# Linear Discriminant Analysis (LDA) based on auxiliary slicing for binary classification data

Xiaojia Ran [1, *], Boxin Nie [2]

[1] School of Public Health, Hainan Medical University, Haikou, China, 571199

[2] School of Mathematics and Statistics, Northeastern University at Qinhuangdao, Shenyang, China, 066099

* Corresponding author: 15185822170@163.com

**Abstract.** Linear Discriminant Analysis (LDA) is a foundational algorithm in the realm of machine learning specifically designed for classification tasks. While effective, LDA's limitation lies in its ability to only reduce the feature space to one dimension in binary classification scenarios. This can result in a loss of vital information from the predictor variables, impacting the model's predictive capabilities. To address this issue, this study introduces an enhanced discriminant analysis model leveraging auxiliary slicing. This novel approach not only removes the constraint of one-dimensional reduction post LDA through equidistant binning but also enhances the extraction of local information by introducing a continuous auxiliary response variable. Through simulation experiments and real data analysis, it is demonstrated that our proposed method outperforms the traditional LDA model in terms of prediction accuracy and robustness.

**Keywords:** High-dimensional data, binary classification problem, Linear Discriminant Analysis (LDA), auxiliary slicing.

## 1. Introduction

With the continuous advancement of data collection and storage technologies, high-frequency recording of research object characteristics has become possible. While this enriches the information available in the data, it also presents new challenges. Specifically, high-dimensional data features often lead to the curse of dimensionality, a problem encountered by prediction models. Additionally, variables frequently exhibit high linear correlation, which reduces the efficiency of certain statistical models. For instance, hyperspectral remote sensing technology, compared to traditional remote sensing, generates a large amount of narrow spectral continuous image data, resulting in increased data volume and redundancy among connected bands [1] [2]. Furthermore, as more data is added and dimensionality increases, the Hughes phenomenon may occur [3].

Various methods exist to address the curse of dimensionality problem. From a modeling perspective, generalized linear models based on penalty estimation, such as LASSO, ridge regression, and elastic net, are available [4] [5] [6]. Additionally, there are dimensionality reduction techniques based on variable selection, including mutual information-based variable selection [7], chi-square test-based variable selection [8], and high-dimensional data prediction models based on neural networks. For example, artificial neural networks have been employed to predict basic physical properties of over 2500 types of hydrocarbons and halogenated hydrocarbon molecules [9]. These methods effectively alleviate the curse of dimensionality but possess limitations. Penalized generalized linear models assume linear relationships between predictor and response variables, rendering them unsuitable for nonlinear scenarios. Dimensionality reduction techniques based on variable selection reduce dimensions by selecting a subset of features from the full set, but the selection may not be unique and computational efficiency is relatively low. Moreover, the lack of interpretability restricts the applicability of neural networks. This paper focuses on feature extraction techniques as an alternative. This approach achieves dimensionality reduction through linear or nonlinear combinations of original variables. Such methods typically have explicit solutions and yield somewhat interpretable dimensionality reduction. Feature extraction methods are primarily divided into two categories: unsupervised dimensionality reduction techniques, such as principal component

analysis and kernel principal component analysis, which respectively handle linear and nonlinear relationships between predictor and response variables. Principal component analysis is a suitable solution when dealing with gas outburst data, which involves many influencing factors and potential autocorrelation [10]. Combining kernel principal component analysis with multiple algorithm optimizations greatly improves accuracy in predicting high-dimensional corroded variables by reducing noise interference [11]. However, the relationship between the extracted feature information and the response variables remains unknown.

In comparison, another type of dimensionality reduction technique, supervised dimensionality reduction, offers several advantages. Representative algorithms in this category include linear discriminant analysis (LDA) and its nonlinear extension method, kernel-based linear discriminant analysis. This method finds widespread applications. For example, Xiong Jianfang et al. [12] utilized the LDA algorithm to reduce the dimensionality of NIR spectral data and combined it with K-nearest neighbor (KNN), random forest (RF), extreme gradient boosting (XGBoost), and logistic regression (LR) algorithms to classify mixed datasets of DSP pollution samples and healthy samples. Wang Minghe et al. [13] employed linear discriminant analysis to improve the accuracy of speech endpoint detection by applying linear discriminant analysis mel-frequency cepstral coefficients (F-MFCC) endpoint detection method, enhancing separability between speech and background noise. In the field of face recognition, a fusion method combining PCA transformation, LDA feature, and SVM classifier has been proposed to reduce high-dimensional features to low dimensions while achieving better recognition performance. The recognition performance improves with increasing facial feature dimensionality until it reaches a stable level of complexity. Preprocessing using Radon transform to remove subjective factors is combined with PCA+LDA+SVM algorithm for face recognition [14] [15].

However, in binary classification tasks, linear discriminant analysis may not perform optimally. This is because linear discriminant analysis has limitations on the dimensionality of the reduction space, which cannot exceed the number of classes. Therefore, in binary classification tasks, the reduced space dimensionality can only be set to 1. However, this setting presents a clear problem when dealing with high-dimensional features, leading to the loss of a significant amount of original effective feature information. Classifying samples based on a single dimensionality perspective oversimplifies the model. To address this issue, this paper proposes an improved discriminant analysis model based on auxiliary slicing. This method overcomes the one-dimensional limitation of the feature space after LDA dimensionality reduction by utilizing equidistant binning techniques. It enhances the ability of LDA to capture effective information from predictor variables by generating a continuous auxiliary response variable. Simulation experiments demonstrate the effectiveness of the proposed method in improving the accuracy and robustness of the original LDA model. Empirical data analysis further proves the higher predictive accuracy of the proposed method compared to LDA.

## 2. Theory and Methods

### 2.1. Based on auxiliary slicing linear discriminant analysis.

LDA is a supervised linear learning algorithm and a method for reducing dimensionality with classification feature labels. Compared to PCA, LDA is an enhanced algorithm, primarily because data reduced by LDA is more readily classifiable. The key characteristic of the LDA algorithm lies in its capacity to comprehensively capture the feature information between each class, particularly in relation to the attribute features of different classes. Its objective is to identify a set of projection directions that maximize the Fisher criterion function. These vectors represent the optimal projection directions, maximizing the separation between samples from different classes while minimizing the separation within the same class.

In a dataset D of size m×n (m dimensions, n samples), the objective is to project the data onto a straight line while satisfying two conditions:

a) Ensure that samples from the same class have close projections.

b) Ensure that samples from different classes have distant projections.

If we consider the dataset D to be binary, represented as $D = \{(x_i, y_i)\}_{i=1}^{m}$ where yi E {0,1}, Xi is the set of independent variables, i= (0, 1), and m is the number of independent variables, the data is projected onto the line ω with $\Sigma_i$ representing the class covariance matrix. Assuming the centroids of the two classes are $u_i(i = 0,1)$, the definition is as follows:

$$\mu_i = \frac{1}{N_i}\sum_{x \in y_i} x \tag{1}$$

The projections of the centroids onto the line are represented as $\omega^T\mu_0$ and $\omega^T\mu_1$ω^T, while the covariances of the two classes are denoted as ω^T Σ_0 ω and ω^T Σ_1 ω. To meet the specified conditions, the objective is to minimize the covariance of projections for samples from the same class and maximize the separation between class centroids: minimize (ω^T Σ_0 ω + ω^T Σ_1 ω) and maximize (||ω^T μ_0 - ω^T μ_1 ||^2).

The within-class scatter matrix:

$$\begin{aligned} S_\omega &= \Sigma_0 + \Sigma_1 \\ &= \sum_{x \in X_0}(x - \mu_0)(x - \mu_0)^T + \sum_{x \in X_1}(x - \mu_1)(x - \mu_1)^T \end{aligned} \tag{2}$$

Between-class scatter matrix:

$$S_b = (\mu_0 - \mu_1)(\mu_0 - \mu_1)^T \tag{3}$$

Considering both conditions a) and b), the objective is to maximize:

$$\begin{aligned} J &= \frac{\|\omega^T\mu_0 + \omega^T\mu_1\|_2^2}{\omega^T\Sigma_0\omega + \omega^T\Sigma_1\omega} \\ &= \frac{\omega^T S_b \omega}{\omega^T S_\omega \omega} \end{aligned} \tag{4}$$

In equation (4), the solution for J is independent of the length of ω and only dependent on the direction. Setting $\omega^T S_\omega \omega$=1, the above equation is equivalent to:

$$min - \omega^T S_b \omega \tag{5}$$

$$s.t.\ \omega^T S_\omega \omega = 1 \tag{6}$$

By utilizing the Lagrange multiplier method, with λ representing the Lagrange multiplier, $S_b\omega$ aligns with the direction of $\mu_0 - \mu_1$. Assuming: $S_b\omega = \lambda(\mu_0 - \mu_1)$. To ensure numerical stability and to account for the singular value decomposition of S_ω in practical data scenarios, we decompose $S_\omega$ as $S_\omega = U\Sigma V^T$ where Σ is a real diagonal matrix with its diagonal elements being the singular values of $S_\omega$. Hence, $S_\omega^{-1} = V\Sigma^{-1}U^T$, which can be expressed as:

$$S_b\omega = \lambda S_\omega \omega \tag{7}$$

As a result, the solution for ω can be obtained:

$$\omega = S_\omega^{-1}(\mu_0 - \mu_1) \tag{8}$$

Expanding on this concept, we broaden the methodology to encompass multi-class classification tasks. Considering the existence of N classes, let $U = \{(x_i, y_i)\}_{i=1}^{m}$ where yi E {0,1,2,…,N}, and each class i has mi independent variables. Here, u represents the overall mean vector, while ui denotes the mean of the independent variables for the i-th class. We redefine equation (2) as the accumulation of within-class scatter matrices for each class, termed $S_\omega$, with distinctions from the definitions in (2) and (3) for $S_\omega$ and $S_b$:

$$S_\omega = \sum_{i=1}^{N} S_{\omega_i} \tag{9}$$

$$S_{\omega_i} = \sum_{x \in X_i}(x - \mu_i)(x - \mu_i)^T \tag{10}$$

Global scatter matrix:

$$S_t = S_b + S_\omega = \sum_{i=1}^{m}(x_i - u)(x_i - u)^T \tag{11}$$

The between-class scatter matrix can be calculated from equations (9) and (11) as:

$$S_b = S_t - S_\omega = \sum_{i=1}^{N} m_i(\mu_i - u)(\mu_i - u)^T \tag{12}$$

The expressions (4) and (7) above, where tr(.) denotes the trace of matrix U, can be expressed as:

$$S_b A = \lambda S_\omega A \tag{13}$$

$$max_A \frac{tr(A^T S_b A)}{tr(A^T S_\omega A)} \tag{14}$$

Expanding on the concept of Assisted Slicing LDA (SLDA), within the traditional binary LDA algorithm, data is typically constrained to a single dimension, potentially resulting in overlooked information that could be vital for exploration and leading to erroneous conclusions. To capture more pertinent data, the initial binary dataset undergoes LDA for one-dimensional reduction, yielding the reduced data set denoted as d, where Xd signifies the independent variables within d. Through an equal division of data d into segments based on Xd, these segmented portions are labeled as di, with Xdi representing the independent variables specific to each segment di. Here, φ denotes the number of equal divisions, and the response variable for data di is:

$$y_{d_i} = \frac{X_{d_i}}{\varphi} \tag{15}$$

Following the acquisition of the new multi-class dataset di, a subsequent dimensionality reduction is conducted using multi-class LDA, as outlined in equations (9)-(14). Within this context, τ signifies the dimensionality of the reduction, where τ≤N-1. Consequently, matrix A functions as the projection matrix, facilitating the transformation of di into a τ-dimensional space to decrease its dimensionality. The process of dimensionality reduction is iteratively applied to the data dj until the final dataset dj is derived. By contrast to traditional LDA methods that restrict reductions to a single dimension, this approach surmounts such limitations, thereby conserving more information and enhancing both accuracy and predictive capacity.

## 3. Data Analysis

### 3.1. Data Sources and Experimental Setup

The simulated data in Table 1 is generated using the make_classification function from the Sklearn library in Python. Real data is drawn from the UCR dataset, specifically the test and train datasets of the ECG200 dataset outlined in Table 2. The ECG200 dataset, a binary classification dataset formatted by R. Olszewski for his 2001 paper at Carnegie Mellon University [16], features X representing the electrical activity during each heartbeat and Y indicating normal heartbeats or myocardial infarctions. The ECG200 test and train datasets, named ECG2001 and ECG2002 respectively, each consist of 100 samples. The test proportion is set at 0.3 (though this may vary), leading to two distinct real data experiments. Following LDA analysis on these binary datasets, a new one-dimensional dataset is created by binning X into equidistant intervals. Subsequently, multi-class LDA analysis is employed to reduce dimensionality and achieve the ultimate goal. Please note: n/sample size, k/number of features, p/test proportion, m/number of slices, f/dimensionality reduction.

**Table 1.** Parameter Settings for Simulated Data

| Experiment | n | k | p | m | f |
|---|---|---|---|---|---|
| 1 | 1000 | 20 | 0.4 | 5 | 3 |
| 2 | 1000 | 15 | 0.5 | 6 | 4 |
| 3 | 1500 | 25 | 0.5 | 5 | 2 |

**Table 2.** Parameters for Real Data

| Real data | n | p | m | f |
|---|---|---|---|---|
| ECG2001 | 100 | 0.3 | 8 | 5 |
| ECG2002 | 100 | 0.7 | 15 | 5 |

The experiment is conducted in the Jupyter Notebook 6.5.4 (Python 3) environment within Anaconda Navigator. In this experiment, we implement both the classic LDA algorithm and the SLDA algorithm with auxiliary slicing improvement. We perform 100 iterations, running these algorithms on simulated and real data at various dimensions. The mean accuracy and standard deviation of testing and prediction results are calculated for each algorithm and displayed in Table 5. A higher mean accuracy reflects better precision, while a lower mean standard deviation indicates greater robustness. Through this analysis, we aim to demonstrate that the improved SLDA algorithm outperforms the classic LDA algorithm.

$$accuracy = \frac{number\ of\ correct\ predictions}{total\ predictions} \times 100\% \tag{16}$$

### 3.2. Comparison of Results

Based on the results from Table 3, which show simulated data, and Table 4, which present real data, it is clear that employing the classic LDA with auxiliary slicing and reusing the LDA algorithm allows for a feature space dimension beyond one dimension. This method enhances the retention of predictive variable information and minimizes the loss of valuable data. Upon reviewing the simulated data results, the spline-LDA method consistently demonstrates higher mean accuracy rates across the three datasets compared to the LDA method, with lower mean standard deviations. On average, there has been a 6.6071% increase in mean accuracy rates and a 1.2844% decrease in mean standard deviations. As a result, based on the simulated data results, the proposed binning-assisted slicing algorithm in this paper proves to be superior to the traditional LDA algorithm.

**Table 3.** Simulated Data Results

| methods | Mean1 | Mean2 | Mean3 | Std1 | Std2 | Std3 |
|---|---|---|---|---|---|---|
| LDA | 0.7529 | 0.7512 | 0.7505 | 0.0757 | 0.0687 | 0.0715 |
| Spline-LDA | 0.8203 | 0.8316 | 0.8009 | 0.0572 | 0.0582 | 0.0619 |

**Table 4.** Real Data Results

| Methods | ECG2001.mean | ECG2002.mean | ECG2001.std | ECG2002.std |
|---|---|---|---|---|
| LDA | 0.7813 | 0.7410 | 0.0531 | 0.0587 |
| Spline-LDA | 0.7980 | 0.7784 | 0.0480 | 0.0525 |

From the results of real data, the mean accuracy of the two real datasets is higher when using the spline-LDA method than when using the LDA method, and the mean standard deviation is lower. The average mean accuracy of the two overall datasets has increased by 2.7021%, and the average mean standard deviation has decreased by 0.5675%.

**Table 5.** The mean and standard deviation of the improvement effect error in 100 experiments under different dimensions

| K | LDA | | | | spline-LDA | | | |
|---|---|---|---|---|---|---|---|---|
| | ECG2001 | | ECG2002 | | ECG2001 | | ECG2002 | |
| | mean | std | mean | std | mean | std | mean | std |
| 5 | 0.6990 | 0.0506 | 0.7019 | 0.0559 | 0.7271 | 0.0404 | 0.8096 | 0.0506 |
| 6 | 0.6934 | 0.0542 | 0.7192 | 0.0545 | 0.7166 | 0.0488 | 0.7992 | 0.0544 |
| 7 | 0.6970 | 0.0559 | 0.7101 | 0.0555 | 0.7221 | 0.0497 | 0.8061 | 0.0476 |
| 8 | 0.6938 | 0.0503 | 0.7040 | 0.0529 | 0.7137 | 0.0522 | 0.7897 | 0.0473 |
| 9 | 0.6980 | 0.0540 | 0.7093 | 0.0584 | 0.7177 | 0.0515 | 0.7857 | 0.0474 |
| 10 | 0.7102 | 0.0476 | 0.7104 | 0.0539 | 0.7276 | 0.0482 | 0.7978 | 0.0424 |
| 15 | 0.7478 | 0.0493 | 0.7645 | 0.0472 | 0.7221 | 0.0483 | 0.8130 | 0.0454 |
| 20 | 0.6886 | 0.0836 | 0.7641 | 0.0449 | 0.7283 | 0.0584 | 0.8072 | 0.0464 |
| 25 | 0.6345 | 0.0930 | 0.7710 | 0.0493 | 0.7290 | 0.0494 | 0.8052 | 0.0455 |
| 30 | 0.6251 | 0.0890 | 0.7779 | 0.0477 | 0.7652 | 0.0515 | 0.8019 | 0.0480 |
| 35 | 0.6667 | 0.0777 | 0.7786 | 0.0441 | 0.7603 | 0.0538 | 0.8020 | 0.0458 |
| 40 | 0.6912 | 0.0750 | 0.7983 | 0.0492 | 0.7651 | 0.0504 | 0.8212 | 0.0449 |
| 45 | 0.7232 | 0.0672 | 0.7886 | 0.0511 | 0.7893 | 0.0417 | 0.8101 | 0.0496 |
| 50 | 0.7410 | 0.0587 | 0.7813 | 0.0531 | 0.7784 | 0.0525 | 0.7980 | 0.0480 |

Figure 1 on the left displays the results of the ECG2001 data, while Figure 1 on the right illustrates the outcomes of the ECG2002 data. These visual representations offer a more intuitive and lucid comparison, clearly demonstrating that the accuracy attained through spline-LDA analysis surpasses that achieved via LDA analysis. Furthermore, the variability range in spline-LDA, depicted by bands, is generally wider than that of LDA, except in cases where the bands overlap and exhibit equality. This observation leads to the inference that the efficacy of the spline-LDA method exceeds that of the conventional LDA approach. Notably, Figure 2 presents a noteworthy exception to this trend within the specific dimension range of 10-20, underscoring the critical significance of selecting an optimal number of slices for dimension reduction.

Upon thorough examination, it becomes apparent that whether utilizing simulated or real data, once appropriate parameters are chosen, employing the spline-LDA method for analysis yields higher accuracy and reduced standard deviation compared to utilizing the traditional LDA method. These findings strongly suggest that the spline-LDA method outperforms the traditional LDA method in binary data analysis.

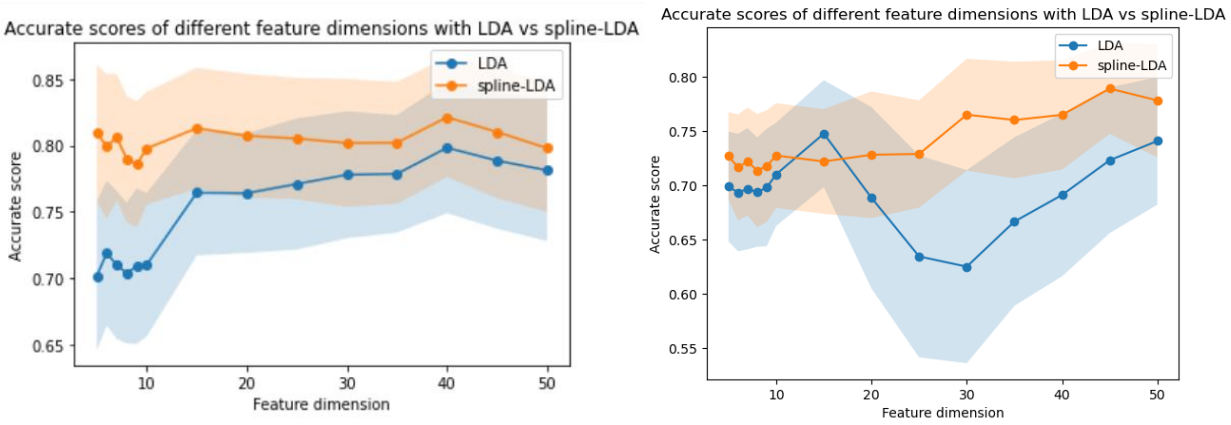


**Figure 1.** Improvement Effect Display Under Different Dimensions (ECG2001 - Left; ECG2002 - Right)

## 4. Conclusion and Prospects

This article introduces linear discriminant analysis based on equidistant auxiliary slices, which theoretically has a better capacity to capture effective feature information compared to classical linear discriminant analysis. The experimental results reveal that the former algorithm achieves higher prediction accuracy and greater robustness. Furthermore, the feature space dimension is no longer limited to one dimension, thereby enhancing the retention of predictive variable information and reducing the loss of effective information.

In terms of future research directions, the determination of the number of auxiliary slices is an area that warrants further investigation. It is worth noting that preliminary findings from simulated experiments indicate that this hyperparameter is not sensitive to classification efficiency. Additionally, the utilization of auxiliary slices requires additional exploration. While equidistant binning techniques are adequate for capturing abundant feature information when variables follow a uniform or symmetric distribution, a more flexible slicing method, such as clustering slicing rules, may be necessary if the data is skewed to the left or right. Moreover, this linear discriminant analysis based on auxiliary slices can readily be expanded to nonlinear discriminant analysis models based on reproducing kernel Hilbert spaces.

## References

[1] Yang Mingli, Fan Yugang, Li Baoyun. Research on dimensionality reduction and classification of hyperspectral images based on LDA and ELM[J]. Journal of Electronic Measurement and Instrumentation, 2020, 32 (5): 190 - 196.

[2] Yang Min, Fu Weishun, Nie Xingxin, et al. Application of hyperspectral remote sensing technology in geological environment investigation of mines [J]. Modern Mining, 2024, 40 (01): 48 - 52.

[3] Ligtenberg A, Wachowicz M, Bregt A K, et al. A design and application of a multi-agent system for simulation of multi-actor spatial planning [J]. Journal of Environmental Management 72 (2004) 1-2. 2004, 72 (1-2): 43 - 55.

[4] Deng Hongtao, Jia Qiong, Li Shaojun, Li Wei. Construction of R-vine copula model based on LASSO regression and its application in fault detection of chemical processes [J]. Journal of Chongqing University, 2023, 46 (1): 27 - 34.

[5] Long Zhenyu, Wang Changquan, Shi Lihong, Liu Yang, Xiang Xiaolong. Study on the model of CO2 solubility in formation water based on kernel ridge regression algorithm [J]. Journal of Xi'an Shiyou University (Natural Science Edition), 2023, 38 (1): 95 - 101.

[6] Zeng Shan, Chen Gang, Qi Fazhi. Research on elastic network of high-performance cloud data centers [J]. Computer Engineering and Applications, 2018, 54 (7): 89 - 95.

[7] Tang Yongbo, Gui Weihua, Peng Tao, Ouyang Wei. Prediction of dissolved gas concentration in transformer oil based on mutual information variable selection [J]. Journal of Instrumentation, 2013, 34 (7): 1492 - 1498.

[8] Tu Peng, Xiong Yiyin, Wu Yu, et al. Abnormal point suppression filtering algorithm based on chi-square test for direction finding [J]. Audio Engineering, 2023, 47 (06): 148 - 152.

[9] Lin Meijin, Dong Xuan, Hong Xiaodong, Liao Zuwei, Sun Jingyuan, Yang Yao, Wang Jingdai, Yang Yongrong. Prediction of basic physical properties of working media based on artificial neural network [J]. Petroleum Refining and Chemical Engineering, 2024, 55 (1): 180 - 188.

[10] Lv Fu, Liang Bing, Sun Weiji, Wang Yan. Prediction of gas outburst in mining face based on principal component regression analysis [J]. Journal of China Coal Society, 2012, 37 (1): 113 - 116.

[11] Jia Haiyun, Hu Lihua, Li Xiaqiao, Qu Zhihao, Wang Zhu, Chang Wei, Zhang Lei. Prediction of corrosion risk in submarine pipelines based on kernel principal component analysis algorithm[J]. Corrosion and Protection, 2023, 44 (3): 82 - 87.

[12] Xiong Jianfang, Liu Yao, Qiao Fu, Liu Zhongyan, Jiang Wei, Lu Liqiong. Detection of diarrhetic shellfish toxins based on LDA dimensionality reduction method [J]. Sensors and Microsystems, 2023, 42 (05): 25 - 28.

[13] Wang Minghe, Zhang Erhua, Tang Zhenmin, Xu Hao. Endpoint detection method for speech signals based on Fisher linear discriminant analysis [J]. Journal of Electronics and Information Technology, 2015, 37 (6): 1343 - 1349.

[14] Wang Yiduo. Application research on face image recognition based on the combination of PCA, LDA, and SVM [D]. Lanzhou Jiaotong University, 2024. DOI: 10.27205/d.cnki.gltec.2023.000893.

[15] Wang Yiduo, Lv Weidong, Hu Chencheng, et al. Research on face image recognition based on PCA, LDA, and LR fusion algorithm [J]. Scientific and Technological Innovation, 2022 (22): 72 - 75.

[16] Olszewski R T. Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data [J]. Time Series Data Ph.D. Dissertation Carnegie Mellon University, 2001.