



## LINEAR DISCRIMINANT ANALYSIS FOR ENHANCED DIMENSIONALITY REDUCTION AND PATTERN RECOGNITION

Pramod Kumar, Ganga Institute of Technology & Management, India (pramod.gill1@gmail.com)

### ABSTRACT

In machine learning and pattern recognition, dimensionality reduction can significantly enhance classifiers' discriminative performance and efficiency. Ratio Sum (RS) is a novel Linear Discriminant Analysis (LDA) variant that aims to optimise each dimension's projection matrix. However, RS does not consider the data's local geometric structure, which may lead to suboptimal solutions. An algorithm called Adaptive Neighbor Local Ratio Sum Linear Discriminant Analysis (ALDA) is proposed to overcome this limitation of RS. This algorithm employs an adaptive neighbour construction method to build the adjacency matrix, preserving the local geometric structure of the data and facilitating the construction of inter-class and intra-class matrices. This approach helps in finding a better representation of the data. Furthermore, the method utilizes an efficient non-parametric neighbourhood assignment strategy to construct the adjacency matrix, eliminating the need to adjust kernel parameters. Comparative experiments on UCI datasets and face datasets validate the effectiveness of this algorithm.

**Keywords:** Dimensionality Reduction, Pattern Recognition, Machine Learning, Linear Discriminated Analysis, Ratio Sum.

### 1. INTRODUCTION

In many data processing tasks in machine learning and pattern recognition, high-dimensional data is involved. High-dimensional data typically possesses an underlying low-dimensional structure that can effectively describe the variations between data points. Therefore, in data preprocessing, dimensionality reduction is considered one of the crucial steps to uncover the low-dimensional structure of data and enhance discriminative performance (Batool et al., 2023). Over the past few decades, there have been numerous unsupervised and supervised dimensionality reduction methods. Among them, Principal Component Analysis (PCA) (Huang et al., 2023) and Linear Discriminate Analysis (LDA) (Cao et al., 2023) are the most well-known unsupervised and supervised feature extraction methods, respectively.

LDA aims to find the optimal representation of data in low dimensions by maximizing the between-class scatter matrix and minimizing the within-class scatter matrix, effectively achieving feature extraction for data classification. The LDA algorithm can be formulated as an optimization problem based on the trace ratio (TR) criterion. However, the TR-based LDA cannot obtain a closed-form global optimal solution. Therefore, some researchers transformed the TR criterion into a ratio trace (RT) criterion, which can be solved using generalized eigenvalue decomposition. However, the solutions obtained through generalized eigenvalue decomposition are suboptimal and may lead to uncertainty in subsequent classification or clustering performance. Additionally, LDA has other drawbacks, such as sensitivity to small sample sizes, susceptibility to outliers, and limitations on maximum dimensionality reduction. The Maximum Margin Criterion (MMC) (Souza et al., 2023) was developed in response to these concerns. It effectively resolves the aforementioned challenges by maximising the average margin between each class.

The incapacity of LDA to handle manifold data is another drawback. In an effort to address this issue, scientists have developed graph-based techniques that make use of the connections among data points in an effort to maintain the local geometric structure of the data. Inspired by Locality Preserving Projection (LPP) (Shi et al., 2023) and Local Fisher Discriminated Analysis (LDA) (Li et al., 2023), Local Fisher Discriminated Analysis (LFDA) creates an adjacency matrix to mimic the local geometric structure of data and uses it to produce between-class and within-class scatter matrices. This makes it easier to optimise the objective function by precisely describing the geometric and discriminative structures of the data. Furthermore, (Li et al., 2021) proposed the Locality Sensitive Discriminate Analysis (LSDA) algorithm, which maximises the distance between data points of different classes within local regions by using predefined graphs, namely, intra-class and inter-class graphs, to compute the projection matrix.

Nevertheless, LFDA has a flaw in that it uses kernel functions to build its adjacency matrix, which leaves it vulnerable to noise and redundant features and leads to less-than-ideal categorization.

Dynamic Maximum Entropy Graph (DMEG) (Wang et al., 2022) and Adaptive Local Linear Discriminated Analysis (ALLDA) (JinlongQu et al., 2023) are two recent studies that scholars have carried out to address this problem. Adjacency matrices for data samples are adaptively constructed using the ALLDA method, which iteratively updates them to minimise the effects of noise and redundant features while maintaining the local data structure. However, ALLDA introduces an L0 norm constraint, which tends to yield trivial solutions. Therefore, the DMEG algorithm was proposed, which imposes a maximum entropy regularization constraint on the adjacency matrix to avoid trivial solutions.

However, all the algorithms mentioned above are still based on the TR criterion for feature extraction. Research indicates that the TR criterion tends to select projection directions with smaller overall sample variances, making it difficult to maximize the representation of the most representative information within the data samples after projection into the subspace.

To address this issue, researchers have proposed the RS criterion (Yan et al., 2022). This method attempts to make samples of the same category as close as possible in each dimension after projection, while keeping different samples as far apart as possible in each dimension. The goal is to maximize the ratio of between-class variance to within-class variance in each dimension. However, the aforementioned algorithms based on the RS criterion only consider the overall variance of the samples, ignoring the local geometric structure of the data.

Therefore, in order to overcome the limitation of the RS criterion not considering the local geometric structure of the data and to better preserve the local geometric structure, this paper proposes an Adaptive Neighbor Local Ratio Sum Linear Discriminate Analysis (ANLRSLDA) algorithm. This algorithm can better preserve the local geometric structure of the data and avoid introducing kernel parameters, thereby finding a more optimal representation of the data. Finally, comparative experiments are conducted on datasets such as YaleB, Pose27, UMIST, and UCI datasets. The experimental results show that this method achieves high classification accuracy and, in most cases, outperforms the compared algorithms.

## 2. THEORETICAL ANALYSIS

### 2.1 Ratio Sum

Let the data matrix be  $Y = [y_1, y_2, \dots, y_n] \in S^{d \times n}$ , where  $n$  represents the total number of samples and each sample  $y_i \in S^{d \times 1}$  has  $d$ -dimensional features. The objective of the TR problem is to find a projection matrix  $Z \in S^{d \times m}$ , which, through  $X = Z^U Y$ , projects high-dimensional data into low-dimensional data  $X \in S^{d \times m}$ . The projection matrix  $Z$  can be obtained by solving the following equation:

$$Z^* = \arg \max \frac{\text{tr}(Z^U T_b Z)}{\text{tr}(Z^U T_w Z)} \quad (1)$$

Where  $T_b, T_w$  represent the between-class scatter matrix and within-class scatter matrix, respectively.  $n_i$  represents the number of samples in the  $i$ -th class,  $\mu_i$  and  $\mu$  represent the means of the samples in the  $i$ -th class and all samples, respectively.

$$T_b = \sum_{i=1}^c \sum_{j=1}^{n_i} (y_{ij} - \mu_i)(y_{ij} - \mu_i)^U \quad (2)$$

$$T_w = \sum_{i=1}^c n_i (\mu_i - \mu)(\mu_i - \mu)^U \quad (3)$$

The TR problem can also be written in the following form:

$$Z^* = \arg \max \frac{\sum_{i=1}^k Z_i^U T_b Z_i}{\sum_{i=1}^k Z_i^U T_w Z_i} \quad (4)$$

When using the TR criterion for dimensionality reduction, this method tends to project onto directions with smaller variance. This leads to a situation where if there is small variance in the projection direction, the resulting subset will have difficulty maximizing the representation of the most representative information within the data.

To address this issue, Nei et al. proposed an RS criterion to avoid selecting directions with smaller variance and instead project onto directions with larger variance. The optimization problem can be formulated as follows:

$$Z^* = \arg \max = \sum_{i=1}^k \frac{Z_i^U T_b Z_i}{Z_i^U T_z Z_i} \tag{5}$$

Theoretically, based on the proof in (Juefei-Xu et al., 2015), it can be known that TR tends to select features with smaller variance, while RS can avoid this situation. To briefly illustrate this advantage of RS, consider the scenario with three independent features, and two of them need to be retained. The numerical relationships for TR can be expressed as:

$$\frac{10+0.1}{1+0.1} < \frac{15+10}{1+1} < \frac{15+0.1}{1+0.1} \tag{6}$$

Therefore, for TR, the first and third projection directions may be chosen. In contrast, the numerical relationships for RS are:

$$\frac{10}{1} + \frac{0.1}{0.1} < \frac{15}{1} + \frac{0.1}{0.1} < \frac{10}{1} + \frac{15}{1} \tag{7}$$

Clearly, the first and second projection directions exhibit stronger discriminative performance, implying that the first and second projection directions should be chosen. However, if we use the TR criterion, then the first and third projection directions would be selected, with the third projection direction having little discriminative capability. The simple example above demonstrates that RS indeed avoids selecting projection directions with smaller variance to best represent the most representative information within the data.

Let the data matrix be  $Y = [y_1, y_2, \dots, y_n] \in S^{d \times n}$ , where  $n$  represents the total number of samples and each sample  $y_i \in S^{d \times 1}$  has  $d$ -dimensional features. The objective of the TR problem is to find a projection matrix  $Z \in S^{d \times m}$ , which, through  $X = ZUY$ , projects high-dimensional data into low-dimensional data  $X \in S^{d \times m}$ . The projection matrix  $Z$  can be obtained by solving the following equation:

$$Z^* = \arg \max \frac{\text{tr}(Z^U T_b Z)}{\text{tr}(Z^U T_w Z)} \tag{1}$$

Where  $T_b, T_z$  represent the between-class scatter matrix and within-class scatter matrix, respectively.  $n_i$  represents the number of samples in the  $i$ -th class,  $\mu_i$  and  $\mu$  represent the means of the samples in the  $i$ -th class and all samples, respectively.

$$T_b = \sum_{i=1}^c \sum_{j=1}^{n_i} (y_{ij} - \mu_i)(y_{ij} - \mu_i)^U$$

(2)

$$T_w = \sum_{i=1}^c n_i (\mu_i - \mu)(\mu_i - \mu)^U \tag{3}$$

The TR problem can also be written in the following form:

$$Z^* = \arg \max \frac{\sum_{i=1}^k Z_i^U T_b Z_i}{\sum_{i=1}^k Z_i^U T_z Z_i} \tag{4}$$

When using the TR criterion for dimensionality reduction, this method tends to project onto directions with smaller variance. This leads to a situation where if there is small variance in the projection direction, the resulting subset will have difficulty maximizing the representation of the most representative information within the data.

To address this issue, Nei et al. proposed an RS criterion to avoid selecting directions with smaller variance and instead project onto directions with larger variance. The optimization problem can be formulated as follows:

$$Z^* = \arg \max = \sum_{i=1}^k \frac{Z_i^U T_b Z_i}{Z_i^U T_z Z_i} \tag{5}$$

Theoretically, based on the proof in (Juefei-Xu et al., 2015), it can be known that TR tends to select features with smaller variance, while RS can avoid this situation. To briefly illustrate this advantage of RS, consider the scenario

with three independent features, and two of them need to be retained. The numerical relationships for TR can be expressed as:

$$v \frac{10+0.1}{1+0.1} < \frac{15+10}{1+1} < \frac{15+0.1}{1+0.1} \tag{6}$$

Therefore, for TR, the first and third projection directions may be chosen. In contrast, the numerical relationships for RS are:

$$\frac{10}{1} + \frac{0.1}{0.1} < \frac{15}{1} + \frac{0.1}{0.1} < \frac{10}{1} + \frac{15}{1} \tag{7}$$

Clearly, the first and second projection directions exhibit stronger discriminative performance, implying that the first and second projection directions should be chosen. However, if we use the TR criterion, then the first and third projection directions would be selected, with the third projection direction having little discriminative capability. The simple example above demonstrates that RS indeed avoids selecting projection directions with smaller variance to best represent the most representative information within the data.

Feature	Projection direction 1	Projection direction 2	Projection direction 3
Distance between classes	15	10	0.1
Intra-class distance	1	1	0.1
Population variance	16	11	0.2
Ratio	15	10	1

While RS can avoid choosing projection directions with smaller variance to best represent the most representative information within the data, it does not consider the local geometric structure of the data. Therefore, considering embedding an adaptive nearest neighbor graph into RS to select the optimal projection direction is worthwhile.

### 2.2 Local Fisher Discriminates Analysis (LFDA)

In manifold data, data from the same class often form several clusters distributed in various locations in the original space. Traditional linear discriminates analysis algorithms tend to overlook the local characteristics of the data when calculating scatter, resulting in poor dimensionality reduction performance when dealing with manifold data. To address this issue, the LFDA algorithm was proposed in (Li et al., 2023).

LFDA constructs within-class adjacency matrix  $F_z$  and between-class adjacency matrix  $F_b$  using a Gaussian kernel function, which is formulated as follows, where  $N_k$  represents the samples of class  $k$ .

$$F_{ij} = \begin{cases} -\frac{\|y_i - y_j\|_2^2}{2\sigma^2}, & y_i, y_j \in M_k \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

Using these two adjacency matrices, LFDA constructs the between-class scatter matrix  $T_b$  and the within-class scatter matrix  $T_z$ .  $T_b$  and  $T_z$  can be represented as follows:

$$T_b = YN_bY^U \tag{9}$$

$$T_z = YN_zY^U \tag{10}$$

Here,  $N_z = F_z - F_w$ ,  $N_b = D_b - F_b$  are  $n$ -dimensional Laplacian matrices, where  $E_b$  and  $E_z$  are degree matrices, with  $E_{b,ii} = \sum_{j=1}^n F_{b,ij}$ ,  $E_{z,ii} = \sum_{j=1}^n F_{z,ij}$ .

Therefore, the objective function of LFDA can be obtained by solving the following equation:

$$Z_{LFDA}^* = \arg \max \left( \frac{Z^U T_b Z}{Z^U T_z Z} \right) = \arg \max \left( \frac{Z^U Y M_b Y^U Z}{Z^U Y M_z Y^U Z} \right) \tag{11}$$

Because LFDA constructs adjacency matrices through a kernel function, this graph construction approach is susceptible to the choice of kernel parameters, leading to suboptimal classification results. To address this issue, an adaptive construction of the data's adjacency matrix is achieved by leveraging the local geometric structure between data points, thereby preserving the data's local information and avoiding the introduction of additional hyper parameters.

### 3. ADAPTIVE NEIGHBOR LOCAL RATIO SUM LINEAR DISCRIMINATES ANALYSIS ALGORITHM

#### 3.1 Adaptive Neighbor Graph Construction

Let  $y_i \in S^{d \times 1}$  neighbors be defined as the  $k$  nearest data points to  $y_i$  in the dataset, using the Euclidean distance as the distance metric, i.e.,  $e_{i,j}^y = \|y_i - y_j\|_2^2$  represents the distance between the  $i$ -th data point,  $y_i$ , and the  $j$ -th data point,  $y_j$ .  $F \in S^{n \times n}$  is the adjacency matrix between data points, where  $F_{ij}$  is the element in the  $i$ -th row and  $j$ -th column of the matrix  $F$ , indicating the neighbor relationship between the  $i$ -th data point and the  $j$ -th data point. All data points  $\{y_1, y_2, \dots, y_n\}$  can be considered as neighbors of the  $i$ -th data point,  $y_i$ , with weights  $F_{ij}$ . Smaller distances,  $d(y_i, y_j)$ , result in larger  $F_{ij}$  values, and vice versa. However, this approach would make the points closest to  $x_i$  have a weight of 1 as neighbors of  $y_i$ . Therefore, a regularization parameter  $\lambda$  is introduced so that each point in the data will have some weight as neighbors of  $y_i$ .  $F_{ij}$  can be computed using the following formula:

$$\min_{G_i^U 1=1, 0 < G_{ij} < 1} \sum_{j=1}^n \left( \|y_i - y_j\|_2^2 F_{ij} + \lambda F_{ij}^2 \right) \tag{12}$$

Where  $G_i^U$  is the  $i$ -th row vector of the matrix  $F$ , and  $\lambda = \frac{k}{2} e_{i,k+1}^y - \frac{1}{2} \sum_{j=1}^k e_{i,j}^y$ . Solving the equation above yields:

$$F_{ij} = \frac{e_{i,k+1}^y - e_{i,j}^y}{k e_{i,k+1}^y - \sum_{j=1}^k e_{i,j}^y} \tag{13}$$

Define  $F_b$  and  $F_z$  as the between-class adjacency matrix and within-class adjacency matrix, respectively:

$$F_b = \begin{cases} F_{ij}, & y_i \neq y_j \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

$$F_w = \begin{cases} F_{ij}, & y_i = y_j = c \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

Where  $y_i$  and  $y_j$  are the class labels of the data. Clearly, the adjacency matrix  $F = F_b + F_z$ .

#### 3.2 Adaptive Neighbor Local Ratio Sum Linear Discriminated Analysis

By obtaining the within-class weight matrix  $F_z$  and between-class weight matrix  $F_b$  through adaptive neighbor graph construction the RS objective function is as shown in Equation (5). However, considering that the objective function is optimized by seeking the maximum value, it is possible to obtain suboptimal results when  $Z_i^U T_z x_i$  is too small to meet the maximization condition. Therefore, an equivalent transformation of Equation (5) is performed, and an orthogonal constraint  $ZUZ = I$  is added to obtain the new objective function as follows:

$$Z^* = \arg \min_{Z^U Z = I} = \sum_{i=1}^m \frac{Z_i^U T_z x_i}{Z_i^U T_b x_i} \tag{16}$$

Where  $T_z$  and  $T_b$  represent the within-class scatter matrix and between-class scatter matrix, defined as follows:

$$T_z = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n F_{z,ij} (y_i - y_j) (y_i - y_j)^U \tag{17}$$

$$T_b = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n F_{b,ji} (y_i - y_j) (y_i - y_j)^U \tag{18}$$

Expanding  $T_z$  and  $T_b$  separately:

$$T_z = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n F_{z,ij} (y_i - y_j) (y_i - y_j)^U = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n F_{z,ij} (y_i y_i^U + y_j y_j^U - y_j y_i^U - y_i y_j^U) = \sum_{i=1}^n (\sum_{j=1}^n F_{z,ij}) y_i y_i^U - \sum_{i,j=1}^n F_{z,ij} y_j y_i^U = Y E_z Y^U - Y E_z Y^U = Y M_z Y^U \tag{19}$$

$$T_c = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n F_{b,ij} (y_i - y_j) (y_i - y_j)^U = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n F_{b,ij} (y_i y_i^U + y_j y_j^U - y_j y_i^U - y_i y_j^U) = \sum_{i=1}^n (\sum_{j=1}^n F_{b,ij}) y_i y_i^U - \sum_{i,j=1}^n F_{b,ij} y_i y_j^U = Y E_b Y^U - Y E_b Y^U = Y M_b Y^U \tag{20}$$

For Equations (19) and (20),  $M_z$  and  $M_b$  are the Laplacian matrices for within-class and between-class, respectively.  $E_z$  and  $E_b$  are  $n$ -dimensional diagonal matrices, where  $E_{z,ii} = \sum_{j=1}^n F_{z,ij}$ ,  $E_{b,ii} = \sum_{j=1}^n F_{b,ij}$  substituting into Equation (16), the final objective function for RS is as follows:

$$Z^* = \arg \min_{z^U_{Z=1}} \sum_{i=1}^m \frac{z_i^U Y U_z Y^U z_i}{z_i^U Y M_b Y^U z_i} \tag{21}$$

In the next section, the process of solving the objective function will be introduced.

### 3.3 Solving the Objective Function

In this section, a method called Greedy RS, as introduced in (Zhang et al., 2020), will be used to solve the objective function. The solving process starts from Equation (22):

$$Z^* = \arg \min_{z_m} \sum_{i=1}^m \frac{z_i^U Y U_z Y^U z_i}{z_i^U Y M_b Y^U z_i} \tag{22}$$

Subject to the constraint:  $z_m^U z_1 = z_m^U z_2 = \dots = z_m^U z_{m-1} = 0$ . Since this objective function is binomial, the solution should depend on the direction rather than the length, so it can be scaled as follows:

$$z_m^U Y M_b Y^U z_m = 1 \tag{23}$$

Considering the above constraints, by introducing the Lagrange operator  $\eta$  and  $\beta = [\beta_1, \beta_2, \dots, \beta_{m-1}]$ , the corresponding Lagrange equation is obtained:

$$\sum_{i=1}^{m-1} \beta_i z_m^U z_i - \eta (z_m^U Y M_b Y^U z_m - 1) \tag{24}$$

Taking the derivative of  $M$  with respect to  $w_m$  and setting the derivative result to 0, as shown below:

$$\frac{\partial L(z_m, \eta, \beta)}{\partial z_m} = 0 \tag{25}$$

Namely:

$$2 Y M_z Y^U z_m - \sum_{i=1}^{m-1} \beta_i z_i - 2 \eta Y M_b Y^U z_m = 0 \tag{26}$$

Pre-multiplying the equation above by  $z_1^U (Y M_b Y^U)^{-1}, \dots, z_{m-1}^U (Y M_b Y^U)^{-1}$  resulting in  $m - 1$  equation, as follows:

$$\left\{ \begin{array}{l} \beta_1 z_1^U (Y M_b Y^U)^{-1} z_1 + \dots + \beta_{m-1} z_{m-1}^U (Y M_b Y^U)^{-1} z_{m-1} = \\ \quad 2 z_2^U (Y^U)^{-1} M_b^{-1} M_z Y^U z_m \\ \beta_1 z_2^U (Y M_b Y^U)^{-1} z_1 + \dots + \beta_{m-1} z_{m-1}^U (Y M_b Y^U)^{-1} z_{m-1} = \\ \quad 2 z_2^U (Y^U)^{-1} M_b^{-1} M_z Y^U z_m \\ \quad \dots \\ \beta_1 z_{m-1}^U (Y M_b Y^U)^{-1} z_1 + \dots + \beta_{m-1} z_{m-1}^U (Y M_b Y^U)^{-1} z_{m-1} = \\ \quad 2 z_{m-1}^U (Y^U)^{-1} M_b^{-1} M_z Y^U z_m \end{array} \right. \tag{27}$$

To simplify the calculations, define the following equations:

$$\beta^{m-1} = [\beta_1, \beta_2, \dots, \beta_{m-1}] \tag{28}$$

$$Z^{m-1} = [z_1, z_2, \dots, z_{m-1}] \tag{29}$$

$$V^{m-1} = Z_{m-1}^U (Y M_b Y^U)^{-1} Z_{m-1} \tag{30}$$

$$V_{ij}^{m-1} = z_i^U (Y M_b Y^U)^{-1} z_j \tag{31}$$

Then, the  $m - 1$  equations in Equation (27) can be represented as follows:

$$\beta^{m-1} = 2 [V^{m-1}]^{-1} [Z^{m-1}]^U (Y^U)^{-1} M_b^{-1} M_z Y^U z_m \tag{32}$$

Pre-multiplying Equation (26) by  $(XL_bXT)^{-1}$  yields:

$$2(Y^U)^{-1}M_b^{-1}L_wY^Uz_m - \sum_{i=1}^{m-1} \beta_i(XM_bY^U)^{-1}w_i - 2\eta z_m = 0 \quad (33)$$

Finally, by combining Equation (32) and Equation (33), we obtain:

$$\begin{aligned} & \{I - (YM_bY^U)^{-1}Z^{m-1}[V^{m-1}]^{-1}[Z^{m-1}]^U\} \\ & (Y^U)^{-1}M_b^{-1}M_zY^Uz_m = \eta z_m \end{aligned} \quad (34)$$

When  $m = 1$ , Equation (34) can be written as  $M_zY^Uz_1 = \eta M_zY^Uz_m$ . Clearly,  $z_1$  is the eigenvector corresponding to the smallest eigenvalue obtained through generalized eigenvalue decomposition of  $(Y^U)^{-1}M_b^{-1}M_zY^U$ . When  $m = 2$ , it is known that  $Z1 = z1$ , and by substituting it into Equation (34), we can obtain the eigenvector  $z2$  corresponding to the next smallest eigenvalue. This process can be iterated to obtain the  $m$  columns of the projection matrix. Below is a summary of the algorithm steps for the Adaptive Neighbor Local Ratio Sum Linear Discriminate Algorithm.

Algorithm 1 (Adaptive Neighbor Local Ratio Sum Linear Discriminate Algorithm):

Input: Data  $Y = [y_1, y_2, \dots, y_n] \in S_d \times n$ , class labels  $X = \{x_i\}_{i=1}^n = 1$ , number of neighbors  $k$ .

Compute the adjacency matrix  $F$ .

Obtain the within-class adjacency matrix  $F_z$  and between-class adjacency matrix  $F_b$  through the adjacency matrix  $F$ .

Compute the within-class scatter matrix  $XL_wXT$  and between-class scatter matrix  $YM_bYU$  using  $E_w$  and  $F_b$ .

Calculate the smallest eigenvector  $w1$  of  $(Y^U)^{-1}M_b^{-1}M_zY^U$ .

Use Equation (34) iteratively to compute the eigenvectors  $z2, z3, \dots, z_m$  corresponding to the smallest eigenvalues.

Output:  $Z^*$ .

The computational complexity of this algorithm mainly lies in constructing the adjacency matrix, matrix inversion, and solving the projection matrix. Matrix multiplication and addition computational complexities are not considered for now. The computational complexity of constructing the adjacency matrix is  $O(n^2e)$ . The computational complexity of matrix inversion is  $O(e^3)$ . The computational complexity of solving the eigenvectors in each iteration is  $O(ne^2 + n^2e)$ , resulting in a computational complexity of  $O(nme^2 + n^2me)$  for obtaining the projection matrix. Since  $n \gg e$ , the overall computational complexity of this algorithm is  $O(nme^2 + n^2me)$ .

## 4. EXPERIMENTS AND ANALYSIS

In this paper's experiments, MATLAB R2016b was used to implement various algorithms on a computer with 128 GB of RAM and an Intel i9-10980XE 3.0 GHz CPU. To validate the effectiveness of this algorithm, experiments were conducted comparing the ANLRSLDA algorithm with MMC, LSDA, LFDA, ALLDA, DMEG, Greedy RS, and other algorithms on the YaleB, Pose27, UMIST face datasets, and three UCI datasets. During the experiments, only preprocessing was applied to the face datasets. The projection matrix  $Z^*$  was obtained through training, and then dimensionality reduction was performed on all data. Finally, a nearest-neighbor classifier based on Euclidean distance was used to classify the reduced data. The experiments were repeated 10 times to obtain the algorithm's accuracy and standard deviation.

### 4.1 Parameter Settings

In this experiment, as both the proposed algorithm and LSDA, LFDA, ALLDA, DMEG, and Greedy RS algorithms involve parameter selection, the parameters were carefully chosen. For LSDA, the weight parameter  $\beta$  for intra-class and inter-class graphs was varied between 0 and 1. Following the empirical values from literature (Wang et al., 2022), the regularization parameter  $\eta$  for DMEG was set to 1 to avoid trivial solutions. Additionally, the number of intra-class neighbor's kin in the proposed algorithm, LSDA, LFDA, ALLDA, DMEG, and Greedy RS algorithms was varied from 1 to  $n_i - 1$ , where  $n_i$  represents the minimum number of samples in each class. The number of non-class neighbor's kout for LSDA was fixed at 10. Finally, hyper parameter search tools were employed to obtain the optimal parameters for each algorithm.

### 4.2 Dataset Description and Data Preprocessing

In this experiment, a total of six datasets were used for comparative analysis, as detailed in Table 2. These datasets include three face datasets and three UCI datasets. The face datasets consist of YaleB, Pose27, and UMIST, while the UCI datasets are Chess, Cancer, and Balance. The YaleB face dataset comprises more than 2,400 color face images of 38 individuals, each with pixel dimensions of 32x32. Each individual has approximately 64 near-frontal images taken under different lighting conditions. The Pose27 dataset is a subset of the PIE dataset, consisting of 3,329 pose photos of 68 individuals, with each photo having pixel dimensions of 64x64. The UMIST face dataset contains 1,012 images of 20 individuals, each with pixel dimensions of 32x32. These images cover a range of poses from side profiles to front views, representing a variety of ethnicities, genders, and appearances.

**Table 2: Dataset Description**

Data Set	Original Dimensions	Number Of Samples	Number Of Categories	Dimensions After PCA Preprocessing
<b>YaleB</b>	32×32	2 414	38	107
<b>Pose27</b>	64×64	3 329	68	68
<b>UMIST</b>	32×32	1 012	20	67
<b>Chess</b>	36	3 196	2	—
<b>Cancer</b>	9	683	2	—
<b>Balance</b>	4	625	3	—

During the experiment, only the face datasets were scaled, and PCA was applied to preprocess the training samples to remove zero space, retaining 95% of the original data's variance. This was done to avoid singularity in scatter matrices, which could occur due to zero space in the samples when comparing algorithms.

**4.3 UCI Dataset Experiments and Analysis**

For the UCI datasets, experiments were conducted on Chess, Balance, and Cancer datasets at training rates of 10%, 20%, 30%, and 40%. Table 3 lists the recognition rates and corresponding dimensions for each algorithm in the three UCI datasets the data in bold represents the highest recognition rate and standard deviation at that training rate. From Table 3, it can be observed that the ANLRSLDA algorithm performs the best among all algorithms on the three UCI datasets, with recognition rates approximately 1% higher than the other algorithms in each case. Furthermore, when using graph embedding methods such as LSDA and LFDA, the proposed algorithm consistently achieves approximately a 2% higher recognition rate. For the adaptive graph construction method ALLDA, the proposed algorithm outperforms it by 3-4% at training rates of 10% and 20%. Additionally, compared to the Greedy RS algorithm, the proposed algorithm consistently exhibits better recognition rates than other algorithms.

In the Chess dataset, the ANLRSLDA algorithm consistently achieves the highest recognition rates across all four training rates, approximately outperforming each of the comparison algorithms by 1-2 percentage points. Notably, at a training rate of 20%, ANLRSLDA surpasses the Greedy RS algorithm by 5.08 percentage points.

In the Cancer and Balance datasets, although the recognition rates of all algorithms are relatively close, ANLRSLDA consistently outperforms the other algorithms by approximately 1 percentage point. In summary, across the three UCI datasets, ANLRSLDA demonstrates significantly superior performance compared to other methods.

**4.4 Face Dataset Experiments and Analysis**

For the face datasets, experiments were conducted on YaleB, Pose27, and UMIST datasets at training rates of 10%, 20%, 30%, and 40%. Table 4 presents the recognition rates and corresponding dimensions for each algorithm across the three face datasets. Additionally, Figures 1, 2, and 3 visualize the accuracy obtained by seven dimensionality reduction methods at various dimensions in the YaleB, Pose27, and UMIST datasets, respectively. Bold data in Table 4 represents the highest recognition rate and standard deviation at each training rate.

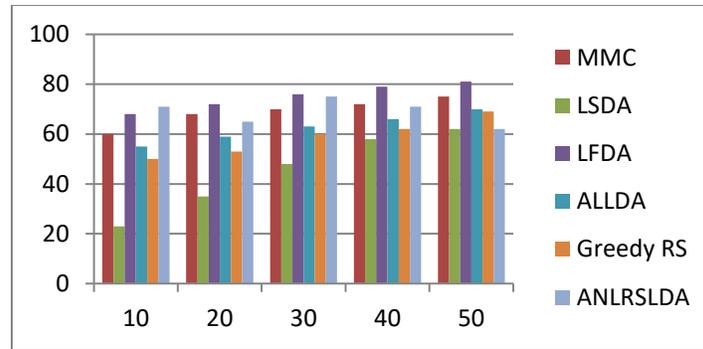
**Table 3: Recognition rate in UCI dataset**

Data set	Algorithm	Accuracy
<b>Chess</b>	MMC	95.05±0.33(9)
	LSDA	95.90±1.00 (13)
	LFDA	94.73±0.89(15)
	ALLDA	95.77±0.50 (11)

	DMEG	95.95±0.74(15)
	Greedy RS	93.79±0.65(11)
	ANLRSLDA	96.92±0.14(15)
<b>Cancer</b>	MMC	90.04±0.42(8)
	LSDA	89.36±0.93(8)
	LFDA	88.94±0.92(7)
	ALLDA	89.00±0.91(9)
	DMEG	89.06±2.39(7)
	Greedy RS	88.45±1.15(9)
	ANLRSLDA	90.53±0.46(8)
<b>Balance</b>	MMC	91.04±1.44(4)
	LSDA	91.31±1.73(3)
	LFDA	91.52±1.84(3)
	ALLDA	86.93±4.53(3)
	DMEG	91.25±2.05(3)
	Greedy RS	91.09±2.07(2)
	ANLRSLDA	91.68±1.97(2)

Data set	Algorithm	Accuracy
<b>YaleB</b>	MMC	93.69±0.75(63)
	LSDA	92.83±0.73 (72)
	LFDA	91.20±1.53 (33)
	ALLDA	93.69±0.50 (68)
	DMEG	92.27±0.85(37)
	Greedy RS	92.06±0.25(28)
	ANLRSLDA	94.41±0.45(59)
<b>Pose27</b>	MMC	96.38±0.39 (56)
	LSDA	97.36±0.08 (55)
	LFDA	96.70±0.01(15)
	ALLDA	97.28±0.29 (41)
	DMEG	97.62±0.29 (41)
	Greedy RS	97.07±0.18(24)
	ANLRSLDA	98.29±0.47 (33)
<b>UMIST</b>	MMC	99.11±0.85(20)
	LSDA	98.84±0.56 (50)
	LFDA	98.98±0.61(28)
	ALLDA	98.48±0.43 (61)
	DMEG	98.78±0.77(32)
	Greedy RS	98.81±0.92(30)
	ANLRSLDA	99.45±0.25(27)

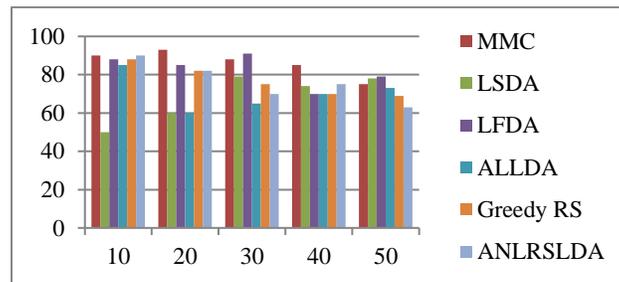
From Table 4, it is evident that ANLRSLDA outperforms other algorithms in most cases across the three face datasets. Furthermore, when using graph embedding methods such as LSDA and LFDA, ANLRSLDA consistently achieves approximately a 2% higher recognition rate. For the adaptive graph construction method ALLDA, ANLRSLDA's recognition rates are consistently 2-3% higher, and in the UMIST dataset, ANLRSLDA outperforms ALLDA by 4-5%. Across datasets with different training rates, ANLRSLDA's recognition rates are significantly superior to Greedy RS.



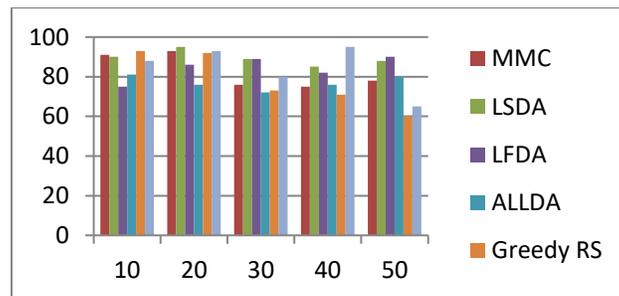
**Figure 1:** Performance of each algorithm on the YaleB data set

In the YaleB dataset, while ANLRSLDA's recognition rate is not significantly different from other algorithms at a training rate of 10%, as the training rate increases, ANLRSLDA starts to exhibit better recognition rates. At each training rate, ANLRSLDA consistently outperforms other algorithms by 1-2 percentage points. Figure 1 with table 5 illustrates that ANLRSLDA achieves the highest recognition rates in most dimensions. However, it is observed that as the dimensionality increases to a certain extent, further dimension increase does not enhance algorithm accuracy. Instead, it rapidly decreases. This phenomenon is due to the fact that adding more features may not necessarily improve classification, potentially leading to a decline in recognition accuracy (Li et al., 2017).

In the Pose27 dataset, when the training rate is 10%, ANLRSLDA's recognition rate is slightly lower than DMEG. However, with an increase in the training rate, ANLRSLDA's recognition rate starts to surpass all other algorithms. Figure 2 shows that ANLRSLDA consistently outperforms other algorithms in most dimensions.



**Figure 2:** Performance of each algorithm on the Pose27 data set.



**Figure 3:** Performance of each algorithm on the UMIST data set.

In the UMIST dataset at various training rates, it is evident that ANLRSLDA consistently outperforms other algorithms, with significantly better performance at a training rate of 10%, surpassing LSDA and ALLDA. Figure 3 demonstrates that on the UMIST dataset, all algorithms perform well, but ANLRSLDA's recognition rate remains consistently higher than other algorithms. In conclusion, the proposed algorithm exhibits excellent performance in face datasets.

## 5. CONCLUSION

In this work, we have introduced a dimensionality reduction algorithm called Adaptive Neighbor Local Ratio Sum Linear Discriminate Analysis (ANLRSLDA). The advantages of this algorithm lie in its ability to adaptively construct data's adjacency matrix based on the relationships between data points, effectively preserving the local geometric structure of the data. Additionally, the algorithm performs adaptive graph construction without introducing heat kernel parameters, and it can efficiently mitigate the influence of noise and redundant features in the original data space. Finally, comparative experiments have demonstrated the effectiveness of the proposed algorithm. Although the ANLRSLDA algorithm has shown promising results compared to other algorithms, there are still areas for further exploration. Future research may delve into discovering more optimal optimization methods for solving the RS problem and investigating how to learn improved adjacency matrices.

## REFERENCES

- Batool A., WasifNisar M., Khan M. A., Shah J. H., Usman T., Robertas D. , (2023), Traffic sign recognition using proposed lightweight twig-net with linear discriminant classifier for biometric application, *Image and Vision Computing*, <https://doi.org/10.1016/j.imavis>. 135 ,ISSN 0262-8856..
- Hongchun Q., Lin L., Zhaoni L., Jian Z., (2021). Supervised discriminant Isomap with maximum margin graph regularization for dimensionality reduction, *Expert Systems with Applications*, <https://doi.org/10.1016/j.eswa>. 180, ISSN 0957-4174.
- Huang P., Yangyang S., Zhangjing Y., Zhang C. , Yang G., (2022). Dual collaborative representation based discriminant projection for face recognition, *Computers and Electrical Engineering*, <https://doi.org/10.1016/j.compeleceng..108281>. 102, ISSN 0045-7906.
- Jinlong Q., Zhao X., Yun X., Xiaojun C., Zhihui L., Wang X., (2023). Adaptive Manifold Graph representation for Two-Dimensional Discriminant Projection, *Knowledge-Based Systems*, <https://doi.org/10.1016/j.knosys>. , 26, ISSN 0950-7051.
- Juefei X. F. & Marios S. (2016) Multi-class Fukunaga Koontz discriminant analysis for enhanced face recognition, *Pattern Recognition*, <https://doi.org/10.1016/j.patcog>. 52, 186-205,ISSN 0031-3203.
- Li S., Hengmin Z., Ruijun M., Jianhang Z., Jie W., Bob Z., (2023).Linear discriminant analysis with generalized kernel constraint for robust image classification, *Pattern Recognition*, <https://doi.org/10.1016/j.patcog> 136, ISSN 0031-3203.
- Li C. N., Yuan-H. S., Wei-Jie C., Zhen W., Nai-Yang D., (2021) Generalized two-dimensional linear discriminant analysis with regularization, *Neural Networks*, <https://doi.org/10.1016/j.neunet>. 142, 73-91, ISSN 0893-6080.
- Li W., Wang Y., Zhen X., Yinyan J., Zongqing L., Qingmin L., (2017). Weighted contourlet binary patterns and image-based fisher linear discriminant for face recognition, *Neurocomputing*, <https://doi.org/10.1016/j.neucom>., 267, 436-446, ISSN 0925-2312,
- .Shi M., Xiaowei Z., Xiaoyan Y., Xiaojun C., Fan N., Jun G., (2023), Multiview Latent Structure Learning: Local structure-guided cross-view discriminant analysis, *Knowledge-Based Systems*, 276, ,110707, ISSN 0950-7051.
- Souza L. S. , Bernardo B. G., Jing-H., Kazuhiro F., (2020) ,Enhanced Grassmann discriminant analysis with randomized time warping for motion recognition, *Pattern Recognition*, , <https://doi.org/10.1016/j.patcog>. 97, ISSN 0031-3203, 107028.
- Shanwen Z., Zhang C., Wang X., (2020), Plant species recognition based on global–local maximum margin discriminant projection, *Knowledge-Based Systems*, <https://doi.org/10.1016/j.knosys>. , 20, ISSN 0950-7051.
- Wang Z.g., Haojie H., Rong W., Qianrong Z. , Feiping N., Xuelong L., (2022) ,Capped  $\ell_p$ -norm linear discriminant analysis for robust projections learning, *Neurocomputing*, <https://doi.org/10.1016/j.neucom>., 511 , 399-409,ISSN 0925-2312.
- Yan T., Dong W., Tangbin X., Jie L., Zhike P., Lifeng X., (2022), Investigation on optimal discriminant directions of linear discriminant analysis for locating informative frequency bands for machine health monitoring, *Mechanical Systems and Signal Processing*, <https://doi.org/10.1016/j.ymssp>. 180 ,ISSN 0888-3270,109424.
- Zhong J. ,Shan Z., Zhang X., Haifeng Lu, Hong Peng, Bin Hu, (2023), Robust discriminant feature extraction for automatic depression recognition, *Biomedical Signal Processing and Control*, <https://doi.org/10.1016/j.bspc>. , 82,ISSN 1746-8094, 104505.
- Ziyi C., Zhang S. , Youlin L., Casey J. S., Alex M. S., Yechan H., Garth J. S., (2023). Spectral classification by generative adversarial linear discriminant analysis, *Analytical ChimicaActa*, <https://doi.org/10.1016/j.aca..341129>. , 1261, ISSN 0003-2670.