA	global–local margin Fisher analysis
OLGPP	orthogonal locality and globality preserving projection
LGBODP	local–global balanced orthogonal discriminant projection
ESBDP	Euler representation-based structural balance discriminant projection
CWRU	Case Western Reserve University
LSPD	local similarity preserving discriminant
SVM	support vector machine

References

1. Liu, Y.; Hu, Z.; Zhang, Y. Symmetric positive definite manifold learning and its application in fault diagnosis. *Neural Netw.* **2022**, *147*, 163–174. [CrossRef] [PubMed]
2. Qi, R.; Zhang, J.; Spencer, K. A Review on Data-Driven Condition Monitoring of Industrial Equipment. *Algorithms* **2022**, *16*, 9. [CrossRef]
3. Zhang, Z.; Huang, W.; Liao, Y.; Song, Z.; Shi, J.; Jiang, X.; Shen, C.; Zhu, Z. Bearing fault diagnosis via generalized logarithm sparse regularization. *Mech. Syst. Signal Process.* **2022**, *167*, 108576. [CrossRef]
4. Fekih, A.; Habibi, H.; Simani, S. Fault Diagnosis and Fault Tolerant Control of Wind Turbines: An Overview. *Energies* **2022**, *15*, 7186. [CrossRef]
5. Wang, W.; Yuan, F.; Liu, Z. Sparsity discriminant preserving projection for machinery fault diagnosis. *Measurement* **2021**, *173*, 108488. [CrossRef]
6. Su, Z.; Tang, B.; Liu, Z.; Qin, Y. Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine. *Neurocomputing* **2015**, *157*, 208–222. [CrossRef]
7. Peng, B.; Bi, Y.; Xue, B.; Zhang, M.; Wan, S. A Survey on Fault Diagnosis of Rolling Bearings. *Algorithms* **2022**, *15*, 347. [CrossRef]
8. Brusa, E.; Cibrario, L.; Delprete, C.; Maggio, L.G.D. Explainable AI for Machine Fault Diagnosis: Understanding Features' Contribution in Machine Learning Models for Industrial Condition Monitoring. *Appl. Sci.* **2023**, *13*, 2038. [CrossRef]
9. Tiboni, M.; Remino, C.; Bussola, R.; Amici, C. A review on vibration-based condition monitoring of rotating machinery. *Appl. Sci.* **2022**, *12*, 972. [CrossRef]

10. Cen, J.; Yang, Z.; Liu, X.; Xiong, J.; Chen, H. A review of data-driven machinery fault diagnosis using machine learning algorithms. *J. Vib. Eng. Technol.* **2022**, *10*, 2481–2507. [CrossRef]
11. Reddy, G.T.; Reddy, M.P.K.; Lakshmanna, K.; Kaluri, R.; Rajput, D.S.; Srivastava, G.; Baker, T. Analysis of dimensionality reduction techniques on big data. *IEEE Access* **2020**, *8*, 54776–54788. [CrossRef]
12. Jia, W.; Sun, M.; Lian, J.; Hou, S. Feature dimensionality reduction: A review. *Complex Intell. Syst.* **2022**, *8*, 2663–2693. [CrossRef]
13. Wang, H.; Ni, G.; Chen, J. Research on rolling bearing state health monitoring and life prediction based on PCA and internet of things with multi-sensor. *Measurement* **2020**, *157*, 107657. [CrossRef]
14. Wang, H.; Huang, H.; Yu, S.; Gu, W. Size and Location Diagnosis of Rolling Bearing Faults: An Approach of Kernel Principal Component Analysis and Deep Belief Network. *Int. J. Comput. Intell. Syst.* **2021**, *14*, 1672–1686. [CrossRef]
15. Zhong, K.; Han, M.; Qiu, T.; Han, B. Fault diagnosis of complex processes using sparse kernel local Fisher discriminant analysis. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 1581–1591. [CrossRef]
16. Tian, J.; Zhang, Y.; Zhang, F.; Ai, X.; Wang, Z. A novel intelligent method for inter-shaft bearing-fault diagnosis based on hierarchical permutation entropy and LLE-RF. *J. Vib. Control* **2022**, *1*, 10775463221134166. [CrossRef]
17. Jiang, L.; Xuan, J.; Shi, T. Feature extraction based on semi-supervised kernel Marginal Fisher analysis and its application in bearing fault diagnosis. *Mech. Syst. Signal Process.* **2013**, *1*, 113–126. [CrossRef]
18. Wang, W.; Aggarwal, V.; Aeron, S. Tensor train neighborhood preserving embedding. *IEEE Trans. Signal Process.* **2018**, *66*, 2724–2732. [CrossRef]
19. Ran, R.; Ren, Y.; Zhang, S.; Fang, B. A novel discriminant locality preserving projections method. *J. Math. Imaging Vis.* **2021**, *63*, 541–554. [CrossRef]
20. He, Y.L.; Zhao, Y.; Hu, X.; Yan, X.N.; Zhu, Q.X.; Xu, Y. Fault diagnosis using novel AdaBoost based discriminant locality preserving projection with resamples. *Eng. Appl. Artif. Intell.* **2020**, *91*, 103631. [CrossRef]
21. Yang, C.; Ma, S.; Han, Q. Unified discriminant manifold learning for rotating machinery fault diagnosis. *J. Intell. Manuf.* **2022**, *91*, 1–12. [CrossRef]
22. Zheng, H.; Wang, R.; Yin, J.; Li, Y.; Lu, H.; Xu, M. A new intelligent fault identification method based on transfer locality preserving projection for actual diagnosis scenario of rotating machinery. *Mech. Syst. Signal Process.* **2020**, *135*, 106344. [CrossRef]
23. Li, Y.; Lekamalage, C.K.L.; Liu, T.; Chen, P.A.; Huang, G.B. Learning representations with local and global geometries preserved for machine fault diagnosis. *IEEE Trans. Ind. Electron.* **2019**, *67*, 2360–2370. [CrossRef]
24. Fu, Y. Local coordinates and global structure preservation for fault detection and diagnosis. *Meas. Sci. Technol.* **2021**, *32*, 115111. [CrossRef]
25. Luo, L.; Bao, S.; Mao, J.; Tang, D. Nonlocal and local structure preserving projection and its application to fault detection. *Chemom. Intell. Lab. Syst.* **2016**, *157*, 177–188. [CrossRef]
26. Tang, Q.; Chai, Y.; Qu, J.; Fang, X. Industrial process monitoring based on Fisher discriminant global-local preserving projection. *J. Process Control* **2019**, *81*, 76–86. [CrossRef]
27. Li, Y.; Ma, F.; Ji, C.; Wang, J.; Sun, W. Fault Detection Method Based on Global-Local Marginal Discriminant Preserving Projection for Chemical Process. *Processes* **2022**, *10*, 122. [CrossRef]
28. Huang, Z.; Zhu, H.; Zhou, J.T.; Peng, X. Multiple marginal fisher analysis. *IEEE Trans. Ind. Electron.* **2018**, *66*, 9798–9807. [CrossRef]
29. Zhao, X.; Jia, M. Fault diagnosis of rolling bearing based on feature reduction with global-local margin Fisher analysis. *Neurocomputing* **2018**, *315*, 447–464. [CrossRef]
30. Su, S.; Zhu, G.; Zhu, Y. An orthogonal locality and globality dimensionality reduction method based on twin eigen decomposition. *IEEE Access* **2021**, *9*, 55714–55725. [CrossRef]
31. Zhang, S.; Lei, Y.; Zhang, C.; Hu, Y. Semi-supervised orthogonal discriminant projection for plant leaf classification. *Pattern Anal. Appl.* **2016**, *19*, 953–961. [CrossRef]
32. Shi, M.; Zhao, R.; Wu, Y.; He, T. Fault diagnosis of rotor based on local-global balanced orthogonal discriminant projection. *Measurement* **2021**, *168*, 108320. [CrossRef]
33. Chou, J.; Zhao, S.; Chen, Y.; Jing, L. Unsupervised double weighted graphs via good neighbours for dimension reduction of hyperspectral image. *Int. J. Remote Sens.* **2022**, *43*, 6152–6175. [CrossRef]
34. Yan, T.; Shen, S.L.; Zhou, A.; Chen, X. Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm. *J. Rock Mech. Geotech. Eng.* **2022**, *14*, 1292–1303. [CrossRef]
35. Liu, Y.; Gao, Q.; Han, J.; Wang, S. Euler sparse representation for image classification. *Proc. AAAI Conf. Artif. Intell.* **2018**, *32*, 3691–3697. [CrossRef]
36. Zhang, X.; Xiao, H.; Gao, R.; Zhang, H.; Wang, Y. K-nearest neighbors rule combining prototype selection and local feature weighting for classification. *Knowl. Based Syst.* **2022**, *243*, 108451. [CrossRef]
37. Su, W.; Wang, Y. Estimating the Gerber-Shiu function in Lévy insurance risk model by Fourier-cosine series expansion. *Mathematics* **2021**, *9*, 1402. [CrossRef]
38. Mekonen, B.D.; Salilew, G.A. Geometric Series on Fourier Cosine-Sine Transform. *J. Adv. Math. Comput. Sci.* **2018**, *28*, 1–4. [CrossRef]
39. Fitch, A.; Kadyrov, A.; Christmas, W.; Kittler, J. Fast robust correlation. *IEEE Trans. Image Process.* **2005**, *70*, 1063–1073. [CrossRef]

40. Peng, L.; Yao, Q. Least absolute deviations estimation for ARCH and GARCH Models. *Biometrika* **2003**, *90*, 967–975. [CrossRef]
41. Zhong, H.; Lv, Y.; Yuan, R.; Yang, D. Bearing fault diagnosis using transfer learning and self-attention ensemble lightweight convolutional neural network. *Neurocomputing* **2022**, *501*, 765–777. [CrossRef]
42. Yang, X.; Shao, H.; Zhong, X.; Cheng, J. Symplectic weighted sparse support matrix machine for gear fault diagnosis. *Measurement* **2021**, *168*, 108392. [CrossRef]
43. Liu, Y.; Hu, Z.; Zhang, Y. Bearing feature extraction using multi-structure locally linear embedding. *Neurocomputing* **2021**, *428*, 280–290. [CrossRef]
44. Altaf, M.; Akram, T.; Khan, M.A.; Iqbal, M.; Ch, M.M.I.; Hsu, C.H. A new statistical features based approach for bearing fault diagnosis using vibration signals. *Sensors* **2022**, *22*, 2012. [CrossRef] [PubMed]
45. Wang, S.; Ding, C.; Hsu, C.H.; Yang, F. Dimensionality reduction via preserving local information. *Future Generat. Comput. Syst.* **2020**, *108*, 967–975. [CrossRef]