electronics

MDPI

Article

Discriminative Projection with Flexible Marginal and Local Relationships Learning for Feature Extraction and Classification

Lin Li [1,*], Jian Zheng [2], Jialiang Xu [1] and Zuliang Lu [1]

[1] College of Mathematics and Statistics, Chongqing Sanxia University of Science and Technology, Chongqing 404000, China; zulianglu@sanxiau.edu.cn (Z.L.)
[2] School of Artificial Intelligence, Chongqing Technology and Business University, Chongqing 400067, China; zhengjian1@ctbu.edu.cn
* Correspondence: 20230028@sanxiau.edu.cn

Abstract

Feature extraction effectively captures meaningful information from raw signals or data to improve the accuracy and real-time performance of detection, classification, and prediction. The selection of an effective discriminant projection method is crucial for extracting discriminant features from the raw signals or data. However, some existing discriminative projection methods face two major limitations. First, the fixed graph structure adopted by some of the methods during feature extraction makes them susceptible to noise or outliers. Second, the predefined zero-one label matrix used by some of the methods to guide the model in learning the discriminative projection matrix leads to the loss of some intrinsic information inherent in the data. To overcome these limitations, we propose a framework called discriminative projection with flexible marginal and local relationship learning (DP-FRL). The key innovations of DP-FRL are primarily reflected in three aspects: (1) utilizing a flexible target matrix to guide the projection matrix in learning marginally structured relationships in the data; (2) employing permutation matrices in the local relationship regularization term reduces the dependence of the projection matrix on the order of features, effectively enhancing its ability to flexibly capture local relationships in the data; and (3) introducing a rank-one quadratic penalty to constrain the projection matrix, which helps learn more discriminative projections. To effectively solve the proposed framework model, we designed an iterative optimization algorithm. Finally, the classification experiments conducted on real-world datasets demonstrate that, compared to related methods, the features extracted by the DP-FRL method are more effective for classification tasks.

Keywords: supervised dimensionality reduction; feature extraction; discriminative projection; flexible relationship learning; classification

Check for updates

Academic Editor: George Angelos Papadopoulos

Received: 4 October 2026
Revised: 23 October 2026
Accepted: 30 October 2026
Published: 3 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The rapid advancement of intelligent sensing and imaging technologies has led to an unprecedented surge in high-dimensional data across domains such as computer vision, biometric recognition, and pattern analysis. Although such data is rich in information, the redundancy and noise present in the data pose serious challenges to storage capacity, computational efficiency, and downstream classification performance. Feature extraction has thus become an indispensable preprocessing step, aiming to project high-dimensional data into a low-dimensional subspace while preserving intrinsic geometric and discriminative structures [1–3]. Existing feature-extraction methods can be broadly categorized into unsupervised and supervised paradigms. The former, including principal component

analysis [1], locality-preserving projections [2], locally linear embedding [3], neighborhood preserving embedding [4], and isometric feature mapping [5], explore data structures without label information. The latter leverage class labels to learn discriminative subspaces, with linear discriminant analysis [6] being a representative example.

Among supervised feature-extraction methods, linear discriminant analysis (LDA) seeks a projection that maximizes between-class scatter while minimizing within-class scatter [6]. Although linear discriminant analysis is theoretically elegant and widely used, three limitations remain: first, the method assumes Gaussian distribution and cannot handle non-Gaussian data; second, the method is relatively sensitive to outliers; third, the method often faces the problem of small sample sizes. To address the first limitation, marginal Fisher analysis (MFA) [7,8] was proposed. In addition, graph embedding was introduced as a general framework that unifies many feature-extraction methods, including principal component analysis, locality-preserving projections, neighborhood preserving embedding, and isometric feature mapping. Within this framework, the quality of the affinity graph is crucial for preserving the geometric and discriminative structures of data. However, traditional graph construction methods, such as K-nearest neighbor (KNN) and ϵ-ball, build the graph in the original high-dimensional space, where Euclidean distances can be severely distorted by noisy or irrelevant features, leading to unreliable graphs. To address this issue, a strategy for adaptively constructing similarity matrices was proposed. Based on this strategy, local adaptive discriminant analysis and submanifold-preserving discriminant analysis (SPDA) [9] were developed. To address the second limitation, Wang et al. [10] developed a robust outlier elimination trace ratio linear discriminant analysis, which employs a binary indicator vector to completely remove outliers while retaining squared ℓ_2 distances. To address the third limitation, Orthogonal linear discriminant analysis [11] was proposed, which avoids the total scatter matrix from being non-singular. Recent efforts have attempted to enhance discriminant analysis from different perspectives. Wang et al. [12] proposed attribute graph adjusted trace ratio linear discriminant analysis (AGATRLDA), which constructs an attribute graph to encode inter-attribute correlations and adjusts the discriminant projection matrix accordingly. Huang et al. [13] proposed reverse neighborhood discriminant analysis (RNDA), which uses reverse nearest neighbors to characterize intrinsic data structures and fuses multi-granularity information with adaptive weighting. In addition, recent research on multi-level feature fusion has provided valuable insights into the use of complementary representations. For example, YOLO-TCS improves the performance of multi-scale traffic sign detection through a triple feature integration module and a dual-view attention mechanism [14]; while MSTF-PatchTST enhances load forecasting accuracy by adaptively fusing temporal and frequency domains [15]. These studies demonstrate that fusing complementary features can effectively capture complex structures.

To mitigate the influence of noisy or irrelevant features, adaptive discriminant analysis has emerged as a promising direction, where the affinity graph is learned jointly with the projection or embedding. Representative methods include locality adaptive discriminant analysis [16], submanifold-preserving discriminant analysis [9], and adaptive locality-preserving regression (ALPR) [17]. These methods dynamically adjust the affinity graph in the low-dimensional subspace, thereby improving robustness against noise and outliers. Nevertheless, many adaptive graph methods still model relationships using projected samples rather than the low-dimensional embedding, which may introduce errors when the manifold is nonlinear. Furthermore, the discrimination of projected samples is often not explicitly enhanced. To overcome these limitations, Li et al. [18] proposed discriminative projection learning with adaptive reversed graph embedding (DP-ARGE), which learns the graph dynamically in the low-dimensional subspace using the low-dimensional embedding rather than projected samples. A discriminative regularization term is designed using a

few pairwise marginal heterogeneous samples to enhance discrimination while keeping training efficient. An automatic strategy is developed to estimate the neighborhood size based on data density and similarity. The method is further extended to a semi-supervised version by introducing a neighborhood preserving regularization term.

Beyond discriminant analysis, discriminative least squares regression has emerged as another powerful paradigm for feature extraction and classification. Discriminative least squares regression (DLSR) improves inter-class separability by relaxing binary labels through the ϵ-dragging technique, which enlarges the margins between different classes [19]. However, discriminative least squares regression overlooks intra-class structure and is sensitive to noise. To address these shortcomings, Yang et al. [20] proposed structural enhanced robust discriminative least squares regression, which integrates projection distance regularization, local sparsity, and local low-rank constraints to strengthen intra-class compactness and suppress noise. Despite its promising performance, structural enhanced robust discriminative least squares regression suffers from high computational complexity, especially for large-scale data. Moreover, discriminant least-squares regression methods exhibit two limitations: (1) graph learning and projection learning are decoupled, and the fixed graph may inadequately capture the intrinsic manifold structure; (2) a predefined binary label matrix guides discriminant projection learning, potentially causing the loss of intrinsic data information.

To address these two limitations, namely, the susceptibility of fixed graph structures to noise and outliers during feature extraction, and the inherent loss of data information caused by the predefined binary label matrix in discriminative projection learning, a framework named discriminative projection with flexible marginal and local relationship learning (DP-FRL) is proposed. To address the first issue, a permutation matrix is introduced into the locality-preserving constraint to guide the projection matrix in flexibly learning local relationships within the data. To address the second issue, a flexible target matrix is utilized to guide the projection matrix in learning marginally structured relationships in the data, which helps preserve more intrinsic information that may be lost by the predefined zero-one label matrix. In addition, a rank-one quadratic penalty is imposed on the projection matrix to constrain the projection matrix, which helps learn more discriminative projections and avoid over-fitting. To effectively solve the proposed framework model, an iterative optimization algorithm is designed. Overall, the contributions of this paper can be divided into the following four parts.

1. A framework named discriminative projection with flexible marginal and local relationship learning (DP-FRL) is proposed. DP-FRL was found to effectively extract discriminative features from data to improve the accuracy of downstream classification tasks.
2. The marginal structures and local relationships flexibly learned from the data help guide the projection matrix in learning the discriminative information within the data.
3. A rank-one quadratic penalty is introduced to constrain the projection matrix, which helps learn more discriminative projection.
4. An iterative optimization algorithm is designed to efficiently integrate all components within the framework.

The remainder of this article is organized as follows. Section 2 reviews the discriminant analysis and related work. Section 3 presents the proposed DP-FRL model and optimization algorithm. Section 4 reports experimental and results. Section 5 concludes the article.

2. Related Work

2.1. Discriminant Projection with Graph Embedding

Linear discriminant analysis is a classical supervised dimensionality reduction method that seeks a projection maximizing the ratio of between-class scatter to within-class scat-

ter [6]. However, linear discriminant analysis can be integrated into the graph embedding framework. Specifically, a low-dimensional representation of a sample, x_i, can be obtained via a linear projection, P. Therefore, this framework can be written in the following form [7]:

$$\begin{aligned} &\min_{P} \frac{1}{2} \sum_{i,j}^{n} \|P^T x_i - P^T x_j\|_2^2 s_{ij} \\ &\text{s.t. } P^T P = Con \text{ or } P^T X D X^T P = Con. \end{aligned} \tag{1}$$

where Con is a constant matrix; D is used to avoid trivial solutions; and s_{ij} is typically computed in advance to represent similarity. Graph embedding provides a unified framework for dimensionality reduction, expressing methods such as locality-preserving projections, neighborhood preserving embedding, linear discriminant analysis, and MFA as graph-based learning problems [7]. To address the shortcoming of linear discriminant analysis in ignoring local manifold structure, Local fisher discriminant analysis combines fisher's discriminant analysis with locality-preserving projections [2] and simultaneously achieves between-class separability and within-class local structure preservation [21]. Local linear discriminant analysis captures local structure using only a small number of training samples [22]. To preserve the discriminative information of the data in the projection matrix, a strong marginal constraint was imposed on local sensitivity discriminant analysis [8]. Wang et al. [12] proposed AGATRLDA, which constructs an attribute graph to encode inter-attribute correlations and adjusts the discriminant projection matrix accordingly. By penalizing the difference between coefficients of similar attributes, AGATRLDA encourages correlated attributes to share similar weights, thereby improving discriminability. Huang et al. [13] proposed RNDA, which uses reverse nearest neighbors to characterize intrinsic data structures and fuses multi-granularity information with adaptive weighting.

2.2. Discriminant Projection with Adaptive Neighbors

In the framework (1), affinity graph quality is critical for preserving data geometry and discrimination. Yet traditional KNN and ϵ-ball methods build graphs in the original high-dimensional space, where noise and irrelevant features distort Euclidean distances, making graphs unreliable. To address this issue, an adaptive graph learning strategy has been proposed. Specifically, SPDA [9] is a classic example, with its objective function expressed as follows:

$$\begin{aligned} &\min_{P,s} \sum_{k=1}^{c} \frac{1}{n_i} \sum_{i,j=1}^{n_i} \|P^T x_i^k - P^T x_j^k\|_2^2 s_{ij}^k \\ &\text{s.t.} P^T S_t P = I, \mathbf{1}^T s_i^k = h, s_{ij}^k \in \{0,1\}. \end{aligned} \tag{2}$$

Recently, adaptive discriminant analysis, which combines affinity graphs with projections or embeddings for joint learning, has been extensively studied and applied. Representative methods include locality adaptive discriminant analysis [16] and SPDA [9]. These methods dynamically adjust the graph structure in a low-dimensional subspace, thereby improving robustness against noise and outliers. However, many of these methods still use projected samples rather than low-dimensional embeddings to model relationships, which may lead to errors in nonlinear manifolds. To address this issue, Li et al. [18] proposed a discriminative projection learning method with adaptive reversed graph embedding, which utilizes low-dimensional embeddings (rather than projected samples) to dynamically learn graph structures in a low-dimensional subspace, thereby enhancing discriminative power.

2.3. Discriminant Projection with Discriminative Regression

By minimizing the distance between the training sample set and the label space, least squares regression enables better learning of discriminative features from data. To further

improve such learning, several least squares regression-based feature-extraction methods have been developed. Some of these methods share a unified objective function, which can be expressed as follows:

$$\min_{W} \|Y - \psi(W)X\|_F^2 + \alpha\varphi(W) + \beta\phi(WX) \tag{3}$$

where α and β are trade-off parameters. Y is typically a strictly zero-one label matrix. $\psi(W)$ denotes a certain deformation with respect to W, while $\varphi(W)$ and $\phi(WX)$ act as regularization terms. To obtain a globally optimal discriminant projection, discriminative sparse subspace learning with manifold regularization (DSSLMR) [23] incorporates sub-manifold and multi-manifold information into $\phi(WX)$. To mitigate the information loss caused by directly mapping data from feature space to label space, discriminative approximate regression projection [24] employs a combination of a transfer matrix and a feature extraction matrix in place of $\phi(W)$. To endow the projection matrix with the ability to select important features, sparse representation techniques are applied to $\varphi(W)$. For instance, capped norm based discriminant robust regression [25] constrains W using the $L_{2,1}$ norm, i.e., $\varphi(W) = \|W\|_{2,1}$, thereby extracting the most important features. To extract the intra-class structures within the data, Yang et al. [20] proposed structural enhanced robust discriminant least squares regression (SERDLSR), which integrates projection distance regularization, local sparsity, and local low-rank constraints to strengthen intra-class compactness and suppress noise. To learn a discriminative projection that facilitates classification, the source data is projected by pairwise relations oriented discriminative regression (PRDR) [26] into a latent space rather than a label space, and this latent space preserves both the pairwise relationships between labels and the relationships among instances within classes. However, most methods based on least-squares regression fix the number of features extracted from the projection matrix to the number of classes c, which limits the flexibility to change the dimensionality of the extracted features.

3. Proposed Method

This section presents the detailed steps of the proposed method, and its concept is illustrated in Figure 1.

3.1. Model of the Method

As presented in the related work, feature-extraction methods based on adaptive neighbors and discriminative regression not only effectively capture local structural relationships in the data but also enhance the ability of the projection matrix to extract the intrinsic discriminative features of the data [9]. In some feature-extraction methods based on discriminant regression, the target vector is a label vector, as indicated by Y in Equation (3). However, using a binary vector as the regression target is too restrictive and does not provide sufficient flexibility for the regression fitting problem. Actually, the binary vector constraint is very difficult to satisfy completely, which may result in high regression error [19]. To address this issue, a more flexible target matrix, featuring a larger margin between the error class and the correct class, is used to enhance the discriminative power of the projection [27,28]. Inspired by adaptive neighbors and discriminative regression methods, for the input data, $X \in \mathbb{R}^{m\times n}$, by combining sub-manifold learning and discriminative regression, the corresponding objective function can be built as follows:

$$\begin{aligned} &\min_{P,S,T} \|T - P^T X\|_F^2 + \alpha \sum_{k=1}^{c} \sum_{i,j=1}^{n_k} \|P^T x_i^k - P^T x_j^k\|_2^2 (s_{ij}^k)^2, \\ &\text{s.t. } T_{i,l_i} - \max_{j\neq l_i} T_{i,j} \geq 1, 0 \leq s_{ij}^k, \sum_{j=1}^{h_1} s_{ij}^k = 1, \end{aligned} \tag{4}$$

where α denotes a trade-off parameter, and s_{ij}^k measures the similarity between two samples in the reduced dimensional feature space. $T \in \mathbb{R}^{c \times n}$ is the target matrix, and $l_i \in \{1, \ldots, c\}$ denotes the index of the true label for the ith sample. The constraints $s_{ij}^k \geq 0$ and $\sum_{j=1}^{h_1} s_{ij}^k = 1$ are introduced to preclude trivial solutions, i.e., to avoid the extreme case in which any entry of s_{ij}^k is zero.

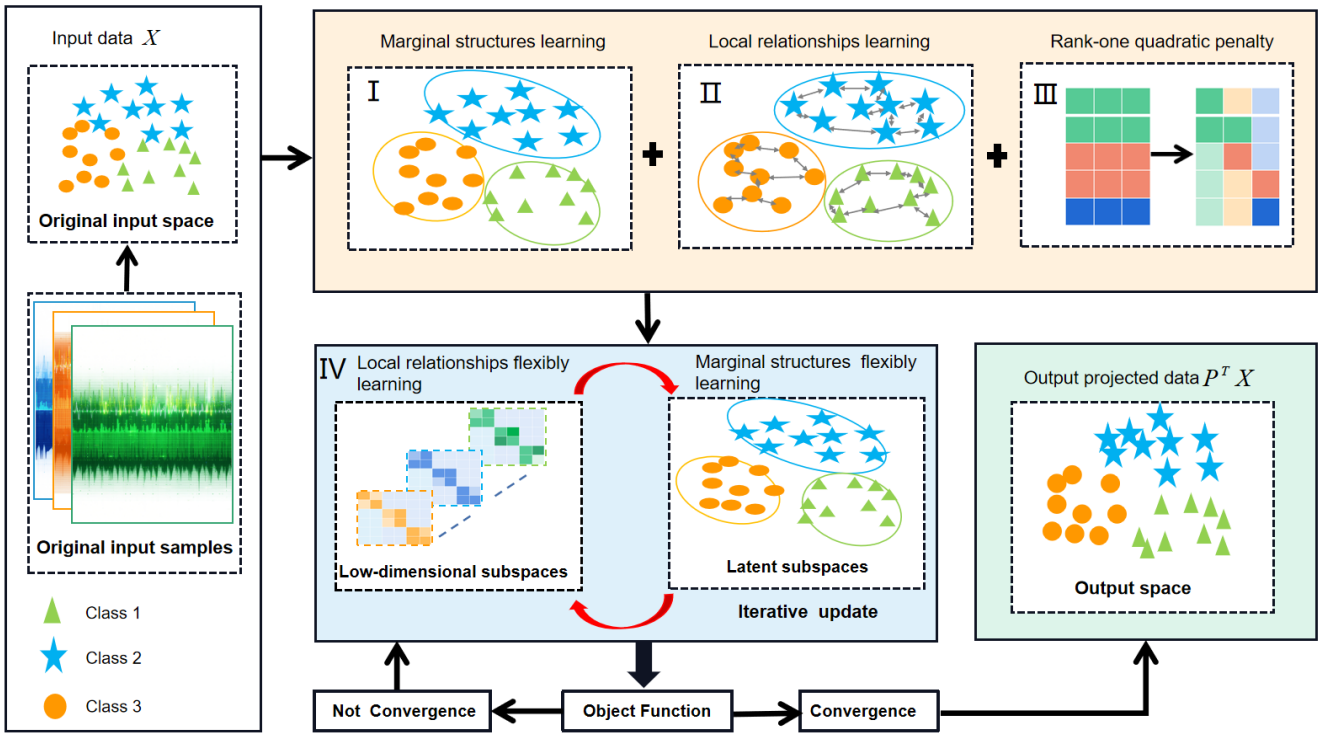


Figure 1. Abstract conceptual overview of the DP-FRL framework. (I) The marginal relationships among data classes can be learned using discriminant regression based on a flexible target matrix. (II) The local relationships in the data can be preserved by utilizing sub-manifold learning. (III) The introduction of a rank-one quadratic penalty constrains the projection matrix, encouraging the model to learn more discriminative projection. (IV) The discriminative projection and similarity matrix are optimized in an alternating manner. During this process, marginal structures and local structural relationships jointly guide the model in extracting discriminative information from the data.

Since the projection matrix $P \in \mathbb{R}^{m \times c}$ in the objective function (4) can extract only c features, this model may extract redundant information when dealing with data whose intrinsic dimension is smaller than the number of classes c. To address this issue, we introduce a projection matrix $R \in \mathbb{R}^{c \times d}$ to enable the objective function (4) to extract discriminative features from the data more flexibly. Then, for the projection matrix $P \in \mathbb{R}^{m \times d}$, the objective function (4) can be rewritten as follows:

$$\begin{aligned} &\min_{P,S,T} \|T - RP^T X\|_F^2 + \alpha \sum_{k=1}^{c} \sum_{i,j=1}^{n_k} \|P^T x_i^k - P^T x_j^k\|_2^2 (s_{ij}^k)^2, \\ &\text{s.t. } T_{i,l_i} - \max_{j \neq l_i} T_{i,j} \geq 1, 0 \leq s_{ij}^k, \sum_{j=1}^{h_1} s_{ij}^k = 1, R^T R = I, \end{aligned} \tag{5}$$

where I is an identity matrix of the appropriate size. Note that, to satisfy the constraint $R^T R = I$, we set $d \leq c$, that is, the dimension of the feature extraction is less than or equal to the number of categories c. Data acquired in real-world applications typically exhibit high dimensionality, along with many redundant features and noise. These redundant features and noise residing in the original data are not only useless but potentially harmful to model training. To mitigate this problem, we additionally enforce a rank-one quadratic penalty

constraint on the projection, that is, $\|P\|_G = \sum_{j=1}^{d}\left(\sum_{i=1}^{m} P_{ij}\right)^2$. Let f_i be a low-dimensional representation of the high-dimensional data, x_i. Then, we expand $f_i = P^T x_i$ as follows:

$$\begin{aligned} f_{1,i} &= p_{1,1}x_{1,i} + p_{2,1}x_{2,i} + \cdots + p_{m,1}x_{m,i} \\ f_{2,i} &= p_{1,2}x_{1,i} + p_{2,2}x_{2,i} + \cdots + p_{m,2}x_{m,i} \\ &\vdots \\ f_{m,i} &= p_{1,m}x_{1,i} + p_{2,m}x_{2,i} + \cdots + p_{m,m}x_{m,i} \end{aligned} \tag{6}$$

Based on the formulation (6), the j-th projected feature can be explicitly expanded as:

$$f_{j,i} = \sum_{k=1}^{m} p_{k,j}x_{k,i} = p_{1,j}x_{1,i} + p_{2,j}x_{2,i} + \cdots + p_{m,j}x_{m,i}. \tag{7}$$

By imposing the G-norm on the projection matrix P with $\min \|P\|_G$, we penalize the sum of the weights in each column, i.e., $(p_{1,j} + p_{2,j} + \cdots + p_{m,j})^2$. This design yields threefold benefits: (1) it prevents any single projected dimension f_j from dominating the representation by constraining the total energy of its corresponding weight vector; (2) it naturally encourages the weights within the same column to have opposite signs (e.g., $p_{1,j}$ and $p_{2,j}$ canceling each other out), thereby constructing contrastive features to enhance discriminability; (3) it induces structural sparsity by suppressing redundant projection directions entirely, yielding a more interpretable subspace. Then, we plug $\|P\|_G$ into (5), leading to

$$\begin{aligned} &\min_{P,S,R,T} \|T - RP^T X\|_F^2 + \alpha \sum_{k=1}^{c} \sum_{i,j=1}^{n_k} \|P^T x_i^k - P^T x_j^k\|_2^2 (s_{ij}^k)^2 + \beta \|P\|_G, \\ &\text{s.t. } T_{i,l_i} - \max_{j \neq l_i} T_{i,j} \geq 1, 0 \leq s_{ij}^k, \sum_{j=1}^{h_1} s_{ij}^k = 1, R^T R = I, \end{aligned} \tag{8}$$

where β denotes a trade-off parameter. Additionally, the order of redundant and noisy features relative to distinguishing features in the data may affect the similarity relationships among samples learned by the similarity matrix. To eliminate the bias caused by feature order, we introduce a permutation matrix into model (8). Finally, the framework of the discriminative projection with flexible marginal and local relationship learning (DP-FRL) can be formed:

$$\begin{aligned} &\min_{P,S,R,T,W} \|T - RP^T X\|_F^2 + \alpha \sum_{k=1}^{c} \sum_{i,j=1}^{n_k} \|P^T W x_i^k - P^T W x_j^k\|_2^2 (s_{ij}^k)^2 + \beta \|P\|_G, \\ &\text{s.t. } T_{i,l_i} - \max_{j \neq l_i} T_{i,j} \geq 1, 0 \leq s_{ij}^k, \sum_{j=1}^{h_1} s_{ij}^k = 1, R^T R = I, W^T W = I, W \geq 0, \end{aligned} \tag{9}$$

where the constraints $W^T W = I$ and $W \geq 0$ jointly guarantee that W, as a permutation matrix, rearranges the features of the input samples [29]. The conceptual of DP-FRL is presented in Figure 1. The first term in the framework (9) is used to flexibly learn the marginal structural relationships between classes of data. The objective is to expand the margin of the regression target and enhance the separation between classes. The second component in this framework (9) is used to flexibly learn local relationships among data. The third term in the framework (9) imposes a rank-one quadratic penalty constraint on the projection matrix, which primarily guides the learned projections to be more discriminative. The introduction of the transformation matrix W helps reduce the impact of feature order on the model's ability to capture similarity relationships among the training data. The components of the framework are tightly interconnected through an iterative optimization algorithm, described in the next subsection, which aims to learn an efficient discriminative projection.

3.2. Solution to the Proposed Method

In this subsection, the solution to the problem (9) is computed. Based on the update rules derived from the theory of alternating optimization [9], the update strategy for variable R is first presented under the assumption that all other variables are fixed. The update strategies for variables T, W, P, and S are then derived in the same manner.

R-step: By fixing the variables T, W, P, and S, R can be solved as follows:

$$\min_{R^TR=I} \|T - RP^TX\|_F^2. \tag{10}$$

Under the orthogonal constraints on variable R, the subproblem (10) can be transformed into a Procrustes problem [30].

$$\max_{R^TR=I} Tr(R^TTX^TP). \tag{11}$$

Therefore, a closed-form solution to Problem (6) can be expressed as follows:

$$R = JV^T. \tag{12}$$

where J, V are both orthogonal matrices. The relationship between J and V is established through the singular value decomposition of matrix TX^TP, that is, $SVD(TX^TP) = JZV^T$. Here, the operator $SVD(\cdot)$ denotes the singular value decomposition operator.

T-step: By fixing the variables R, W, P, and S, we can solve for T by solving the following constrained quadratic programming problem:

$$\min_{T_{i,l_i} - \max_{j \neq l_i} T_{i,j} \geq 1} \|T - RP^TX\|_F^2. \tag{13}$$

To solve Problem (13), the problem is decomposed into the following subproblems:

$$\min_{T_{i,l_i} - \max_{j \neq l_i} T_{i,j} \geq 1} \|T - Q\|_2^2. \tag{14}$$

where $Q = RP^TX$, and t_i and q_i are the i-th elements of T and Q, respectively. Then, the subproblems (14) can be solved using the following Theorem.

Theorem 1. *The optimal solution of problem*

$$\min_{t_h - \max_{i \neq h} t_i} \|t - q\|_2^2,$$

is

$$t_i = \begin{cases} q_i + \Delta, & \text{if } i = h, \\ q_i + \min(\Delta - v_i, 0), & \text{otherwise,} \end{cases}$$

where the vector $q = [q_1, q_2, \ldots, q_n]$, $v_i = 1 + q_i - q_h, \Delta = \frac{\sum_{i \neq h} v_i \Upsilon(\Gamma'(v_i), 0)}{1 + \sum_{i \neq h} \Upsilon(\Gamma'(v_i), 0)}$. $\Upsilon(\cdot)$ *is a logical function. The specific definition is* $\Upsilon(a, b) = \begin{cases} 1, & \text{if } a > b \\ 0, & \text{otherwise} \end{cases}$. *The function* $\Gamma'(x) = 2\left(x + \sum_{i \neq h} \min(x - v_i, 0)\right)$.

Proof of Theorem 1. This proof process is similar to that for $C = 1$ in Algorithm 1 in the reference [28]. □

W-step: By fixing the variables R, T, P, and S, we can calculate W by the following subproblem:

$$\begin{aligned} &\min_{W} \sum_{k=1}^{c} \sum_{i,j=1}^{n_k} \|P^TWx_i^k - P^TWx_j^k\|_2^2 (s_{ij}^k)^2, \\ &\text{s.t. } W^TW = I, W \geq 0. \end{aligned} \tag{15}$$

The Lagrangian function of the problem (15) is

$$\mathcal{L}(W, \eta, \Lambda) = Tr(P^T WXLX^T W^T P) + \frac{\eta}{2}\|W^T W - I\|_F^2 + Tr(\Lambda W^T). \tag{16}$$

where η_i is the Lagrangian multiplier. The Laplacian matrix L is defined as $L = B - S$, where the B is the degree matrix whose diagonal element $b_{ii} = \sum_j s_{ij}^k$. Taking the derivative of Equation (16) with regard to W and setting it to zero, we can obtain:

$$W \leftarrow W \odot \frac{W}{PP^T WXLX^T + \eta WW^T W}, \tag{17}$$

where $\odot$ is matrix dot product.

P-step: By fixing R, T, W, and S, and updating P by the following problem:

$$\min_P \|T - RP^T X\|_F^2 + \beta\|P\|_G + \alpha \sum_{k=1}^{c} \sum_{i,j=1}^{n_k} \|P^T W x_i^k - P^T W x_j^k\|_2^2 (s_{ij}^k)^2. \tag{18}$$

To solve subproblem (18), all terms in the problem (18) should be transformed into trace form. Under the condition $R^T R = I$, the first term in subproblem (18) can be transformed into the following trace form:

$$\|T - RP^T X\|_F^2 = Tr\left(P^T XX^T P\right) - 2Tr\left(R^T TX^T P\right) + Tr\left(TT^T\right), \tag{19}$$

and the second term in subproblem (18) can be transformed into the following trace form:

$$\|P\|_G = \sum_{j=1}^{d} \left(\sum_{i=1}^{m} P_{ij} \right)^2 = \mathrm{Tr}\left(P^T \mathbf{1}\mathbf{1}^T P\right). \tag{20}$$

where $\mathbf{1}$ represents a vector in which all elements are 1. Then, sub-optimization problem (18) can be transformed into

$$\min_P Tr(P^T(\beta \mathbf{1}\mathbf{1}^T + \alpha WXLX^T W^T + XX^T)P) - 2Tr\left(R^T TX^T P\right). \tag{21}$$

The partial derivative of (21) with respect to P is given as:

$$2\left(\beta \mathbf{1}\mathbf{1}^T + \alpha WXLX^T W^T + XX^T\right)P - 2XT^T R = 0.$$

Thus, the solution of P can be computed as:

$$P = \left(\beta \mathbf{1}\mathbf{1}^T + \alpha WXLX^T W^T + XX^T\right)^{-1} XT^T R. \tag{22}$$

S-step: By fixing R, T, W, and P, we can update S by the following sub-optimization problem:

$$\begin{aligned} &\min_S \sum_{k=1}^{c} \sum_{i,j=1}^{n^k} \|P^T W(x_i^k - x_j^k)\|_2^2 (s_{ij}^k)^2 \\ &\text{s.t. } s_{ij}^k \geq 0, \sum_{j=1}^{h_1} s_{ij}^k = 1. \end{aligned} \tag{23}$$

Since s_{ij}^k does not depend on the class, we can solve it via the sub-optimization problem presented below:

$$\begin{aligned} &\min_{S^k} \sum_{i,j=1}^{n_k} \|P^T W(x_i^k - x_j^k)\|_2^2 (s_{ij}^k)^2 \\ &\text{s.t. } s_{ij}^k \geq 0, \sum_{j=1}^{h_1} s_{ij}^k = 1, \end{aligned} \tag{24}$$

where S^k represents the similarity matrix of samples in the k-th class. For brevity, we define $\xi_{ij}^k = \|P^T W(x_i^k - x_j^k)\|_2^2$. Because of the constraint $\sum_{j=1}^{h_1} s_{ij}^k = 1$, only the similarities between x_i^k

and its h_1 nearest neighbors need to be updated. We therefore set all entries of $s_i^k \in \mathrm{R}^{n_i \times 1}$ to zero, except those corresponding to the h_1-nearest neighbors of x_i^k. As a result, problem (24) can be transformed into the following equivalent form:

$$\min_{S^k} \sum_{i=1}^{n_k} \xi_i^k (s_i^k)^2 \quad \text{s.t. } \mathbf{1}^T (s_i^k) = 1, s_{ij}^k \geq 0. \tag{25}$$

The solution to problem (25) can be obtained in a manner similar to [16]. Specifically, for each element in S^k, the result is given as follows:

$$s_{ij}^k = \begin{cases} \frac{1}{\xi_{ij}^k} (\sum_{e=1, e \neq j}^{h_1} \frac{1}{\xi_{ie}^k})^{-1}. & \text{if} \quad x_i^k \in N(x_j^k) \\ 0, & \text{otherwise.} \end{cases} \tag{26}$$

where $N(x_j^k)$ is h_1-nearest neighbor set of x_j^k. Algorithm 1 outlines the detailed optimization procedure for the framework in problem (9). Based on the update scheme of Algorithm 1, namely, the variables R, T, W, P, and S in problem (9) are updated alternately while the remaining ones are held fixed, both the similarity matrix S and the target matrix T in problem (9) are produced in a dynamic manner. More concretely, once Algorithm 1 has reached convergence, S and T are updated dynamically through Equation (26) and Theorem 1, respectively.

Algorithm 1 DP-FRL (Solving Problem (9))

Input: Data matrix $X \in \mathbb{R}^{m \times n}$, class label $Y \in \mathbb{R}^{c \times n}$, reduced dimension d, parameters α, β, h_1.

Initialization: $W = rand(m, m) \geq 0$ with orthogonal, $P \in \mathbb{R}^{m \times d}$ with PCA, $S \in \mathbb{R}^{n \times n}$ by using (26) with $\xi_{ij}^k = \|x_i^k - x_j^k\|_2^2$;

while not converged **do**

1: Update R by using (12);
2: Update T by using Theorem 1;
3: Update W by using (17);
4: Update P by using (22);
5: Update S by using (26).

end while

Output: P, S.

3.3. Convergence Analysis

In this section, we prove the convergence of Algorithm 1 when applied to problem (9), as stated in the theorem below.

Theorem 2. *Algorithm 1 is employed to solve problem (9), throughout the iterative process, the objective function value exhibits a monotonic decrease at each step.*

Proof. To streamline the presentation, let the objective function value of problem (9) at iteration l be given by:

$$\begin{aligned} \mathcal{F}(R^l, T^l, W^l, P^l, S^l) =& \|T^l - R^l (P^l)^T X\|_F^2 + \beta \|P^l\|_G \\ &+ \alpha \sum_{k=1}^{c} \sum_{i,j=1}^{n_k} \|(P^l)^T W^l x_i^k - (P^l)^T W^l x_j^k\|_2^2 ((s_{ij}^k)^l)^2. \end{aligned} \tag{27}$$

In the R-step, with T^l, W^l, P^l, and S^l held fixed, subproblem (10) implies the following inequality.

$$\begin{aligned} &\mathcal{F}(R^{l+1}, T^{l}, W^{l}, P^{l}, S^{l}) \\ &\leq \mathcal{F}(R^{l}, T^{l}, W^{l}, P^{l}, S^{l}). \end{aligned} \tag{28}$$

In the T-step, with P^{l+1}, W^{l}, P^{l}, and S^{l} fixed, the following inequality can be derived.

$$\begin{aligned} &\mathcal{F}(R^{l+1}, T^{l+1}, W^{l}, P^{l}, S^{l}) \\ &\leq \mathcal{F}(R^{l+1}, T^{l}, W^{l}, P^{l}, S^{l}). \end{aligned} \tag{29}$$

In the W-step, if R^{l+1}, T^{l+1}, P^{l}, and S^{l} are fixed, we can get:

$$\begin{aligned} &\mathcal{F}(R^{l+1}, T^{l+1}, W^{l+1}, P^{l}, S^{l}) \\ &\leq \mathcal{F}(R^{l+1}, T^{l+1}, W^{l}, P^{l}, S^{l}). \end{aligned} \tag{30}$$

In the P-step, if R^{l+1}, T^{l+1}, W^{l+1}, and S^{l} are fixed, we can obtain:

$$\begin{aligned} &\mathcal{F}(R^{l+1}, T^{l+1}, W^{l+1}, P^{l+1}, S^{l}) \\ &\leq \mathcal{F}(R^{l+1}, T^{l+1}, W^{l+1}, P^{l}, S^{l}). \end{aligned} \tag{31}$$

In the S-step, if R^{l+1}, T^{l+1}, W^{l+1}, and P^{l+1} are fixed, we can obtain:

$$\begin{aligned} &\mathcal{F}(R^{l+1}, T^{l+1}, W^{l+1}, P^{l+1}, S^{l+1}) \\ &\leq \mathcal{F}(R^{l+1}, T^{l+1}, W^{l+1}, P^{l+1}, S^{l}). \end{aligned} \tag{32}$$

Finally, it follows from Equations (27) to (32) that the following inequality holds:

$$\begin{aligned} &\mathcal{F}(R^{l+1}, T^{l+1}, W^{l+1}, P^{l+1}, S^{l+1}) \\ &\leq \mathcal{F}(R^{l}, T^{l}, W^{l}, P^{l}, S^{l}). \end{aligned} \tag{33}$$

Therefore, at each iteration, the objective function value of (9) decreases monotonically. Thus, Theorem 2 is proved. □

Furthermore, the objective function in problem (9) is lower-bounded. In combination with Theorem 2, the objective function values of proposed method is convergent. Furthermore, Section 4.6 provides additional convergence analysis experiments to verify the convergence of the objective function values. One should note that the combination of monotonically decreasing and bounded properties only guarantees the convergence of the objective function values, but does not guarantee that the algorithm converges to a local optimal solution.

3.4. Computational Time Complexity

This section focuses on evaluating the time complexity of Algorithm 1. The overall computational expense of the proposed method is mainly attributed to the five steps performed inside the loop. Neglecting the lower-order cost of matrix addition and multiplication, the runtime of the proposed method is dominated by two operations: the matrix inversion in the P-step and the singular value decomposition in the R-step. Specifically, for the matrix $(\mathbf{1}\mathbf{1}^T + \alpha WXLX^TW^T + XX^T) \in \mathbb{R}^{m\times m}$, its inverse can be computed in $O(m^3)$. For the matrix $TX^TP \in \mathbb{R}^{c\times d}$, computing its SVD requires $O(d^3)$. Since these operations are repeated at each iteration, the complexity for the proposed method to reach convergence after τ iterations is $O(\tau(mn^2 + m^3 + d^3))$.

4. Experiments

This subsection presents a series of experiments designed to validate the effectiveness of DP-FRL for feature extraction and classification. These experiments aim to explore the following five aspects: (1) to evaluate whether the features extracted by DP-FRL can improve the performance of upstream and downstream classification tasks on the original datasets; (2) to assess the robustness of DP-FRL on noisy datasets; (3) to further evaluate the effectiveness of this method compared to discriminant regression methods; (4) to validate the effectiveness of each component in the DP-FRL framework through ablation experiments; and (5) to examine the parameter sensitivity of DP-FRL and the convergence of its iterative optimization algorithm. In the classification experiment, features are extracted from the test data X_{test} by the trained projection matrix P, i.e., $P^T X_{test}$, and the extracted features are then input to the downstream classifier. To assess classification accuracy, we employed the 1-nearest neighbor (1-NN), Naive Bayes (NB), and support vector machine (SVM) classifiers. These classifiers were implemented using the MATLAB functions fitcknn, fitcnb, and fitcecoc, respectively, with their default settings. Here, classification accuracy (Acc (%)) is adopted as the commonly used evaluation metric, which is defined as:

$$Acc(\%) = \frac{N_{correct}}{N_{total}} \times 100\%,$$

where $N_{correct}$ denotes the number of correctly classified samples and N_{total} denotes the total number of samples. In addition, the macro-F1 [31] metric is also used to evaluate the performance of the classification. All experiments were conducted in MATLAB R2019a on a laptop equipped with an Intel Core i5-8250U 1.6 GHz CPU and Windows 10.

4.1. Comparison Methods

We conducted a comparative evaluation of our proposed method against six relevant feature-extraction methods, namely MFA [7], SPDA [9], AGATRLDA [12], RNDA [13], PRDR [26], and DSSLMR [23]. In experiment 4.4, we used the following methods as comparison benchmarks: raw features, LDA [6], DP-RGE [18], and discriminative regression-based methods. The discriminative regression-based methods include ALPR [17], marginally structured representation learning (MSRL) [28], Re-targeted Least Squares Regression (ReLSR) [27], and DLSR [19]. The classification results of the raw features are denoted as the baseline for comparison with the aforementioned methods. A brief overview of these baseline methods is provided below. In MFA, a similarity graph is constructed to maintain connectivity among samples within a class, and a penalty graph is employed to ensure separation between classes. In SPDA, weights are assigned to data points according to their nearest-neighbor relationships within each class. AGATRLDA is designed such that an attribute graph is constructed to encode inter-attribute correlations and the discriminant projection matrix is adjusted accordingly. RNDA is designed such that reverse nearest neighbors are used to characterize intrinsic data structures and multi-granularity information is fused with adaptive weighting. In PRDR, the source data is regressed into a latent space rather than the label space, and pairwise label relations as well as intra-class instance relations are simultaneously preserved in this latent space. DSSL-MR is formulated as a supervised linear sparse subspace learning method, in which manifold regularization is incorporated and the $L_{2,1}$-norm is imposed for discriminative feature selection. The number of nearest neighbors for MFA, PRDR, and SPDA is selected from the set of candidate values $[4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15]$. All other parameters of the compared methods are kept at the default values specified in their respective papers.

4.2. Datasets

This subsection briefly introduces the seven benchmark datasets used in the experiments, with their basic information summarized in Table 1. In these datasets, BreastCancer and Isolet are obtained from the UCI repository (https://archive.ics.uci.edu/, accessed on 10 September 2026). AR, YaleB, and ORL are face datasets composed of 32×32 pixel images, which can be downloaded from the website (http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html, accessed on 10 September 2026). A popular subset of the USPS (http://www.cad.zju.edu.cn/home/dengcai/Data/MLData.html, accessed on 1 September 2026) handwritten digit database is adopted. To evaluate robustness to noise, random noise is added to some samples to simulate environmental interference. The tr11 dataset (https://pages.nist.gov/trec-browser/, accessed on 7 September 2026) is derived from the TREC-5 corpus. To avoid the singularity problem, principal component analysis [32] is applied to the image datasets (i.e., ORL, AR, YaleB, USPS, Tr11, and Isolet), retaining 98% of the energy of the original data.

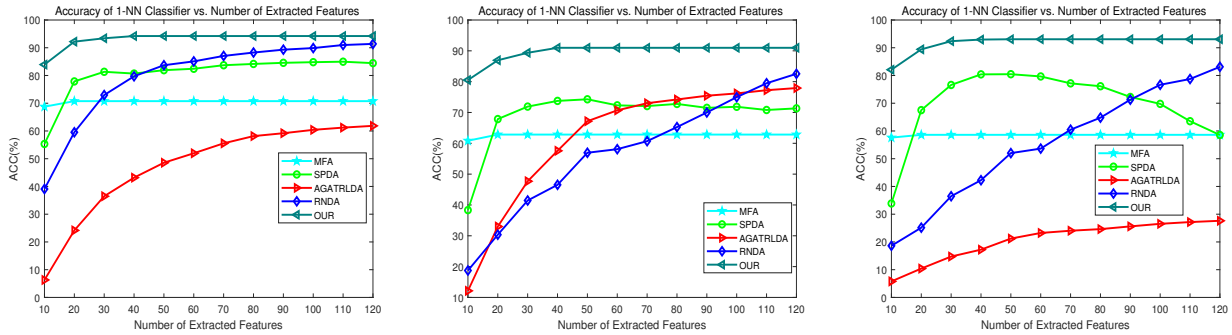
Table 1. Attributes of the experimental datasets.

No.	Dataset	# Samples	# Features	# Classes
1	BreastCancer	569	30	2
2	ORL	400	1024	40
3	USPS	9298	256	10
4	AR	2600	1024	100
5	Isolet	1560	617	26
6	Tr11	414	6429	9
7	YaleB	2414	1024	38

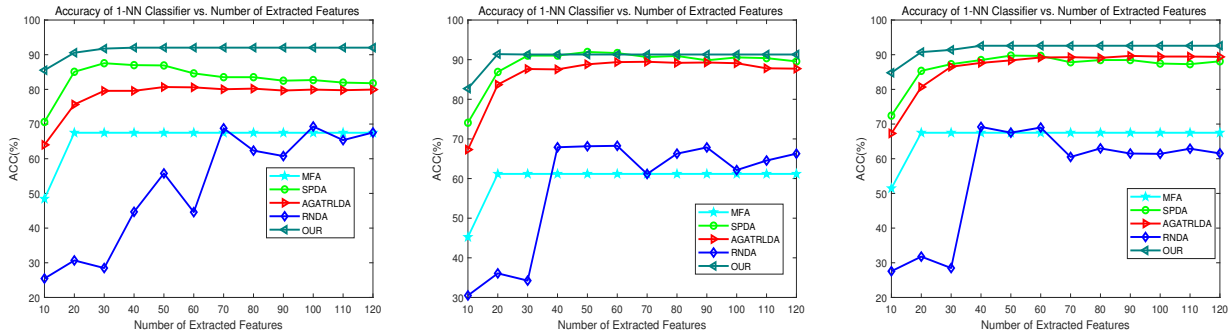
4.3. Classification Experiments on the Original Datasets

4.3.1. Experimental Setup

Comparative classification experiments are performed on the proposed DP-FRL method and the comparison methods to verify whether the extracted features improve downstream classification performance. For the proposed DP-FRL method, the nearest neighbor parameters (e.g., h_1) are simply selected from the candidate set $[4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15]$. Subsequently, a grid search is adopted to tune α and β in DP-FRL. The recommended range for parameters α and β is $[0.001, 0.01, 0.1, 1, 10, 100]$. One should note that the number of features extracted by methods PRDR and DSSL-MR is fixed at c. Figure 2 illustrates how classification accuracy varies with the number of extracted features for the compared methods. Note that the classification results for the proposed method that exceed those of class c are all set to the classification results obtained from feature c. In all experiments, the feature dimension of the proposed method was set to c. Next, each dataset was divided into a training set and a test set according to the following scheme: (1) For the ORL, Tr11, and AR datasets, because the number of samples in each class is limited, 6 samples were randomly selected from each class and assigned to the training set, while the remaining samples constituted the test set. (2) For the USPS, Isolet, and YaleB datasets, since the number of samples per class is relatively large, 18 samples were randomly selected from each class for the training set, and the remaining samples were used for the test set. (3) For the remaining datasets, 30 samples were randomly selected from each class for the training set, and the remaining samples were used for testing. Finally, the classification experiment was repeated 10 times by employing the optimal combination of parameters. The mean and standard deviation of the classification accuracy were then calculated and summarized in Table 2. One should note that the value with the highest classification accuracy is shown in bold, and the value with the second-highest classification accuracy is underlined.



(**a**) Results of the classification task on the YaleB dataset.



(**b**) Results of the classification task on the Isolet dataset.

Figure 2. Effect of the number of extracted feature dimensions (d) on the classification accuracy for YaleB and Isolet datasets. (**a**) Acc (%) of all classifiers on the YaleB dataset. (**b**) Acc (%) of all classifiers on the Isolet dataset.

Table 2. Acc (%) (mean ± standard deviation) achieved by related methods on the seven datasets.

Dataset	Classifier	MFA	SPDA	AGATR-LDA	RNDA	PRDR	DSSLMR	OUR (DP-FRL)
BreastCancer	1-NN	89.72 ± 1.81	91.36 ± 2.16	89.97 ± 1.87	90.25 ± 3.61	89.38 ± 2.47	91.79 ± 1.49	**93.24** ± 1.89
	NB	90.69 ± 1.65	90.37 ± 8.97	87.84 ± 2.62	91.55 ± 1.30	89.53 ± 0.48	93.20 ± 0.94	**93.88** ± 2.02
	SVM	89.81 ± 2.55	91.74 ± 1.47	86.84 ± 0.66	90.86 ± 1.48	85.20 ± 1.83	84.49 ± 0.92	**92.77** ± 0.36
ORL	1-NN	73.81 ± 2.83	91.50 ± 1.58	81.25 ± 1.50	63.13 ± 5.38	78.63 ± 2.50	93.31 ± 1.19	**93.88** ± 1.50
	NB	60.69 ± 2.80	87.44 ± 1.69	76.06 ± 1.35	57.38 ± 6.78	72.81 ± 3.06	69.19 ± 2.39	**88.31** ± 1.98
	SVM	68.19 ± 3.44	89.63 ± 2.20	60.44 ± 2.28	28.88 ± 6.55	69.75 ± 5.13	92.44 ± 1.88	**93.00** ± 1.28
USPS	1-NN	63.39 ± 3.73	83.53 ± 1.25	83.40 ± 1.12	72.88 ± 2.51	82.51 ± 1.54	80.67 ± 1.33	**86.34** ± 0.90
	NB	64.35 ± 3.14	82.60 ± 1.80	80.77 ± 1.19	68.71 ± 2.86	80.47 ± 1.56	80.45 ± 1.90	**82.79** ± 1.46
	SVM	64.06 ± 3.35	82.20 ± 2.11	82.05 ± 1.61	67.77 ± 2.86	79.99 ± 2.29	78.23 ± 1.58	**86.07** ± 1.26
AR	1-NN	39.99 ± 5.11	65.44 ± 0.94	45.24 ± 0.86	74.29 ± 0.92	71.16 ± 0.97	84.08 ± 0.69	**94.09** ± 0.45
	NB	41.44 ± 4.27	68.24 ± 1.09	41.88 ± 1.26	61.72 ± 1.23	64.38 ± 1.67	79.74 ± 0.70	**88.17** ± 0.96
	SVM	39.16 ± 4.29	67.60 ± 1.00	39.07 ± 0.81	69.29 ± 1.46	67.99 ± 1.78	76.54 ± 0.72	**94.31** ± 0.67
Isolet	1-NN	67.54 ± 2.63	82.82 ± 1.36	68.75 ± 1.14	69.25 ± 0.76	81.03 ± 1.02	89.18 ± 0.60	**92.02** ± 0.47
	NB	61.64 ± 2.82	86.79 ± 0.51	71.97 ± 1.00	68.71 ± 0.61	85.68 ± 0.98	86.67 ± 0.52	**87.19** ± 0.62
	SVM	67.15 ± 2.12	88.32 ± 0.66	71.25 ± 1.33	69.92 ± 0.40	87.37 ± 1.13	84.97 ± 0.91	**92.03** ± 0.82
Tr11	1-NN	60.56 ± 9.89	52.86 ± 2.16	68.67 ± 4.78	77.75 ± 2.48	54.92 ± 3.75	51.31 ± 3.86	**78.08** ± 2.23
	NB	47.89 ± 7.13	52.67 ± 3.17	61.86 ± 3.16	57.94 ± 8.97	52.22 ± 4.72	51.94 ± 6.78	**62.83** ± 5.27
	SVM	63.36 ± 9.19	57.17 ± 5.64	72.61 ± 4.76	70.17 ± 3.16	65.14 ± 4.19	52.75 ± 4.70	**75.69** ± 2.67
YaleB	1-NN	67.94 ± 1.58	83.21 ± 0.54	61.08 ± 0.65	87.95 ± 0.57	78.80 ± 1.09	90.97 ± 0.67	**93.53** ± 0.76
	NB	60.81 ± 1.97	72.42 ± 1.57	65.97 ± 1.50	67.42 ± 1.42	67.24 ± 2.21	87.28 ± 0.68	**89.62** ± 0.95
	SVM	56.11 ± 2.55	74.74 ± 1.47	56.84 ± 0.66	66.86 ± 1.48	57.20 ± 1.83	84.49 ± 0.92	**92.62** ± 0.90

Bold indicates the highest accuracy; underline indicates the second highest.

4.3.2. Experimental Results and Analysis

As shown in Figure 2, the evaluation results clearly indicate that each method is highly dependent on the feature dimension parameter. Therefore, finding a strategy that can deter-

mine the optimal dimension reduction settings for all compared methods is an extremely challenging task. In addition, the classification accuracy of all the comparison methods increases with an increase in the number of feature dimensions, and then tends to stabilize. In comparison with other competitive methods, the features extracted by our method achieved higher classification accuracy on 1-NN, NB, and SVM classifiers. Nevertheless, methods based on adaptive graph updates, such as SPDA, do not always yield superior accuracy on 1-NN, NB, and SVM. This suggests that preserving solely local structural information in feature extraction may not consistently improve classification performance. As shown in Table 2, the comparison of our method and SPDA with MFA reveals that MFA consistently yields unsatisfactory results. The reason for this phenomenon is that MFA preserves the local structure of the data using a predefined, fixed graph, whereas our method and SPDA employ adaptive strategies. When comparing our method and DSSLMR with AGATRLDA and RNDA, we found that, in the majority of cases, our method and DSSLMR consistently achieved better results. This implies that the discriminative regression technique can effectively extract features from the data. Overall, the classification accuracy values for our method are always highlighted in bold or underlined, indicating that our method has achieved a competitive advantage. The excellent classification performance of the proposed method is primarily attributed to the fact that DP-FRL provides a unified framework that combines the learning of adaptive local neighborhood relationships within classes, the learning of marginal structural relationships between classes, and rank-one quadratic penalty constraints on the projection matrix. Within this framework, these components enhance one another, thereby driving the learned projection matrix to capture discriminative features from the data.

4.4. Classification Experiments on the Noise Datasets

4.4.1. Experimental Setup

To comprehensively evaluate the noise resilience of the proposed method, we conducted classification experiments on datasets with added noise. Before beginning the experiments, we randomly selected a subset of samples from each of the eight original datasets and introduced noise into them. The purpose of this operation was to simulate the noise present in real-world environments, thereby testing the performance of the proposed method under conditions closer to real-world applications. In this way, the experiment revealed the extent to which the proposed method maintains its stability and effectiveness when the data is affected by noise. To generate the noise-corrupted datasets, we randomly selected samples from the original datasets and applied Gaussian noise with a variance of 0.05 to 20% of the attributes of the selected samples. The noise is distributed over the range $[x_{min}, x_{max}]$, where x_{min} and x_{max} represent the minimum and maximum values of the attributes in that sample, respectively.The following settings were applied regarding the number of samples corrupted by noise in each dataset: (1) For the ORL and AR datasets, since the number of samples in each class is limited, 6 samples were randomly selected from each class and corrupted with noise. (2) For the USPS, Isolet, Tr11, and YaleB datasets, since the number of samples in each class is relatively large, 18 samples were randomly selected from each class to be corrupted. (3) For the remaining datasets, 30 samples were randomly selected from each class to be corrupted. The datasets were divided into training and test sets in the same manner as described in Section 4.3. Each experiment was run 10 times, and the corresponding classification accuracies are summarized in Table 3.

4.4.2. Experimental Results and Analysis

Table 3 presents the classification accuracy of all methods for the three classifiers on the noisy datasets. By observing the results in Tables 2 and 3, we find that, the classification

accuracy of all methods tends to decrease as noise is introduced. On the BreastCancer, ORL, USPS, AR, Isolet, and YaleB datasets, it can be clearly seen that adaptive neighborhood based methods, including SPDA and our method, are more robust to noise than the MFA method under the 1-NN and SVM classifiers, showing a smaller decline in classification accuracy. This is mainly because pre-computed graph based methods such as MFA rely on a graph constructed in advance; noise can distort this graph and thus degrade classification performance. Adaptive neighborhood assignment, by contrast, can revise neighborhood relationships to reduce the impact of noise. In comparison with methods based on discriminant regression (e.g., DSSLMR, PRDR), our method achieves better classification accuracy on the majority of datasets. One reason for this is that the similarity matrices in the manifold regularization terms of DSSLMR and PRDR are pre-computed, making them susceptible to noise. Overall, the proposed method outperforms other competing methods in most scenarios across all datasets, regardless of whether a 1-NN, NB, or SVM classifier is used. This advantage stems from two factors: (1) The proposed method is built based on a unified objective function and flexibly adjusts the similarity matrix during the optimization process, thereby enhancing robustness. (2) By flexibly learning both the local neighborhood relationships within classes and the marginal structural relationships between classes, the model guides the projection matrix to extract discriminative information beneficial for classification.

Table 3. Acc (%) (mean ± standard deviation) achieved by related methods on the seven noise datasets.

Dataset	Classifier	MFA	SPDA	AGATR-LDA	RNDA	PRDR	DSSLMR	OUR (DP-FRL)
BreastCancer	1-NN	84.70 ± 1.87	85.40 ± 3.24	85.38 ± 1.36	86.29 ± 1.14	84.03 ± 1.94	87.25 ± 1.20	**87.90** ± 1.28
	NB	86.78 ± 2.26	80.79 ± 4.55	84.64 ± 4.41	86.68 ± 1.39	79.74 ± 9.54	87.84 ± 0.85	**87.87** ± 2.07
	SVM	82.44 ± 3.17	83.06 ± 9.96	68.90 ± 9.85	81.22 ± 6.21	77.58 ± 9.89	84.95 ± 4.86	**88.80** ± 0.98
ORL	1-NN	49.56 ± 4.69	56.06 ± 3.53	56.81 ± 2.06	41.38 ± 4.75	55.31 ± 2.94	39.25 ± 3.38	**57.44** ± 3.85
	NB	28.75 ± 4.75	27.50 ± 3.25	**47.63** ± 2.58	26.13 ± 2.40	46.06 ± 2.18	26.94 ± 3.19	46.44 ± 2.84
	SVM	44.13 ± 5.00	48.88 ± 4.25	48.00 ± 2.23	26.06 ± 4.15	44.56 ± 2.23	24.25 ± 1.78	**53.31** ± 2.03
USPS	1-NN	59.53 ± 4.05	81.21 ± 1.43	83.45 ± 1.10	82.96 ± 1.31	83.41 ± 0.86	78.64 ± 1.24	**84.72** ± 1.00
	NB	59.86 ± 3.92	81.30 ± 1.35	80.62 ± 0.99	78.46 ± 2.29	81.35 ± 1.53	78.95 ± 1.54	**82.93** ± 1.19
	SVM	60.1 ± 3.39	79.46 ± 1.57	81.64 ± 1.37	79.86 ± 1.98	81.70 ± 1.45	76.29 ± 1.96	**84.93** ± 0.85
AR	1-NN	27.39 ± 3.21	55.64 ± 1.05	23.27 ± 0.74	37.79 ± 0.99	26.06 ± 0.67	66.87 ± 0.82	**73.99** ± 0.54
	NB	24.83 ± 2.47	53.91 ± 0.98	19.64 ± 0.69	24.62 ± 1.24	22.52 ± 1.26	52.24 ± 0.98	**64.53** ± 0.84
	SVM	25.04 ± 2.92	50.19 ± 1.68	38.74 ± 0.61	27.96 ± 1.12	27.23 ± 1.31	63.09 ± 0.94	**74.50** ± 0.61
Isolet	1-NN	40.37 ± 3.92	76.36 ± 0.92	63.02 ± 1.41	26.25 ± 2.33	75.41 ± 0.95	63.97 ± 1.11	**79.67** ± 0.38
	NB	32.52 ± 4.22	77.58 ± 1.25	66.96 ± 1.15	18.53 ± 1.69	77.48 ± 0.52	53.49 ± 1.86	**78.71** ± 1.26
	SVM	38.73 ± 3.45	79.44 ± 0.88	66.89 ± 1.37	19.89 ± 2.57	79.93 ± 0.72	59.20 ± 0.96	**80.25** ± 0.60
Tr11	1-NN	59.50 ± 8.79	33.69 ± 4.19	68.11 ± 5.89	71.89 ± 2.01	52.86 ± 6.54	12.39 ± 4.94	**74.83** ± 3.32
	NB	19.22 ± 8.86	19.97 ± 2.31	58.89 ± 6.39	60.25 ± 6.69	52.25 ± 4.69	38.75 ± 6.11	**61.72** ± 6.33
	SVM	56.03 ± 9.08	32.64 ± 4.69	70.44 ± 5.93	72.22 ± 1.89	60.06 ± 7.5	11.97 ± 4.76	**74.47** ± 3.69
YaleB	1-NN	50.31 ± 3.55	80.58 ± 0.83	50.64 ± 0.87	56.55 ± 0.85	53.54 ± 1.14	72.76 ± 0.96	**88.23** ± 0.93
	NB	44.71 ± 2.41	70.16 ± 1.16	15.43 ± 1.27	34.90 ± 0.95	34.34 ± 2.78	64.48 ± 1.63	**82.71** ± 1.37
	SVM	45.53 ± 3.24	80.34 ± 1.17	46.50 ± 0.59	32.02 ± 1.08	54.64 ± 1.39	70.40 ± 1.25	**87.43** ± 1.02

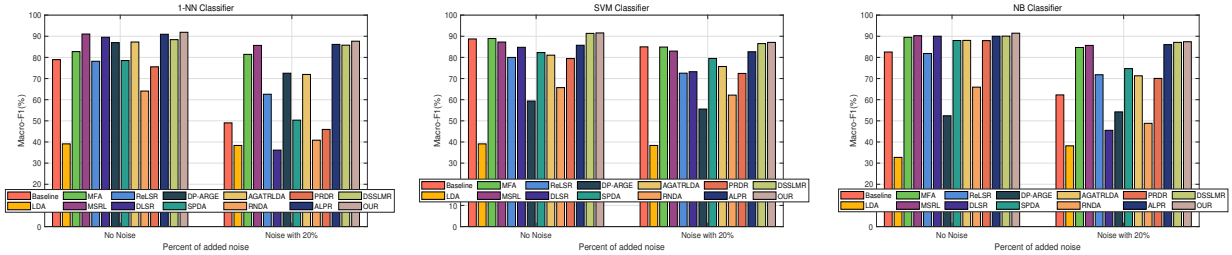
Bold indicates the highest accuracy; underline indicates the second highest.

4.5. Comparison with Discriminative Regression Methods

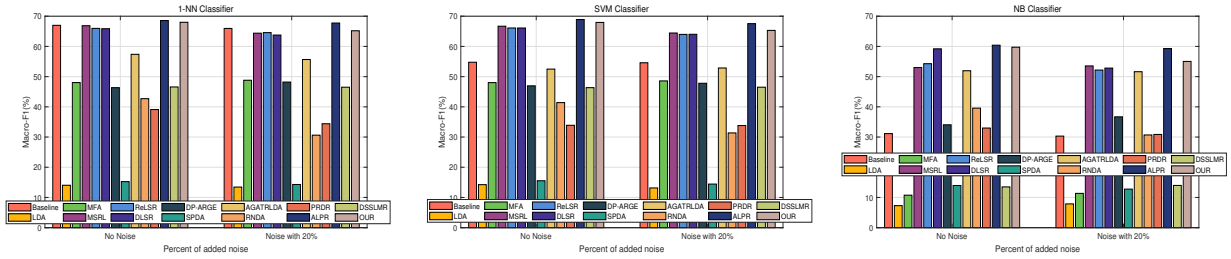
4.5.1. Experimental Setup

This subsection further compares the proposed method with discriminative regression methods, raw features, LDA [6], and DP-ARGE [18]. The discriminative regression methods include ALPR [17], MSRL [28], ReLSR [27], and DLSR [19]. The classification results for raw features are denoted as the Baseline. The training and test set splits for the datasets are the same as those in Section 4.3. Table 4 shows the classification results for all comparison methods on the seven datasets. To evaluate the performance of the models on imbalanced data, the Macro-F1 metric was included in the experiments on the BreastCancer and Tr11

datasets. The classification results on the BreastCancer and Tr11 datasets are presented in Figure 3 in terms of Macro-F1. To further demonstrate the noise tolerance of all the models, noise was added to 20%, 40%, and 60% of the features in the BreastCancer and Tr11 datasets, respectively. Figure 4 presents the results of the above noise experiments.



(**a**) Results of the classification task on the BreastCancer dataset.



(**b**) Results of the classification task on the Tr11 dataset.

Figure 3. Classification results for BreastCancer and Tr11 datasets. (**a**) Macro-F1 (%) of all classifiers on the BreastCancer dataset. (**b**) Macro-F1 (%) of all classifiers on the Tr11 dataset.

Table 4. Acc (%) (mean ± standard deviation) achieved by discriminative regression methods on the seven datasets.

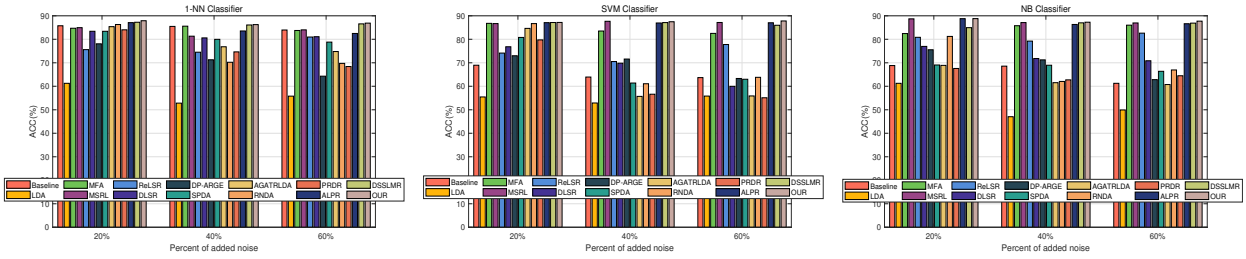
Dataset	Classifier	Baseline	LDA	ALPR	MSRL	ReLSR	DLSR	DP-ARGE	OUR (DP-FRL)
BreastCancer	1-NN	88.11 ± 1.87	70.00 ± 9.24	91.28 ± 1.56	87.88 ± 1.83	90.20 ± 2.72	86.13 ± 2.00	91.92 ± 3.98	**93.24** ± 1.89
	NB	87.09 ± 3.49	70.00 ± 9.24	92.34 ± 0.89	91.81 ± 0.49	87.13 ± 6.44	90.75 ± 0.42	90.09 ± 5.40	**93.88** ± 2.02
	SVM	82.04 ± 4.39	78.61 ± 5.13	**93.34** ± 1.16	91.30 ± 0.65	91.10 ± 2.68	90.63 ± 0.48	88.49 ± 4.05	92.77 ± 0.36
ORL	1-NN	89.00 ± 1.45	77.50 ± 3.50	83.75 ± 2.25	93.79 ± 1.31	75.19 ± 2.86	93.31 ± 1.50	93.31 ± 1.13	**93.88** ± 1.50
	NB	69.00 ± 1.95	44.31 ± 3.30	58.63 ± 4.50	79.44 ± 3.69	88.21 ± 2.98	87.06 ± 1.40	87.94 ± 3.13	**88.31** ± 1.98
	SVM	80.63 ± 1.13	77.38 ± 3.88	82.69 ± 2.19	92.31 ± 1.18	72.50 ± 4.13	90.31 ± 2.06	92.19 ± 1.44	**93.00** ± 1.28
USPS	1-NN	82.31 ± 0.96	83.27 ± 0.85	84.00 ± 1.48	84.55 ± 0.91	80.39 ± 2.60	85.91 ± 0.68	85.65 ± 1.74	**86.34** ± 0.90
	NB	78.33 ± 0.91	82.30 ± 1.22	82.31 ± 1.30	82.81 ± 1.35	74.40 ± 4.24	82.85 ± 1.08	**85.11** ± 2.12	82.79 ± 1.46
	SVM	81.77 ± 1.84	82.09 ± 1.28	81.90 ± 1.53	82.75 ± 1.14	75.36 ± 3.41	82.37 ± 0.88	82.93 ± 2.10	**86.07** ± 1.26
AR	1-NN	37.21 ± 0.89	93.30 ± 0.48	93.92 ± 0.52	91.86 ± 0.51	90.95 ± 3.06	87.04 ± 0.82	93.22 ± 0.95	**94.09** ± 0.45
	NB	71.15 ± 1.25	87.62 ± 1.08	87.66 ± 1.66	71.91 ± 1.80	85.90 ± 2.79	81.69 ± 0.96	87.31 ± 1.98	**88.17** ± 0.96
	SVM	23.43 ± 1.77	93.13 ± 0.69	93.54 ± 0.58	90.58 ± 1.18	89.10 ± 9.06	86.69 ± 1.42	94.20 ± 0.46	**94.31** ± 0.67
Isolet	1-NN	79.63 ± 0.77	89.77 ± 0.83	85.68 ± 1.15	85.18 ± 0.85	81.11 ± 1.99	91.90 ± 0.54	89.44 ± 2.28	**92.02** ± 0.47
	NB	79.81 ± 0.71	87.18 ± 0.54	81.26 ± 1.07	61.06 ± 9.80	45.10 ± 4.28	87.08 ± 0.58	85.47 ± 3.73	**87.19** ± 0.62
	SVM	87.06 ± 0.80	85.73 ± 0.97	86.01 ± 1.19	82.22 ± 1.27	46.12 ± 8.75	88.37 ± 0.81	89.04 ± 2.60	**92.03** ± 0.82
Tr11	1-NN	63.08 ± 3.64	17.5 ± 3.17	**82.03** ± 1.69	75.56 ± 2.28	74.69 ± 2.76	74.86 ± 2.78	67.64 ± 5.67	78.08 ± 2.23
	NB	35.03 ± 5.75	10.94 ± 5.63	**68.61** ± 12.94	53.44 ± 10.3	50.19 ± 6.97	59.44 ± 6.78	47.06 ± 7.63	62.83 ± 5.27
	SVM	81.39 ± 2.17	17.36 ± 3.14	**83.08** ± 1.75	75.56 ± 2.22	73.33 ± 2.89	73.44 ± 3.06	66.61 ± 5.39	75.69 ± 2.67
YaleB	1-NN	64.69 ± 0.91	92.65 ± 0.42	93.51 ± 0.49	92.72 ± 0.56	85.12 ± 1.53	92.68 ± 0.63	93.47 ± 0.35	**93.53** ± 0.76
	NB	62.21 ± 2.22	89.60 ± 0.96	89.32 ± 0.83	85.74 ± 1.12	28.36 ± 8.39	86.43 ± 1.06	88.92 ± 1.20	**89.62** ± 0.95
	SVM	27.33 ± 2.5	91.42 ± 0.78	92.52 ± 0.48	91.84 ± 0.74	48.18 ± 15.03	89.28 ± 0.77	92.11 ± 1.26	**92.62** ± 0.90

Bold indicates the highest accuracy; underline indicates the second highest.

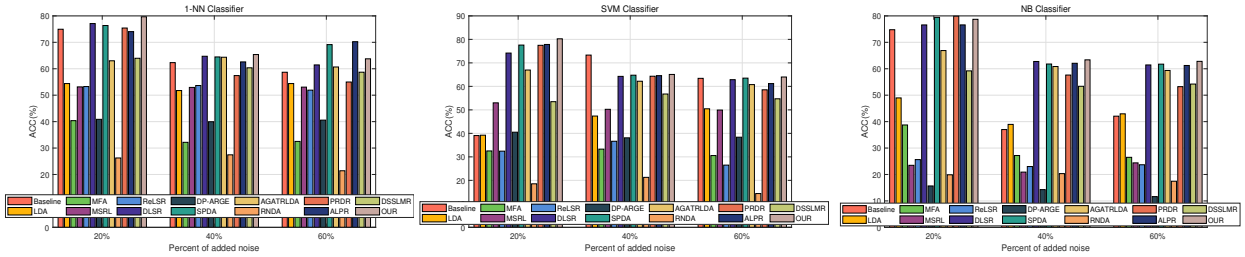
4.5.2. Experimental Results and Analysis

Table 4 shows the classification accuracy of all methods across 7 datasets and 3 classifiers (1-NN, NB, SVM). The proposed method achieved the highest accuracy in 18 out of 21 experimental scenarios. Compared to Baseline and LDA, DP-FRL demonstrated significant improvements across all datasets (e.g., USPS with 1-NN classifier: Baseline 82.31%, LDA 83.27%, OUR 86.34%). Compared with discriminant regression methods (ALPR,

MSRL, ReLSR, DLSR), DP-FRL maintains a lead in most scenarios and achieves an accuracy exceeding 90% when combined with SVM on the AR and Isolate datasets, demonstrating strong discriminative capability. Furthermore, the proposed method generally exhibits a smaller standard deviation, indicating good stability. Although ALPR performs better than 1-NN and NB classifiers on Tr11, DP-FRL still outperforms the other regression methods and baseline models, and its overall generalization ability across the seven datasets is satisfactory. Figure 3 shows the Macro-F1 results for no noise and 20% feature noise. Both in the absence of noise and when 20% noise is introduced, ALPR performs exceptionally well and serves as the primary strong-baseline comparison method. However, under 20% noise interference, while comparison methods such as ALPR exhibited significant performance degradation (e.g., a sharp drop in ALPR accuracy under the Tr11 with NB classifier setting), the proposed method maintained extremely high stability (remaining above 80% on BreastCancer with NB classifier). This indicates that, although ALPR is competitive under specific conditions, the proposed method possesses a significant advantage in noise robustness and delivers superior overall performance. Figure 4 illustrates the impact of varying degrees of corrosion (20% to 60%) on classification accuracy. As the degree of corrosion increases, the performance of most comparison methods (such as Baseline and LDA) deteriorates significantly, whereas the proposed method shows a gradual decline, demonstrating exceptional robustness. Under severe 60% corruption, the advantages of the proposed method are most evident (for example, it still maintains an accuracy of nearly 80% on the BreastCancer dataset with 1-NN classifier, far exceeding that of Baseline). Furthermore, although ALPR performs well under low corruption levels, its performance degrades more rapidly than DP-FRL under high corruption levels (40% and 60%). Overall, the proposed DP-FRL method maintains the highest classification stability even under severe sample corruption.



(**a**) Results of the classification task on the BreastCancer dataset.



(**b**) Results of the classification task on the Tr11 dataset.

Figure 4. Effect of sample corruption degree on classification accuracy. (**a**) Acc (%) of all classifiers on the BreastCancer dataset. (**b**) Acc (%) of all classifiers on the Tr11 dataset.

4.6. Ablation Study

4.6.1. Experimental Setup

To evaluate the contributions of each component in this framework to DP-FRL, we constructed three comparative variants of the proposed DP-FRL. As shown in the objective

function of DP-FRL in framework (9), this objective function consists of three terms. Therefore, these variants are defined as follows: (1) eliminating the second term in the objective function (9) to form VM1; (2) removing the third term from the objective function (9) to form VM2; and (3) eliminating the matrix W from the third term of the objective function (9) to form VM3. Subsequently, we conducted experiments on all the datasets using the same setup as in Sections 4.3 and 4.4, and the results are summarized in Table 5.

Table 5. Acc (%) (mean ± standard deviation) on all the datasets by using 1-NN, NB and SVM classifiers.

Dataset	Classifier	VM1	VM2	VM3	DP-FRL
BreastCancer	1-NN	92.00 ± 1.88	87.09 ± 2.10	87.25 ± 2.10	**93.24** ± 1.89
	NB	91.98 ± 1.87	87.03 ± 2.08	87.09 ± 2.13	**93.88** ± 2.02
	SVM	92.08 ± 2.01	87.07 ± 2.09	87.25 ± 2.10	**92.77** ± 0.36
BreastCancer with noise	1-NN	87.17 ± 1.15	85.72 ± 2.25	85.83 ± 2.15	**87.90** ± 1.28
	NB	87.82 ± 1.19	86.37 ± 2.28	86.56 ± 2.36	**87.87** ± 2.07
	SVM	88.55 ± 1.05	85.48 ± 2.49	85.82 ± 2.41	**88.80** ± 0.98
ORL	1-NN	88.00 ± 1.90	75.06 ± 1.56	76.25 ± 1.63	**93.88** ± 1.50
	NB	74.00 ± 1.93	54.06 ± 2.63	52.75 ± 2.88	**88.31** ± 1.98
	SVM	87.56 ± 1.94	75.44 ± 1.56	76.25 ± 1.63	**93.00** ± 1.28
ORL with noise	1-NN	55.38 ± 2.88	41.00 ± 2.73	41.31 ± 2.80	**57.44** ± 3.85
	NB	46.63 ± 2.93	27.13 ± 3.63	25.81 ± 4.19	<u>46.44</u> ± 2.84
	SVM	53.00 ± 3.18	40.56 ± 2.69	41.06 ± 2.81	**53.31** ± 2.03
USPS	1-NN	84.90 ± 1.13	83.72 ± 1.23	84.12 ± 1.21	**86.34** ± 0.90
	NB	84.13 ± 0.97	83.07 ± 0.89	83.35 ± 0.98	**82.79** ± 1.46
	SVM	84.82 ± 1.13	83.75 ± 1.27	84.09 ± 1.3	**86.07** ± 1.26
USPS with noise	1-NN	84.54 ± 0.72	83.10 ± 1.02	83.43 ± 1.01	**84.72** ± 1.00
	NB	83.52 ± 0.97	82.19 ± 1.24	82.49 ± 1.23	**82.93** ± 1.19
	SVM	83.95 ± 0.98	82.55 ± 1.03	82.85 ± 1.10	**84.93** ± 0.85
AR	1-NN	91.44 ± 1.05	91.76 ± 0.89	92.71 ± 0.85	**94.09** ± 0.45
	NB	79.50 ± 1.39	79.75 ± 1.63	81.86 ± 1.41	**88.17** ± 0.96
	SVM	91.56 ± 0.76	91.82 ± 0.46	92.31 ± 0.55	**94.31** ± 0.67
AR with noise	1-NN	70.14 ± 0.61	58.33 ± 0.45	59.06 ± 0.46	**73.99** ± 0.54
	NB	51.63 ± 2.63	33.01 ± 1.89	29.55 ± 1.70	**64.53** ± 0.84
	SVM	69.75 ± 0.45	58.31 ± 0.46	58.74 ± 0.34	**74.50** ± 0.61
Isolet	1-NN	88.18 ± 0.72	85.44 ± 0.92	85.94 ± 0.93	**92.02** ± 0.47
	NB	80.80 ± 1.07	76.75 ± 1.13	77.95 ± 0.90	**87.19** ± 0.62
	SVM	88.14 ± 0.69	85.54 ± 0.96	85.87 ± 0.88	**92.03** ± 0.82
Isolet with noise	1-NN	67.02 ± 0.70	49.98 ± 1.96	49.92 ± 2.00	**79.67** ± 0.38
	NB	54.88 ± 1.29	33.81 ± 1.89	34.37 ± 2.33	**78.71** ± 1.26
	SVM	66.05 ± 0.80	48.97 ± 1.67	48.75 ± 1.79	**80.25** ± 0.60
Tr11	1-NN	70.28 ± 2.44	77.33 ± 2.72	77.61 ± 2.39	**78.08** ± 2.23
	NB	47.50 ± 8.72	57.39 ± 9.72	53.69 ± 9.19	**62.83** ± 5.27
	SVM	70.28 ± 2.50	74.94 ± 2.44	75.06 ± 2.30	**75.69** ± 2.67
Tr11 with noise	1-NN	70.36 ± 4.12	72.69 ± 2.19	73.00 ± 2.38	**74.83** ± 3.32
	NB	39.61 ± 8.52	49.06 ± 9.17	46.22 ± 9.92	**61.72** ± 6.33
	SVM	70.14 ± 4.33	73.58 ± 2.47	73.61 ± 2.53	**74.47** ± 3.69
YaleB	1-NN	93.30 ± 0.89	93.34 ± 0.80	93.24 ± 0.78	**93.53** ± 0.76
	NB	89.61 ± 1.14	**89.73** ± 1.00	89.30 ± 0.97	89.62 ± 0.95
	SVM	92.41 ± 1.00	**92.86** ± 1.08	92.21 ± 1.03	92.62 ± 0.90
YaleB with noise	1-NN	77.06 ± 0.91	69.26 ± 1.16	69.57 ± 1.11	**88.23** ± 0.93
	NB	71.98 ± 1.22	65.50 ± 1.58	65.39 ± 1.54	**82.71** ± 1.37
	SVM	76.90 ± 0.81	68.90 ± 1.11	69.34 ± 1.09	**87.43** ± 1.02

Bold indicates the highest accuracy; underline indicates the second highest.

4.6.2. Experimental Results and Analysis

As shown in Table 5, a comparison of DP-FRL and VM1 reveals significant differences between them. More specifically, on the Isolet dataset, omitting the second term results in a greater decline in recognition accuracy than when using alternative settings. This observation strongly suggests that imposing rank-one quadratic penalty constraints on the projection matrix plays a positive role, as it effectively guides the learned projection matrix to capture the class-discriminative features inherent in the data. At the same time, in most cases, DP-FRL, VM1, and VM3 consistently outperform VM2 in terms of accuracy, suggesting that preserving local relationships among data points helps guide the projection matrix to extract discriminative features from the data. When all the datasets were corrupted by noise, the classification accuracy of the three variant methods performed poorly. These results confirm that DP-FRL is a unified framework in which components mutually reinforce one another and jointly guide the learned projection matrix to extract discriminative information critical for classification.

4.7. Convergence and Parameters Analysis

To assess the convergence behaviour of the proposed iterative optimization algorithm, a series of experiments were conducted on all the datasets. The purpose was to examine how the objective function value of DP-FRL varies with the number of iterations. The corresponding results are presented in Figure 5. As shown, DP-FRL achieves satisfactory convergence on all the datasets. More specifically, the objective function value decreases rapidly and stabilizes within 30 iterations. This finding provides strong evidence that the proposed iterative updating algorithm is highly efficient and capable of approaching a stable solution quickly.

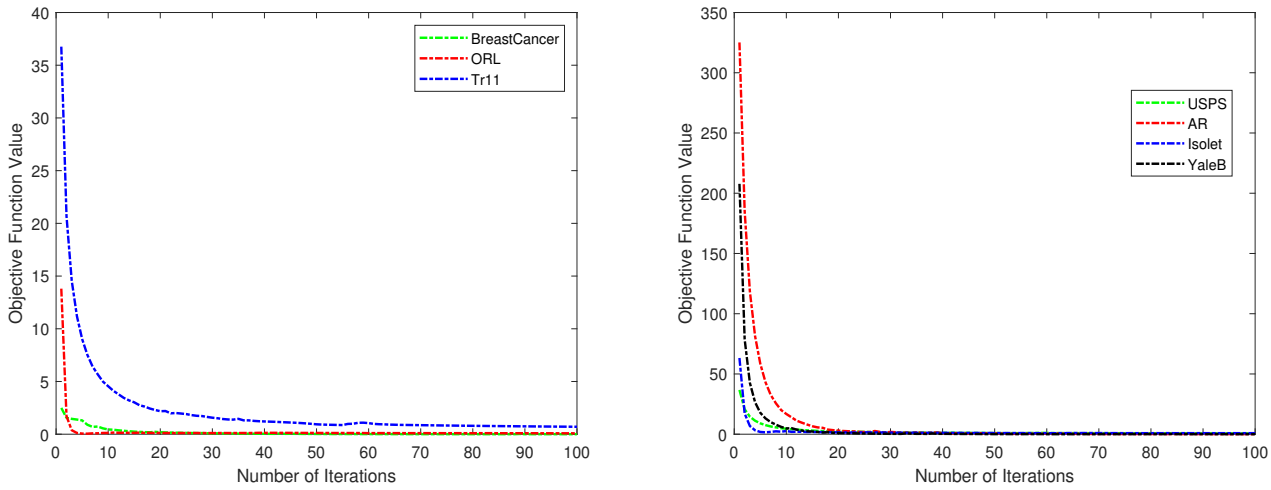
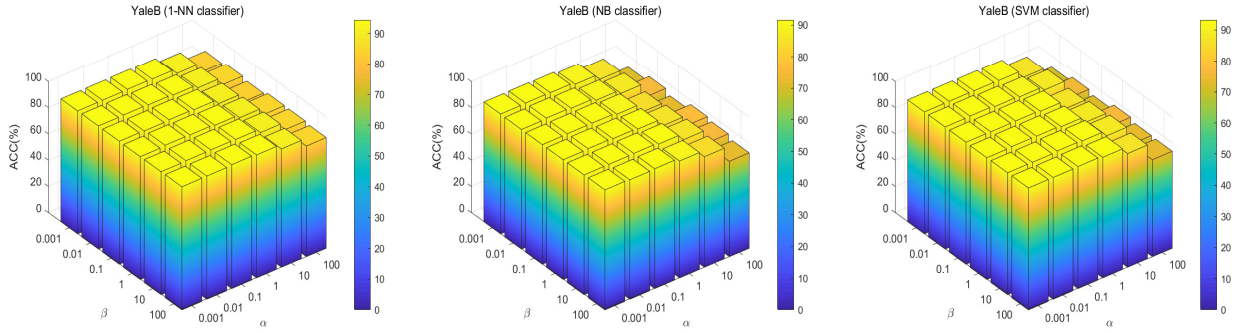


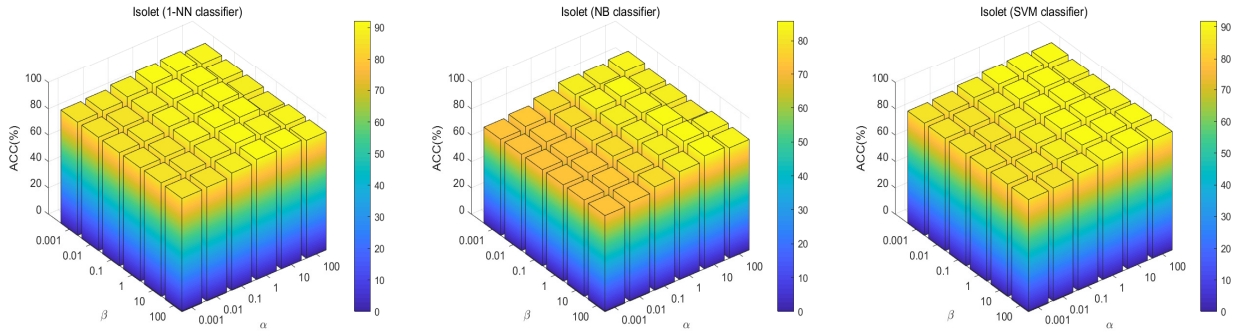
Figure 5. Objective function curves of DP-FRL versus the number of iterations.

The proposed method has two regularization parameters, namely α and β, which affect the performance of the method. To analyze whether the method is sensitive to these two parameters, we first defined a candidate set $[0.001, 0.01, 0.1, 1, 10, 100]$. Then, experiments with DP-FRL are conducted on the YaleB and Isolet datasets under different settings of the two parameters. Figure 6 presents the variation in classification accuracy with respect to the three parameters on the two datasets. The observations suggest that classification performance is affected by all two parameters. As can be seen from the results in Figure 6a,b, the optimal parameter combination is not fixed but varies depending on the dataset and classifier. This indicates that parameter selection is to some extent dependent on both the dataset and the classifier, and adjustments must be made based on the specific

task in practical applications. If the practical application does not require the highest possible classification accuracy, setting parameter α to the range $[1, 10]$ and parameter β to the range method is sensitive to these two parameters, we first defined a candidate set $[0.001, 0.01, 0.1, 1, 10, 100]$ is recommended. Within these ranges, the model typically achieves relatively good and stable classification performance, while also helping to reduce the computational overhead of parameter search. This setting can serve as an empirical parameter selection strategy for practical applications, providing a reference for subsequent experiments and model deployment.



(**a**) Results of the three classifiers on the YaleB dataset.



(**b**) Results of the three classifiers on the Isolet dataset.

Figure 6. Variation in Acc (%) of DP-FRL under different settings of the parameters α and β. The Acc (%) of DP-FRL (**a**) by using 1-NN, NB, and SVM classifiers on the YaleB dataset, (**b**) by using 1-NN, NB, and SVM classifiers on the Isolet dataset.

5. Conclusions

In this paper, we propose a supervised feature-extraction method called DP-FRL for classification tasks. The objective function of DP-FRL can be viewed as a unified framework that combines flexible edge structure learning, local relationship learning, and sparse regularization to guide the learning of the discriminative projection matrix. To mitigate the limitations imposed on related methods by fixed graph structures and predefined binary label matrices, both the flexible target matrix and the similarity matrix in the model are obtained through adaptive learning. Following an alternating optimization algorithm, all components within the framework reinforce one another, thereby producing a discriminative projection that effectively extracts discriminative information from the data. Comprehensive experiments demonstrate that, compared to commonly used methods, the features extracted by DP-FRL are more beneficial for classification tasks. Although the proposed flexible relationship learning strategy helps mitigate the effects of fixed graph structures and rigid label matrices, it remains susceptible to outliers and noise in complex scenarios. Therefore,

in future research, we will further explore more robust discriminative projection methods and conduct a theoretical analysis of the stability of the proposed method.

Author Contributions: Conceptualization, L.L.; methodology, L.L. and J.Z.; software, L.L. and J.Z.; validation, J.X. and Z.L.; formal analysis, Z.L.; writing—original draft preparation, L.L. and J.Z.; writing—review and editing, L.L.; visualization, J.X.; funding acquisition, L.L. and Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJQN202501226, KJZD-M202501201), the Doctoral "Direct Train" Scientific Research Project, Wanzhou District of Chongqing (Grant No. wzstc-20240007).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LDA	Linear Discriminant Analysis
ALPR	Adaptive Locality-Preserving Regression
DLSR	Discriminative Least Squares Regression
MSRL	Marginally Structured Representation Learning
ReLSR	Re-targeted Least Squares Regression
DP-ARGE	Discriminative Projection Learning with Adaptive Reversed Graph Embedding
DP-FRL	Discriminative Projection with Flexible Marginal and Local Relationship Learning
MFA	Marginal Fisher Analysis
KNN	K-Nearest Neighbor
SPDA	Submanifold-Preserving Discriminant Analysis
AGATRLDA	Attribute Graph Adjusted Trace Ratio Linear Discriminant Analysis
RNDA	Reverse Neighborhood Discriminant Analysis
DSSLMR	Discriminative Sparse Subspace Learning with Manifold Regularization
PRDR	Pairwise Relations Oriented Discriminative Regression
SVD	Singular Value Decomposition
1-NN	1-Nearest Neighbor
NB	Naive Bayes
SVM	Support Vector Machine

References

1. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [CrossRef]
2. He, X.; Niyogi, P. Locality preserving projections. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2004; pp. 153–160.
3. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [CrossRef]
4. He, X.; Cai, D.; Yan, S.; Zhang, H.-J. Neighborhood preserving embedding. In *Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, 17–21 October 2005*; IEEE: Piscataway, NJ, USA, 2005; Volume 2, pp. 1208–1213.
5. Tenenbaum, J.B.; de Silva, V.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319–2323. [CrossRef]
6. Fukunaga, K. *Introduction to Statistical Pattern Recognition*, 2nd ed.; Academic: New York, NY, USA, 1990.
7. Yan, S.; Xu, D.; Zhang, B.; Zhang, H.-J.; Yang, Q.; Lin, S. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 40–51. [CrossRef]
8. Cai, D.; He, X.; Zhou, K.; Han, J.; Bao, H. Locality sensitive discriminant analysis. In *IJCAI'07: Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI 2007)*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 2007; pp. 708–713.
9. Nie, F.; Wang, Z.; Wang, R.; Li, X. Submanifold-preserving discriminant analysis with an auto-optimized graph. *IEEE Trans. Cybern.* **2020**, *50*, 3682–3695. [CrossRef]

10. Wang, J.; Yin, H.; Nie, F.; Li, X. Robust outlier elimination trace ratio LDA for dimensionality reduction. *Pattern Recognit*. **2026**, *173*, 112937. [CrossRef]
11. Ye, J.; Xiong, T. Null space versus orthogonal linear discriminant analysis. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; pp. 1073–1080.
12. Wang, Q.; Lei, H.; Wang, F.; Wen, X.; Lin, Z.; Nie, F. Attribute graph adjusted trace ratio linear discriminant analysis for feature extraction. *Pattern Recognit*. **2026**, *176*, 113136. [CrossRef]
13. Huang, Y.; Lai, Z.; Kong, H.; Xu, J. Reverse neighborhood discriminant analysis for feature extraction. *Inf. Sci*. **2026**, *756*, 123831. [CrossRef]
14. Yu, X.; Zhao, X. YOLO-TCS: An enhanced multi-scale network for traffic sign detection integrating multi-level feature fusion and attention. *Multimed. Syst*. **2026**, *32*, 110. [CrossRef]
15. Yu, X.; Liu, M. Short-term load forecasting using an improved PatchTST model with time–frequency feature fusion. *Elec. Power Syst. Res*. **2027**, *265*, 114041. [CrossRef]
16. Li, X.; Chen, M.; Nie, F.; Wang, Q. Locality adaptive discriminant analysis. In *IJCAI'17: Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017*; AAAI Press: Palo Alto, CA, USA, 2017; pp. 2201–2207.
17. Wen, J.; Zhong, Z.; Zhang, Z.; Fei, L.; Lai, Z.; Chen, R. Adaptive locality preserving regression. *IEEE Trans. Circuits Syst. Video Technol*. **2020**, *30*, 75–88. [CrossRef]
18. Li, L.; Qu, H.; Li, Z.; Zheng, J.; Guo, F. Discriminative projection learning with adaptive reversed graph embedding for supervised and semi-supervised dimensionality reduction. *IEEE Trans. Circuits Syst. Video Technol*. **2022**, *32*, 8223–8237. [CrossRef]
19. Xiang, S.; Nie, F.; Meng, G.; Pan, C.; Zhang, C. Discriminative least squares regression for multiclass classification and feature selection. *IEEE Trans. Neural Netw. Learn. Syst*. **2012**, *23*, 1738–1754. [CrossRef]
20. Yang, Z.; Miao, J.; Wang, Y.; Huang, P.; Zhang, F. Structural enhanced robust discriminative least squares regression for image classification. *Expert Syst. Appl*. **2025**, *267*, 126154. [CrossRef]
21. Sugiyama, M. Local Fisher discriminant analysis for supervised dimensionality reduction. In *Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh PA, USA, 25–29 June 2006*; Association for Computing Machinery: New York, NY, USA, 2006; pp. 905–912.
22. Fan, Z.; Xu, Y.; Zhang, D. Local linear discriminant analysis framework using sample neighbors. *IEEE Trans. Neural Netw*. **2011**, *22*, 1119–1132. [CrossRef]
23. Feng, W.; Wang, Z.; Cao, X.; Cai, B.; Guo, W.; Ding, W. Discriminative sparse subspace learning with manifold regularization. *Expert Syst. Appl*. **2024**, *249*, 123831. [CrossRef]
24. Liu, Z.; Zhu, F.; Vasilakos, A.; Chen, X.; Zhao, Q. Discriminative approximate regression projection for feature extraction. *Inf. Fusion* **2025**, *120*, 103088. [CrossRef]
25. Liu, N.; Lai, Z.; Zhang, J.; Gao, C.; Kong, H. Capped norm based discriminant robust regression learning. *Pattern Recognit*. **2025**, *161*, 111288. [CrossRef]
26. Zhang, C.; Li, H.; Qian, Y.; Chen, C.; Gao, Y. Pairwise relations oriented discriminative regression. *IEEE Trans. Circuits Syst. Video Technol*. **2021**, *31*, 2646–2660. [CrossRef]
27. Zhang, X.; Wang, L.; Xiang, S.; Liu, C. Retargeted least squares regression algorithm. *IEEE Trans. Neural Netw. Learn. Syst*. **2015**, *26*, 2206–2213. [CrossRef]
28. Zhang, Z.; Shao, L.; Xu, Y.; Liu, L.; Yang, J. Marginal representation learning with graph structure self-adaptation. *IEEE Trans. Neural Netw. Learn. Syst*. **2018**, *29*, 4645–4659. [CrossRef]
29. Wang, S.; Pedrycz, W.; Zhu, Q.; Zhu, W. Subspace learning for unsupervised feature selection via matrix factorization. *Pattern Recognit*. **2015**, *48*, 10–19. [CrossRef]
30. Zou, H.; Hastie, T.; Tibshirani, R. Sparse principal component analysis. *J. Comput. Graph. Stat*. **2006**, *15*, 265–286. [CrossRef]
31. Hinojosa Lee, M.C.; Braet, J.; Springael, J. Performance metrics for multilabel emotion classification: Comparing Micro, Macro, and weighted F1-scores. *Appl. Sci*. **2024**, *14*, 9863. [CrossRef]
32. Nie, F.; Zhao, X.; Wang, R.; Li, X. Adaptive maximum entropy graph-guided fast locality discriminant analysis. *IEEE Trans. Cybern*. **2023**, *53*, 3574–3587. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.