

Article

Near-Infrared Spectroscopy Combined with Fuzzy Improved Direct Linear Discriminant Analysis for Nondestructive Discrimination of Chrysanthemum Tea Varieties

Jiawei Zhang ¹, Xiaohong Wu ^{2,3,*} , Chengyu He ², Bin Wu ^{4,*}, Shuyu Zhang ¹ and Jun Sun ²

¹ Mengxi Honors College, Jiangsu University, Zhenjiang 212013, China; 3210608050@stmail.ujs.edu.cn (J.Z.); 3210608017@stmail.ujs.edu.cn (S.Z.)

² School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China; 2222107003@stmail.ujs.edu.cn (C.H.); sun2000jun@ujs.edu.cn (J.S.)

³ High-Tech Key Laboratory of Agricultural Equipment and Intelligence of Jiangsu Province, Jiangsu University, Zhenjiang 212013, China

⁴ Department of Information Engineering, Chuzhou Polytechnic, Chuzhou 239000, China

* Correspondence: wxh419@ujs.edu.cn (X.W.); wubin@chzc.edu.cn (B.W.)

Abstract: The quality of chrysanthemum tea has a great connection with its variety. Different types of chrysanthemum tea have very different efficacies and functions. Moreover, the discrimination of chrysanthemum tea varieties is a significant issue in the tea industry. Therefore, to correctly and non-destructively categorize chrysanthemum tea samples, this study attempted to design a novel feature extraction method based on the fuzzy set theory and improved direct linear discriminant analysis (IDLDA), called fuzzy IDLDA (FIDLDA), for extracting the discriminant features from the near-infrared (NIR) spectral data of chrysanthemum tea. To start with, a portable NIR spectrometer was used to collect NIR data for five varieties of chrysanthemum tea, totaling 400 samples. Secondly, the raw NIR spectra were processed by four different pretreatment methods to reduce noise and redundant data. Thirdly, NIR data dimensionality reduction was performed by principal component analysis (PCA). Fourthly, feature extraction from the NIR spectra was performed by linear discriminant analysis (LDA), IDLDA, and FIDLDA. Finally, the K-nearest neighbor (KNN) algorithm was applied to evaluate the classification accuracy of the discrimination system. The experimental results show that the discrimination accuracies of LDA, IDLDA, and FIDLDA could reach 87.2%, 94.4%, and 99.2%, respectively. Therefore, the combination of near-infrared spectroscopy and FIDLDA has great application potential and prospects in the field of nondestructive discrimination of chrysanthemum tea varieties.

Keywords: chrysanthemum tea; near-infrared spectroscopy; dimensionality reduction; feature extraction



Citation: Zhang, J.; Wu, X.; He, C.; Wu, B.; Zhang, S.; Sun, J. Near-Infrared Spectroscopy Combined with Fuzzy Improved Direct Linear Discriminant Analysis for Nondestructive Discrimination of Chrysanthemum Tea Varieties. *Foods* **2024**, *13*, 1439. <https://doi.org/10.3390/foods13101439>

Academic Editor: Charis R Theocharis

Received: 12 April 2024

Revised: 2 May 2024

Accepted: 2 May 2024

Published: 7 May 2024



Copyright: © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Chrysanthemum tea is a valuable flower crop in China, and it is widely used in traditional Chinese medicine for its high medicinal value [1]. It has many beneficial chemical components, including flavonoids, polysaccharides, and unsaturated fatty acids [2], as well as luteolin and luteoloside [3]. It has been proven that chrysanthemum tea can be used to fight cancer, inflammation, and obesity, protect the liver and kidneys, and guard against liver-fire hyperactivity syndrome [4]. The quality and efficacy of chrysanthemum tea are closely related to its geographical origin [5]. As a result, the market is susceptible to fraudulent substitutes of lower value, which would be detrimental to the health and interests of consumers. Therefore, it is crucial to develop a quick and effective method to identify the chrysanthemum tea varieties.

In recent years, many researchers have actively explored some identification methods for chrysanthemum tea varieties. For example, Luo et al. applied gas chromatography–mass spectrometry and olfactometry and an electronic nose combined with principal

component analysis (PCA) to identify the geographical origins of Chinese chrysanthemum flower teas [6]. DNA barcoding analysis based on *PsbA-trnH*, *matK*, and *trnI* has been proven to be effective in the identification of chrysanthemum varieties living in different geographic populations [7]. Hao et al. successfully classified nine geographically distinct chrysanthemum varieties using laser-induced breakdown spectroscopy and chemometrics [8]. However, these techniques are complex in terms of data preprocessing and are relatively costly and slow, so they are unsuitable for the rapid non-invasive detection of chrysanthemum tea varieties.

Currently, near-infrared (NIR) spectroscopy technology is developing rapidly due to the advantages of miniaturized NIR spectrometers [9], and it has good application prospects in the field of nondestructive food detection with its advantages of simplicity, efficiency, and low cost [10–14]. Nowadays, the widespread application of NIR spectroscopy technology appears in the agriculture and food industry [15–22], chemical and material science [23], pharmaceutical industry, and many other fields [24]. For example, Ma et al. combined NIR spectroscopy with partial least squares and an artificial neural network for the rapid detection of sugarcane stalk bending characteristics [25]. Wu et al. utilized a novel fuzzy feature extraction algorithm to process the NIR data of Chunmee tea and established an effective classification model [26]. NIR spectroscopy was combined with chemometrics to identify different tea varieties, and the classification accuracy reached 98.33% in [27]. Chen et al. designed a classification method using NIR spectroscopy and a random forest algorithm to accurately classify tea quality [28].

NIR spectra are characterized by high dimensionality, overlap, and nonlinearity, so the accuracy is low if the NIR spectra are classified directly. A common solution is to first pretreat the NIR spectra and then perform feature extraction on the spectra. Feature extraction algorithms are important for solving small-sample-size (SSS) problems [29]. When linear discriminant analysis (LDA) processes NIR spectra with high dimensionality, SSS problems always arise. In recent years, many approaches have been proposed for solving this SSS problem [30–32], and one of them is direct LDA (DLDA). High-dimensional spectral data usually need to be downsampled by PCA, but some feature information may be lost in this process. The DLDA algorithm can avoid this problem as it can directly extract features from high-dimensional data [33]. But DLDA discards the zero space of the interclass scattering matrix in its computation, and the zero space may have useful information for categorization. This reduces the classification accuracy, and improved DLDA (IDLDA) was proposed to solve the drawbacks of DLDA [34]. However, the classification performance of the IDLDA algorithm may suffer from data overlap. To solve this problem, fuzzy IDLDA (FIDLDA) is proposed in this study to extract the NIR spectra of chrysanthemum tea.

Zadeh et al. introduced fuzzy set theory, which could be a good solution to the data overlap problem [35]. Some feature extraction algorithms have been combined with fuzzy set theory for spectral information extraction. Fuzzy improved null LDA (FiNLDA) was employed to attain the near-infrared spectral discrimination of milk, and an effective model for milk brand discrimination was developed [36]. Fuzzy uncorrelated discriminant transformation (FUDT) was utilized to process the NIR spectrum of milk and achieved a classification accuracy of 98.67% in identifying the geographical sources of milk [37]. Therefore, it is feasible to combine a fuzzy algorithm, feature extraction methods, and NIR spectroscopy for discriminative information extraction. In this experiment, a classification model using NIR spectroscopy and FIDLDA was designed for the nondestructive discrimination of chrysanthemum tea varieties.

2. Materials and Methods

2.1. Sample Preparation

Five types of chrysanthemum tea, including chuju (CJ), hangbaiju (HBJ), huaiju (HJ), huangshangongju (HSGJ), and wuyuanhuangju (WYHJ), originated from Chuzhou, Tongxiang, Jiaozuo, Huangshan, and Wuyuan in China, respectively. The distinguishing differ-

ences between these types were the contents of several functional components, which are shown in Table 1 [38,39]. The tea had a golden or light brown appearance, a clear odor, a good even size, no mold, and intact inflorescences. To keep them dry and cool, they were stored in sealed food preservation bags until NIR analysis was performed.

Table 1. The contents of several functional components of five varieties of chrysanthemum tea (%).

	CJ	HBJ	HJ	HSGJ	WYHJ
Flavone	16.08	14.22	10.25	13.39	4.13
Soluble sugar	23.18	16.35	19.04	23.05	15.01
Chlorogenic acid	3.75	3.47	2.16	1.09	-

A total of 400 samples were used for the spectral data collection. The same number of samples were procured for each category, and they were divided into five groups according to their varieties, so each group had 80 samples. Subsequently, all chrysanthemum tea samples were partitioned into a training set and a test set based on a specific ratio in the discriminant experiment. Spectral acquisition was performed at about 20 °C and 60% relative humidity.

2.2. NIR Spectra Collection

The NIR spectra of the samples were collected in Hadamard mode using a portable spectrometer, NIR-M-F1-C (Shenzhen Puyan Internet Technology, Shenzhen, China). Using a spectrometer in Hadamard mode can improve the signal-to-noise ratio (SNR), and a higher optical energy can be captured by an InGaAs detector. The spectrometer operated at wavelengths from 900 to 1700 nm. The ratio of signal to noise and the optical resolution were set to 6000:1 and 10 nm for the acquisition process, respectively. The spectrometer was equipped with a humidity and temperature sensor. Each spectrum consisted of 400 data points with a wavelength interval of 800 nm.

The scans were performed 8 times, and each scan had an exposure time of 0.625 ms. In this experiment, non-invasive reflectance detection was utilized. Figure 1 displays the raw NIR spectra of the chrysanthemum tea samples.

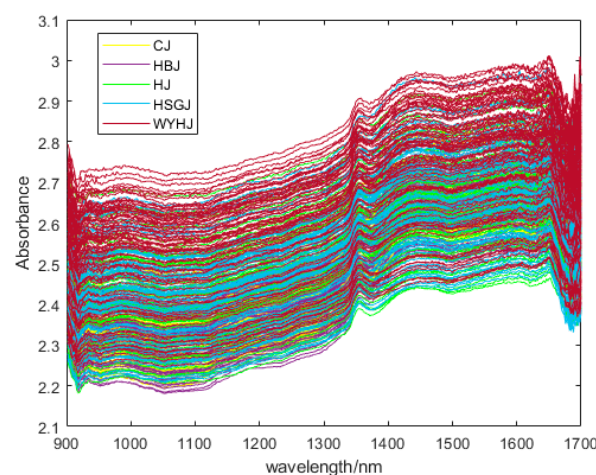


Figure 1. The raw spectra of chrysanthemum tea samples.

2.3. Preprocessing

By using a NIR-M-F1-C spectrometer to analyze the samples, the raw NIR spectra of the chrysanthemum tea varieties could be obtained. However, the direction of light changes due to the effect of small inhomogeneity on the surface of chrysanthemum tea when collecting spectral data, and the noise-generated scatter may affect the raw NIR spectra [40], and, therefore, preprocessing the spectral data is important for the subsequent

processing of the NIR spectra. In this experiment, several preprocessing algorithms were applied to pretreat the NIR spectral data, including standard normal variation (SNV), multiplicative scattering correction (MSC), Savitsky–Golay (SG) filtering [41], and mean centering (MC), which improved the spectral data. Combined pretreatment methods were also tried, like SG + MSC and SG + SNV, but the effects were not very satisfactory in this experiment.

MSC can reduce the negative effects of uneven particle sizes, optical path variations, varying sample compactness, and spectral noise and bias. SNV can correct scattering effects and baseline shifts in spectral data and reduce inter-sample variation. SG filtering can also remove the spectral noise and enhance the smoothness of spectral data. MC can improve the comparability between variables, amplify weak signals, and reduce the collinearity between spectral data. Figure 2 shows the NIR spectral data preprocessed by the four single pretreatment methods.

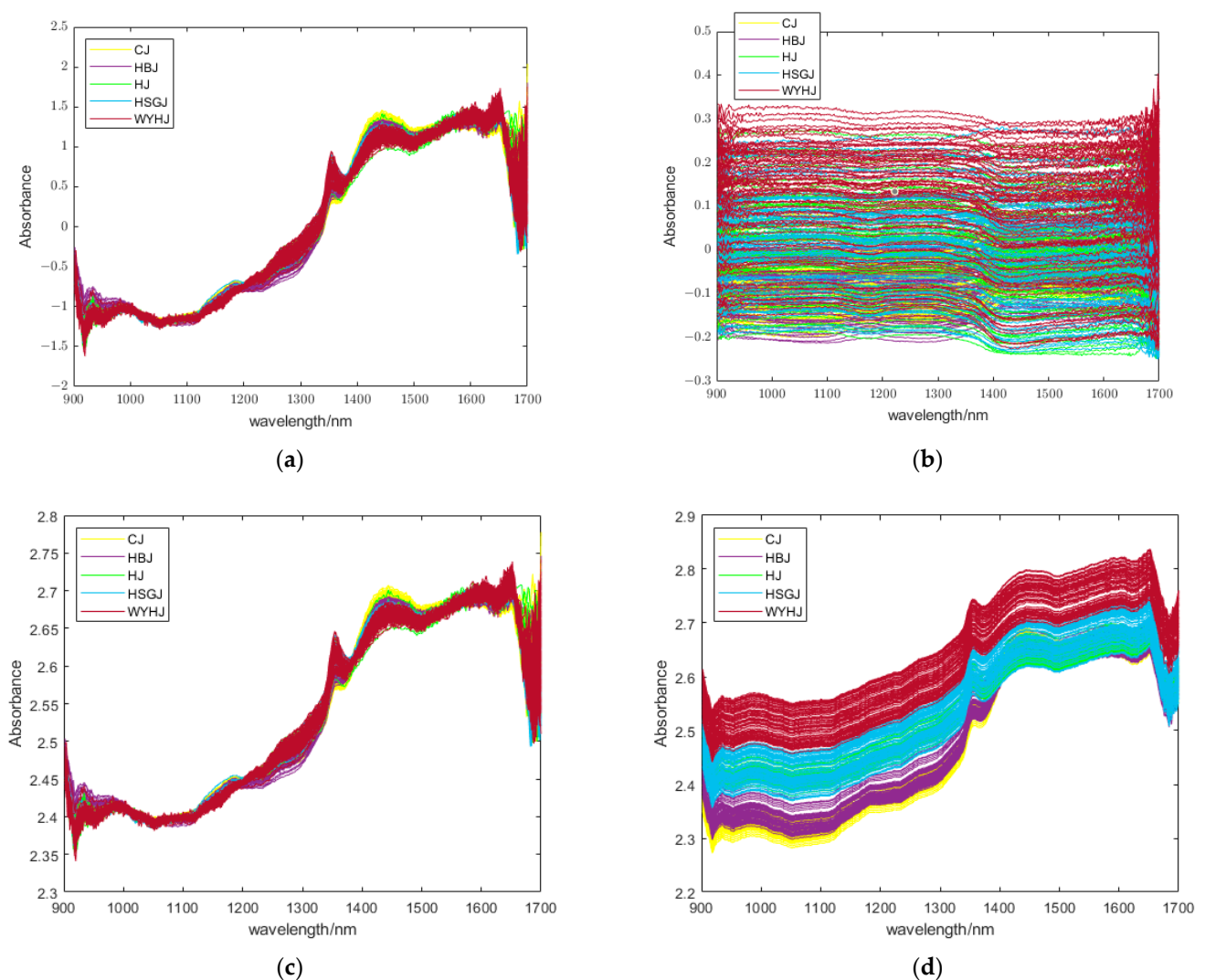


Figure 2. NIR spectra of chrysanthemum tea samples preprocessed by four pretreatment methods: (a) SNV, (b) MC, (c) MSC, and (d) S-G filtering. SNV, standard normal variation; MC, mean centering; MSC, multiplicative scattering correction; S-G, Savitsky–Golay.

2.4. Data Processing Algorithms

2.4.1. Principal Component Analysis

The collected infrared spectra of the chrysanthemum tea samples had a dimension of 400, which contained a large amount of redundant information and noisy data, which

may increase the computational cost and decrease the classification accuracy. Therefore, to acquire high-quality spectral data, it was necessary to perform dimensionality reduction and redundancy removal on the pretreated spectral data. PCA is one of the commonly used methods, which operates by identifying a collection of orthogonal eigenvectors that make their corresponding eigenvalues as large as possible, and dimensionality can be reduced by choosing a meaningful set of eigenvectors. Because these eigenvectors correspond to larger variance, the most significant information in the raw data can be retained while reducing the dimensionality.

2.4.2. LDA

LDA is a classical machine learning algorithm that is utilized for both the extraction of features and reducing data dimensionality [42]. LDA can reduce the complexity of spectral data by finding the most representative features in the data. In order to be able to distinguish between different classes of data, the primary objective of the LDA algorithm is to determine the ideal projection direction to make the inter-class spacing as large as possible and minimize the intra-class spacing as much as possible.

2.4.3. IDLDA

IDLDA is another important technique for the extraction of features in the widespread use of small-sample problem solving [34]. The steps of IDLDA are described as follows (Algorithm 1).

Algorithm 1: The Steps of IDLDA Algorithm

Step 1. Build the matrices S_t , S_b , and S_w ;
 Step 2. Singular value decomposition of S_w as $S_w = U_w D_w^2 U_w^T$;
 Step 3. Find $\partial = \max(\text{diag}(D_w))$ and build $D_\partial = \partial I_{d \times d} - D_w$;
 Step 4. Use D_∂ and U_w to diagonalize S_{fb} , as $D_\partial U_w^T S_b U_w D_\partial = F \Sigma^2 F^T$, where $F = [F_r, F_n]$ such that F_r corresponds to the range space of S_b and F_n corresponds to the null space of S_b ;
 Step 5. Calculate the transformation matrix $W_{IDLDA} = U_w D_\partial F_r$, and project samples into the feature space.

In Step 1, S_t represents the total scatter matrix; S_b represents the between-class matrix; and S_w represents the within-class matrix. They are listed as follows:

$$S_t = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T \quad (1)$$

$$S_b = \sum_{j=1}^c n_j (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})^T \quad (2)$$

$$S_w = \sum_{j=1}^c \sum_{x_i \in c_j} (x_i - \bar{x}_j)(x_i - \bar{x}_j)^T \quad (3)$$

where n represents the sample number; n_j is the sample number in the j th class; and c denotes the number of the variety.

2.4.4. FIDLDA

FIDLDA is a novel fuzzy DLDA algorithm generated based on the combination of fuzzy set theory and the IDLDA algorithm. The specific algorithm execution steps are shown as follows (Algorithm 2).

Algorithm 2: The Steps of FIDLDA Algorithm

Step 1. Build the matrices S_{ft} , S_{fb} , and S_{fw} ;
 Step 2. Singular value decomposition of S_{fw} as $S_{fw} = U_{fw} D_{fw}^2 U_{fw}^T$;
 Step 3. Find $\partial_f = \max(\text{diag}(D_{fw}))$ and build $D_{\partial_f} = \partial_f I_{d \times d} - D_{fw}$;
 Step 4. Use D_{∂_f} and U_{fw} to diagonalize S_{fb} , as $D_{\partial_f} U_{fw}^T S_{fb} U_{fw} D_{\partial_f} = F_f \Sigma^2 F_f^T$, where $F_f = [F_{fr}, F_{fn}]$ such that F_{fr} corresponds to the range space of S_{fb} and F_{fn} corresponds to the null space of S_{fb} ;
 Step 5. Calculate the transformation matrix $W_{FIDLDA} = U_{fw} D_{\partial_f} F_{fr}$, and project samples into the feature space.

In Step 1, S_{ft} represents the fuzzy total scatter matrix; S_{fb} represents the fuzzy between-class matrix; and S_{fw} represents the fuzzy within-class matrix. They can be calculated as follows:

$$S_{ft} = \sum_{j=1}^c \sum_{i=1}^n u_{ij}^m (x_i - \bar{x})(x_i - \bar{x})^T \quad (4)$$

$$S_{fb} = \sum_{j=1}^c \sum_{i=1}^n u_{ij}^m (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})^T \quad (5)$$

$$S_{fw} = \sum_{j=1}^c \sum_{i=1}^n u_{ij}^m (x_i - \bar{x}_j)(x_i - \bar{x}_j)^T \quad (6)$$

where m is the fuzzy weight index, and u_{ij} represents the fuzzy membership (FM) value indicating the belongingness of the j th sample data to the i th class. For the calculation formula for u_{ij} , see Formula (4) in ref. [26].

2.4.5. KNN

K-nearest neighbor (KNN) is one of the common classifiers and was used for the categorization of the chrysanthemum varieties in this experiment. As a supervised machine learning algorithm, its basic principle can be described as follows: Firstly, calculate the distances between a given test sample and each training sample. Then, find the K training samples with the closest distance, and, finally, predict the test sample class based on the class that occurs most frequently among the K samples.

PCA + LDA, PCA + IDLDA, and PCA + FIDLDA were used for extracting the discriminant information from the chrysanthemum tea samples' spectra, and then the chrysanthemum tea varieties were classified by the KNN algorithm. The identification result of KNN is strongly related to the value of K. Therefore, the appropriate K was selected by computing the prediction accuracy under variant K values.

2.5. Software

In this study, the mathematics software we utilized was MATLAB (The Mathworks Inc., Natick, MA, USA) 2019a.

3. Results

3.1. NIR Spectral Analysis

The NIR spectra of the chrysanthemum tea samples in this experiment were within the wavelength range of 900–1700 nm. The original NIR spectra of the samples are shown in Figure 1, and the NIR spectra encompassed a large amount of information about molecular bonding and characteristic functional groups, such as C-H, O-H and N-H, which are likely to be associated with flavonoids, amino acids, and polysaccharides [43]. The absorption regions observed in the NIR spectra primarily originated from the band of groups containing hydrogen and its overtones. In Figure 1, the absorption bands are mainly concentrated in three regions, 1350 nm to 1370 nm, 1400 nm to 1470 nm, and 1630 nm to 1660 nm, respectively. From 920 nm to 940 nm, weak absorption bands can also be observed. The

absorbance of the chrysanthemum tea dramatically changes after 1300 nm and reaches a peak at 1354 nm. This phenomenon may be related to the stretching vibration of the C-H and O-H groups in the amino acids and polysaccharides [44]. The absorption bands from 1400 nm to 1470 nm are ascribed to the first overtone of the O-H stretching vibrations alongside the N-H band [37]. The peak at 1652 nm is due to the C-H stretching first overtone of $-\text{CH}_2$ and the binary combination bands involving C-H stretching modes [45,46].

3.2. Spectral Preprocessing

Figure 2 shows the NIR spectra of the chrysanthemum tea samples by different preprocessing methods. In this study, four single preprocessing methods were utilized: SNV, SG filtering, MSC, and MC, as well as two combined pretreatment methods, namely, SG + SNV and SG + MSC. The NIR spectra preprocessed by MC have no evident troughs and peaks in Figure 2b compared with the other spectra. We conducted experiments using six different pretreatment methods on the NIR spectra. Among them, it was observed that SG filtering had the best preprocessing effect, while the accuracy of the two mixed pretreatment methods combined with the proposed system was only about 80%, so we chose SG filtering as the preprocessing method in this study.

3.3. Dimensionality Reduction by PCA

After preprocessing, the spectral data contained some redundant information and had high dimensionality. Such data were not conducive to the classification of the chrysanthemum tea varieties. Hence, it was essential to use PCA to extract the principal components (PCs) and mitigate redundant information. In this study, the total contribution of the first six PCs exceeded 99.98%, which proved that they retained the vast majority of the features in the NIR spectral data and eliminated a substantial quantity of redundant information. To be specific, the first six eigenvalues were listed as follows: $\lambda_1 = 552.9266$, $\lambda_2 = 25.0565$, $\lambda_3 = 0.3454$, $\lambda_4 = 0.1449$, $\lambda_5 = 0.0371$, and $\lambda_6 = 0.0182$. Hence, the 400-dimensional NIR spectra were projected into a six-dimensional feature space. Since the total contribution of the first three PCs reached 99.9%, a three-dimensional feature space was established to observe the distribution of the spectral data of the different kinds of chrysanthemum tea samples. Due to the four preprocessing methods used in this experiment, the spectral data obtained after the PCA processing were different. The distribution of the spectra processed by SG filtering and PCA in the three-dimensional feature space is shown in Figure 3, and it can be seen that the clustering of the data of the different kinds of samples is distinct, thus proving that PCA can effectively improve NIR spectral data. In addition, it is easy to see that the data after dimensionality reduction using PCA alone were still not good enough to identify the chrysanthemum tea samples, so more feature information needed to be extracted.

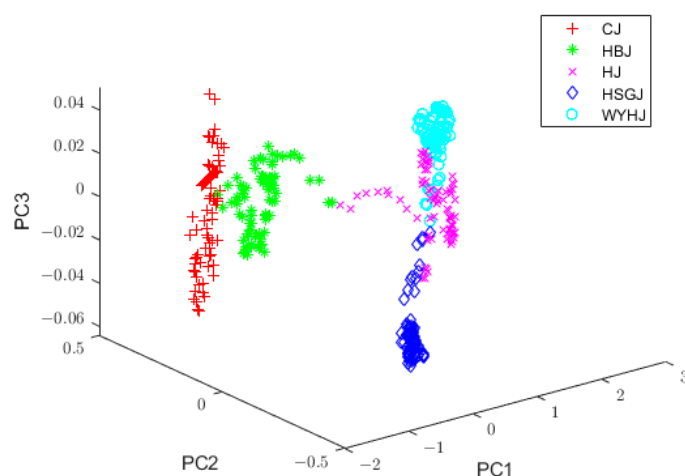


Figure 3. Distribution plot of vectors with PC1, PC2, and PC3 after SG filtering + PCA. PC1, PC2, and PC3, the first three principal components; SG, Savitsky–Golay; PCA, principal component analysis.

The subsequent sections cover the discussion of classification models, namely, PCA + LDA, PCA + IDILDA, and PCA + FIDILDA, applied to different chrysanthemum tea varieties.

3.4. Extraction of Features by LDA

Following the PCA dimensionality reduction process, the 400 chrysanthemum tea samples were partitioned into a training set, which comprised 55 training samples for each variety (totaling 275 samples), and a test set containing 25 test samples for each variety (totaling 125 samples). The LDA algorithm was utilized for feature information extraction from the training set, and, subsequently, the test samples were projected onto the eigenvectors generated by the LDA. The rank of the inter-class scatter matrix was maximized by the number of classes minus one, so the number of eigenvectors and eigenvalues was four. Those four eigenvalues were listed as follows: $\lambda_1 = 64.5485$, $\lambda_2 = 16.2678$, $\lambda_3 = 12.1063$, and $\lambda_4 = 4.9531$. The six-dimensional feature data were projected onto the first three eigenvectors (DV1, DV2, and DV3) of the LDA, and the three-dimensional data distribution is shown in Figure 4. It is clear that PCA + LDA could distinguish the sample varieties to some extent, but there were two varieties of chrysanthemum tea sample data (HJ and HBJ) that overlapped with each other, and its classification accuracy was 87.2%. Therefore, a more effective feature extraction algorithm was imperative to improve the accuracy of the sample classification.

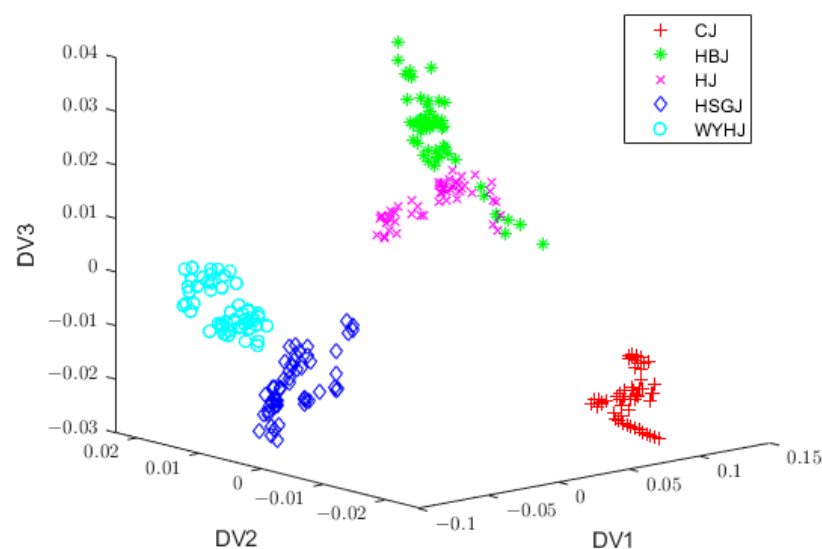


Figure 4. Three-dimensional data distribution by S-G filter + PCA + LDA. S-G, Savitsky–Golay; PCA, principal component analysis; LDA, linear discriminant analysis.

3.5. Discriminant Feature Extraction by IDLDA

After IDLDA extracted feature discriminative vectors from the six-dimensional data, it produced four discriminative vectors after processing the 275 training sets, and the PCA-processed data of the training samples were projected onto the first three discriminative vectors (DV1, DV2, and DV3). Figure 5 shows the scores plot of three discriminant eigenvectors of the IDLDA, and it can be seen that each sample datum had a more pronounced boundary profile. However, there was still some overlap between the two samples (HJ and WYHJ). Nevertheless, compared with the PCA + LDA algorithm, its classification accuracy was improved to 94.4%.

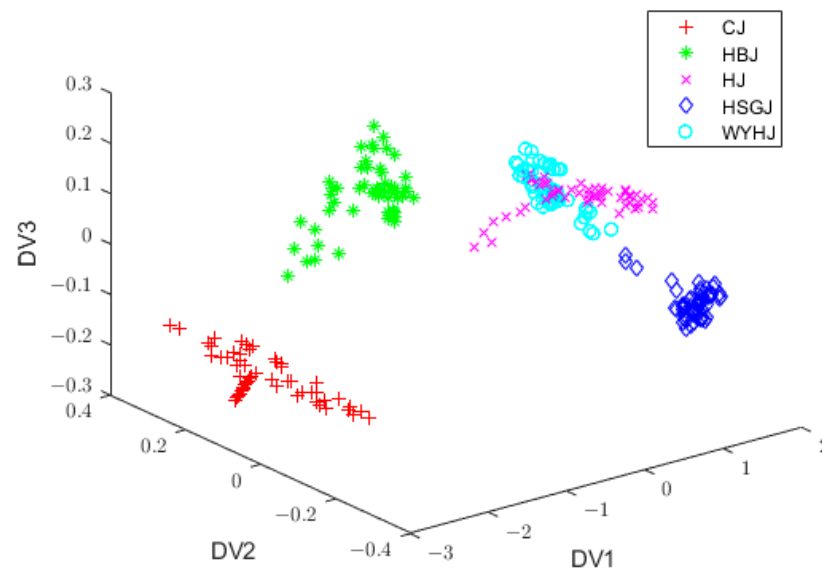


Figure 5. Three-dimensional data distribution by SG filtering + PCA + IDLDA. SG, Savitsky–Golay; PCA, principal component analysis; IDLDA, improved direct linear discriminant analysis.

3.6. Feature Extraction by FIDLDA

FIDLDA performed feature extraction to transform the data into a feature space where the data were correctly classified. The results show that FIDLDA could address the limitations of IDLDA and improve the classification accuracy. All of the parameters related to FIDLDA were listed: the fuzzy weighting factor $m = 1.6$ and the number of sample varieties $c = 5$. The initial cluster center was represented by the mean of each variety of the chrysanthemum tea samples, and it is shown in Equation (7).

$$V^{(0)} = \begin{pmatrix} v_1^{(0)} \\ v_2^{(0)} \\ v_3^{(0)} \\ v_4^{(0)} \\ v_5^{(0)} \end{pmatrix} = \begin{pmatrix} -1.1010 & 0.2820 & 0.0042 & -0.0290 & 0.0032 & -2.6439 \\ -1.1897 & 0.0823 & 0.0063 & 0.0344 & 0.0035 & -4.7939 \\ 0.0358 & -0.3044 & 0.0247 & -0.0069 & -0.0121 & 0.0040 \\ 0.3013 & -0.1799 & -0.0493 & -0.0016 & 0.0029 & -0.0021 \\ 1.6900 & 0.1031 & 0.0159 & -4.3758 & 0.0070 & -0.0011 \end{pmatrix}_{5 \times 6} \quad (7)$$

Figure 6 displays the initial FM values, where the horizontal coordinate represents the chrysanthemum tea training sample and the vertical coordinate represents the FM values. Each little figure represents one chrysanthemum tea variety, namely, CJ, HBJ, HJ, HSGJ, and WYHJ, so there is a total of five little figures. If the FM degree of the k th sample was found to be the highest one within the j th category, it could be determined that the k th sample was attached to the corresponding j th category. The FM values of the HJ and HSGJ samples partially overlapped, which was due to calculating the FM values with the means of the sample data. Figure 3 shows that the score plots of HJ and HSGJ overlapped after PCA pretreatment, indicating that the means of two sample varieties were near, which negatively affected the calculation of the FM degrees.

Figure 7 displays the three-dimensional data distribution by SG filtering + PCA + FIDLDA. It can be seen that the samples of HJ and HSGJ were well separated, which indicated that FIDLDA significantly improved the recognition ability compared with LDA and IDLDA.

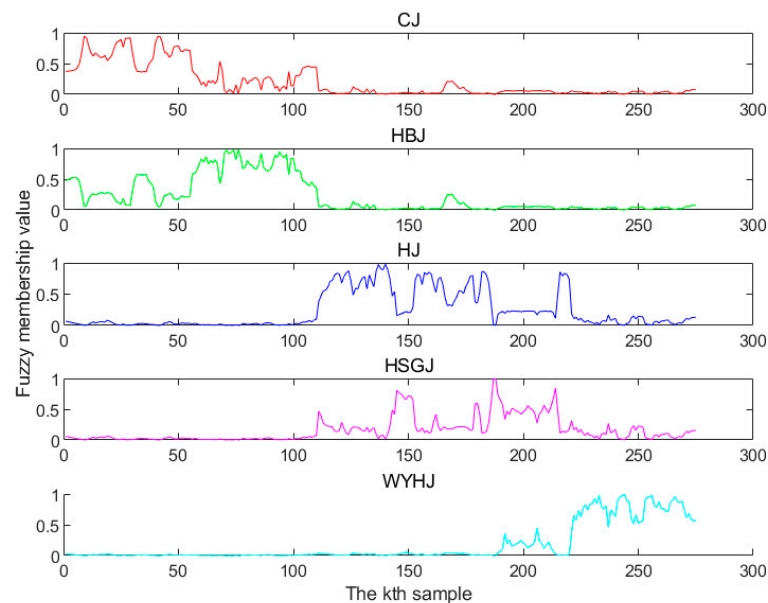


Figure 6. Initial fuzzy membership degrees.

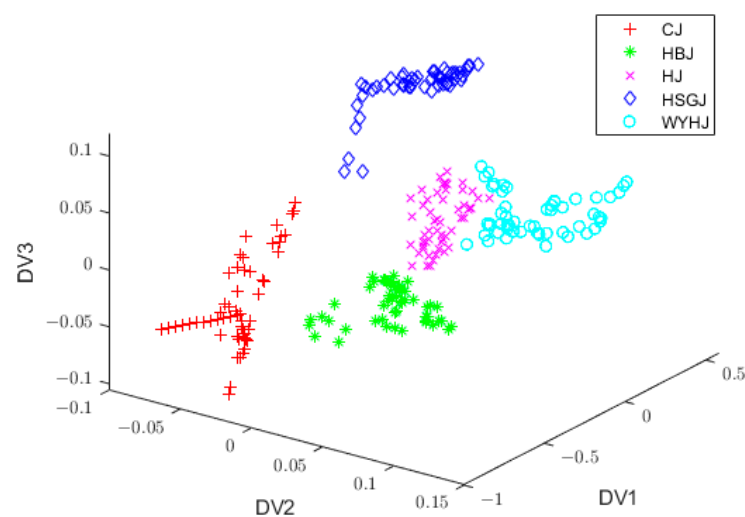


Figure 7. Three-dimensional data distribution by SG filtering + PCA + FIDLDA. S-G, Savitsky–Golay; PCA, principal component analysis; FIDLDA, fuzzy improved direct linear discriminant analysis.

3.7. Classification Results of KNN

The KNN algorithm was employed as a classifier for the identification of the chrysanthemum tea varieties in the data after using the feature extraction algorithms. Since the K-value can affect the classification accuracy of KNN, in order to obtain the K-value for optimal identification accuracy, we employed KNN using different K-values (1, 3, 5, 7, 9, 11, and 13) with three feature extraction methods (LDA, IDLDA, and FIDLDA) for the calculation of the prediction accuracy. The training sample set consisted of 275 samples, and the test sample set comprised 125 samples. The classification accuracies with different K-values are shown in Figure 8. In comparison with LDA and IDLDA, the FIDLDA algorithm had the highest classification accuracy of 99.2% when the value of K was nine. Thus, it was proved that the FIDLDA algorithm combined with the KNN classifier had a great classification ability.

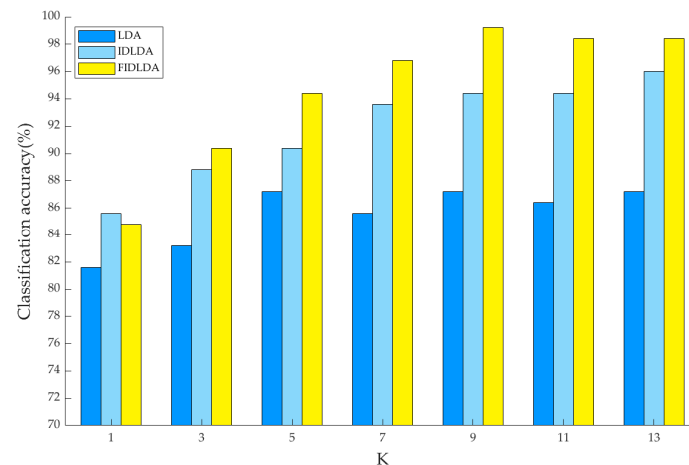


Figure 8. Classification accuracies of LDA, IDLDA, and FIDLDA with different K-values. LDA, linear discriminant analysis; IDLDA, improved direct linear discriminant analysis; FIDLDA, fuzzy improved direct linear discriminant analysis.

4. Discussion

Firstly, the NIR spectra of chrysanthemum tea samples were obtained by a portable spectrometer, and then SG filtering was used for noise reduction, PCA for data dimensionality reduction, and LDA, IDLDA, and FIDLDA for feature information extraction. Finally, KNN was utilized as a classifier to categorize the sample varieties. In Figure 8, it is obvious that using different feature extraction algorithms obtained different classification accuracies. When the traditional LDA algorithm was employed to extract features, the classification accuracy was below 90%. In comparison, when the FIDLDA was applied as a feature extraction algorithm, the highest identification accuracy achieved a value of 99.2%.

The fuzzy weight index m has a strong correlation with the feature extraction effect of FIDLDA. We conducted the experiments using different values of m and recorded the classification accuracies accordingly. In particular, the value of m could not be lower than 1, so m ranged between 1.2 and 5.0. Figure 9 shows the classification accuracy of FIDLDA with different m -values, and it reached the highest classification accuracy when the value of m was 1.6.

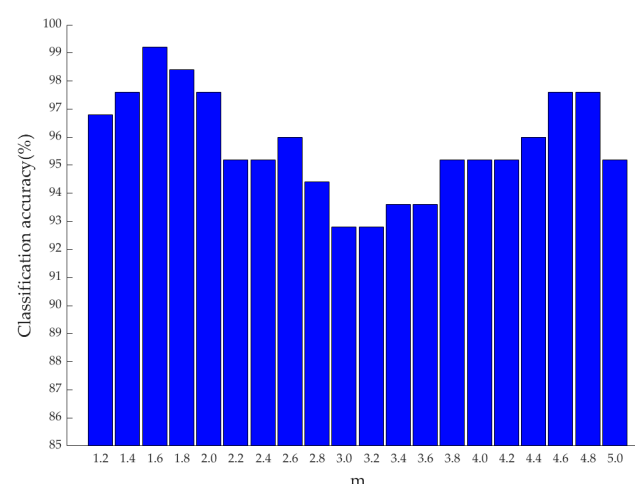


Figure 9. Classification accuracies of FIDLDA with different values of fuzzy weight index m . FIDLDA, fuzzy improved direct linear discriminant analysis.

The data quantities in the training set and test set also affect the classification accuracy of the classification model. Other things being equal, we observed the classification accuracies obtained from three different combinations of training and test samples. Table 2

shows the categorization accuracies using LDA, IDLDA, and FIDLDA with different data quantities for the training and test sets for the chrysanthemum tea varieties. Table 2 shows that FIDLDA produced higher classification accuracies than LDA and IDLDA. When the data quantities for the training set and test set were 275 and 125, respectively, the FIDLDA algorithm reached the highest accuracy of 99.20%.

Table 2. Classification accuracies of LDA, IDLDA, and FIDLDA with different n_{training} and n_{test} combinations (%).

n_{training}	n_{test}	LDA	IDLDA	FIDLDA
250	150	87.33	94.00	95.33
275	125	87.20	94.40	99.20
300	100	90.00	91.00	95.00

Abbreviations: LDA, linear discriminant analysis; IDLDA, improved direct linear discriminant analysis; FIDLDA, fuzzy improved direct linear discriminant analysis.

To show the superiority of the FIDLDA model for chrysanthemum tea varieties, the FIPLDA-KNN model and the FIPLDA-SVM model [44], which have been applied for chrysanthemum tea identification, were used for comparison. When the S-G filtering algorithm was also used for preprocessing, and PCA was used for dimensionality reduction, the FIPLDA-KNN model achieved the maximum classification accuracy of 98.33% when the fuzzy weight coefficient was 2.7 and K was 7, while the FIPLDA-SVM model had the maximum classification accuracy of 90.83%. The specific results can be found in ref. [44]. In contrast, the FIDLDA model reached a classification accuracy of 99.2% in the identification of chrysanthemum tea. Therefore, the proposed nondestructive discrimination system for chrysanthemum tea varieties in this study had a better performance than the models used in the previous research.

5. Conclusions

In order to be able to quickly, non-destructively, and effectively discriminate chrysanthemum tea varieties, a classification system combining NIR spectroscopy with the FIDLDA algorithm was presented in this study. The proposed FIDLDA algorithm is a unique fusion of the fuzzy set and the IDLDA algorithm, and it provides a novel approach for extracting features from chrysanthemum tea spectral data after PCA reduces the data dimensionality. At first, the NIR spectra of the chrysanthemum tea samples were acquired by a portable spectrometer. Secondly, SG filtering, PCA, LDA, IDLDA, and FIDLDA were utilized for data denoising, dimensional reduction, and feature extraction from the data, respectively. Finally, the KNN algorithm was used to classify the chrysanthemum tea varieties.

The results show that the FIDLDA algorithm had the highest accuracy in the classification of chrysanthemum tea varieties compared with the LDA and IDLDA algorithms. This study illustrates that the combination of NIR spectroscopy and the FIDLDA algorithm has great potential for the nondestructive discrimination of chrysanthemum tea varieties.

Author Contributions: Conceptualization, B.W. and X.W.; methodology, X.W. and C.H.; software, C.H., B.W. and J.Z.; validation, B.W., J.S. and X.W.; formal analysis, S.Z. and C.H.; investigation, J.Z. and S.Z.; resources, C.H. and X.W.; data curation, J.S.; writing—original draft preparation, J.Z. and S.Z.; writing—review and editing, X.W. and B.W.; visualization, J.S.; supervision, J.S. and B.W.; project administration, X.W. and B.W.; funding acquisition, B.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Major Natural Science Research Projects of the Colleges and Universities in Anhui Province (2022AH040333), the Youth and Middle-Aged Teachers Cultivation Action Project in Anhui Province (JNFX2023136), and the Undergraduate Innovation and Entrepreneurship Training Program of Jiangsu Province (202310299363X).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article and supplementary materials, further inquiries can be directed to the corresponding author.

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

References

- Chen, L.; Liu, Y.; Huang, X.J.; Zhu, Y.Y.; Li, J.X.; Miao, Y.H.; Du, H.Z.; Liu, D.H. Comparison of chemical constituents and pharmacological effects of different varieties of Chrysanthemum Flos in China. *Chem. Biodivers.* **2021**, *18*, e2100206. [\[CrossRef\]](#)
- Hadizadeh, H.; Samiei, L.; Shakeri, A. Chrysanthemum, an ornamental genus with considerable medicinal value: A comprehensive review. *S. Afr. J. Bot.* **2022**, *144*, 23–43. [\[CrossRef\]](#)
- Sun, J.H.; Wang, Z.D.; Lin, C.; Xia, H.; Yang, L.G.; Wang, S.K.; Sun, G.J. The hypolipidemic mechanism of chrysanthemum flavonoids and its main components, luteolin and luteoloside, based on the gene expression profile. *Front. Nutr.* **2022**, *9*, 952588. [\[CrossRef\]](#)
- Wang, Y.; Li, Y.Y.; Guo, W.; Yang, X.; Qu, J.M.; Gao, M.; Chen, S.T.; Dong, J.R.; Li, Q.; Wang, T.J. Comparison of the chemical components, efficacy and mechanisms of action of chrysanthemum morifolium flower and its wild relative chrysanthemum indicum flower against liver-fire hyperactivity syndrome of hypertension via integrative analyses. *Int. J. Mol. Sci.* **2022**, *23*, 13767. [\[CrossRef\]](#)
- Mei, H.; Nie, J.; Wang, S.; Zhang, Y.; Li, C.; Shao, S.; Shao, S.; Rogers, K.M.; Yuan, Y. Geographical origin authentication of edible chrysanthemum morifolium Ramat. (Hangbaiju) using stable isotopes. *Separations* **2023**, *10*, 287. [\[CrossRef\]](#)
- Luo, D.S.; Chen, J.; Gao, L.; Liu, Y.P.; Wu, J.H. Geographical origin identification and quality control of Chinese chrysanthemum flower teas using gas chromatography–mass spectrometry and olfactometry and electronic nose combined with principal component analysis. *Int. J. Food Sci. Technol.* **2017**, *52*, 714–723. [\[CrossRef\]](#)
- Chen, F.R.; Wang, T.; Guo, Q.S.; Zhu, Z.B.; Yang, F.; Zou, Q.J.; Zhang, Y.J. Identification of chrysanthemum indicum in different geographical populations and Ch. morifolium based on DNA barcodes of psbA-trnH, matK and trnL. *China J. Chin. Mater. Med.* **2019**, *44*, 660–665.
- Hao, N.; Gao, X.; Zhao, Q.; Miao, P.Q.; Cheng, J.W.; Li, Z.; Liu, C.Q.; Li, W.L. Rapid origin identification of chrysanthemum morifolium using laser-induced breakdown spectroscopy and chemometrics. *Postharvest Biol. Technol.* **2023**, *197*, 112116. [\[CrossRef\]](#)
- Bec, K.B.; Grabska, J.; Huck, C.W. Principles and applications of miniaturized near-infrared (NIR) spectrometers. *Chem. Eur. J.* **2020**, *27*, 1514–1532. [\[CrossRef\]](#)
- Jiang, H.; Chen, Q.S. Chemometric models for the quantitative descriptive sensory properties of green tea (*Camellia sinensis* L.) using Fourier transform near infrared (FT-NIR) spectroscopy. *Food Anal. Methods* **2015**, *8*, 954–962. [\[CrossRef\]](#)
- Zhang, Y.M.; Guo, W.C. Moisture content detection of maize seed based on visible/near-infrared and near-infrared hyperspectral imaging technology. *Int. J. Food Sci. Technol.* **2019**, *55*, 631–640. [\[CrossRef\]](#)
- Zhang, H.; Jiang, H.; Liu, G.H.; Mei, C.L.; Huang, Y.H. Identification of radix puerariae starch from different geographical origins by FT-NIR spectroscopy. *Int. J. Food Prop.* **2017**, *20*, 1567–1577. [\[CrossRef\]](#)
- Ripoll, G.; Lobón, S.; Joy, M. Use of visible and near infrared reflectance spectra to predict lipid peroxidation of light lamb meat and discriminate dam's feeding systems. *Meat Sci.* **2018**, *143*, 24–29. [\[CrossRef\]](#)
- Biswas, A.; Chaudhari, S.R. Exploring the role of NIR spectroscopy in quantifying and verifying honey authenticity: A review. *Food Chem.* **2024**, *445*, 138712. [\[CrossRef\]](#)
- Lapcharoensuk, R.; Jirawat, P.; Wongpromrat, P. Classification of saline water for irrigated agriculture using near infrared spectroscopy coupled with pattern recognition techniques. *E3S Web Conf.* **2019**, *116*, 00043. [\[CrossRef\]](#)
- Zhan, Y.; Peng, Y.F.; Peng, H.G.; Luo, H.P. Application of near-infrared spectroscopy nondestructive testing of jujube in south xinjiang sugar content. *J. Agric. Mech. Res.* **2014**, *36*, 179–183.
- Guo, Z.; Barimah, A.O.; Shujat, A.; Zhang, Z.; Chen, Q. Simultaneous quantification of active constituents and antioxidant capability of green tea using NIR spectroscopy coupled with swarm intelligence algorithm. *LWT Food Sci. Technol.* **2020**, *129*, 109510. [\[CrossRef\]](#)
- Sun, F.; Chen, Y.; Wang, K.Y.; Wang, S.M.; Liang, S.W. Identification of genuine and adulterated pinellia ternata by mid-infrared (MIR) and near-infrared (NIR) spectroscopy with partial least squares-discriminant analysis (PLS-DA). *Anal. Lett.* **2020**, *53*, 937–959. [\[CrossRef\]](#)
- Wu, J.Z.; Zareef, M.; Chen, Q.S.; Qu Yuan, Q. Application of visible-near infrared spectroscopy in tandem with multivariate analysis for the rapid evaluation of matcha physicochemical indicators. *Food Chem.* **2023**, *421*, 136185. [\[CrossRef\]](#)
- Ouyang, Q.; Rong, Y.N.; Wu, J.Q.; Wang, Z.; Lin, H.; Chen, Q.S. Application of colorimetric sensor array combined with visible near-infrared spectroscopy for the matcha classification. *Food Chem.* **2023**, *420*, 136078. [\[CrossRef\]](#)
- Lin, H.; Kang, W.C.; Han, E.; Chen, Q.S. Quantitative analysis of colony number in mouldy wheat based on near infrared spectroscopy combined with colorimetric sensor. *Food Chem.* **2021**, *354*, 129545. [\[CrossRef\]](#) [\[PubMed\]](#)
- Liu, T.; He, J.; Yao, W.; Jiang, H.; Chen, Q. Determination of aflatoxin B1 value in corn based on Fourier transform near-infrared spectroscopy: Comparison of optimization effect of characteristic wavelengths. *LWT Food Sci. Technol.* **2022**, *164*, 113657. [\[CrossRef\]](#)

23. Larrechi, M.S.; Callao, M.P. Strategy for introducing NIR spectroscopy and multivariate calibration techniques in industry. *TrAC Trends Anal. Chem.* **2003**, *22*, 634–640. [\[CrossRef\]](#)
24. Alaoui Mansouri, M.; Ziemons, E.; Sacré, P.Y.; Kharbach, M.; Barra, I.; Cherrah, Y.; Hubert, P.; Marini, R.D.; Bouklouze, A. Classification of polymorphic forms of fluconazole in pharmaceuticals by FT-IR and FT-NIR spectroscopy. *J. Pharmaceut. Biomed. Anal.* **2021**, *196*, 113922. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Ma, F.M.; Wang, M.Y.; Yan, N.; Adnan, M.; Jiang, F.H.; Hu, Q.; He, G.Y.; Shen, Y.J.; Wan, Y.; Yang, Y.X.; et al. A fast and efficient phenotyping method to estimate sugarcane stalk bending properties using near-infrared spectroscopy. *Eur. J. Agron.* **2024**, *154*, 127107. [\[CrossRef\]](#)
26. Wu, X.H.; He, F.; Wu, B.; Zeng, S.P.; He, C.Y. Accurate classification of chunmee tea grade using NIR spectroscopy and fuzzy maximum uncertainty linear discriminant analysis. *Foods* **2023**, *12*, 541. [\[CrossRef\]](#)
27. Wang, Y.J.; Li, M.H.; Li, L.Q.; Ning, J.M.; Zhang, Z.Z. Green analytical assay for the quality assessment of tea by using pocket-sized NIR spectrometer. *Food Chem.* **2021**, *345*, 128816. [\[CrossRef\]](#)
28. Chen, G.K.; Zhang, X.C.; Wu, Z.B.; Su, J.H.; Cai, G.R. An efficient tea quality classification algorithm based on near infrared spectroscopy and random Forest. *J. Food Process Eng.* **2021**, *44*, e13604. [\[CrossRef\]](#)
29. Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 711–720. [\[CrossRef\]](#)
30. Liu, J.; Chen, S.C.; TAN, X.Y.; Zhang, D.Q. Efficient pseudoinverse linear discriminant analysis and its nonlinear form for face recognition. *Int. J. Pattern Recognit. Artif. Intell.* **2007**, *21*, 1265–1278. [\[CrossRef\]](#)
31. Ye, J.P. Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems. *J. Mach. Learn. Res.* **2005**, *6*, 483–502.
32. Yu, H.; Yang, J. A direct LDA algorithm for high-dimensional data—With application to face recognition. *Pattern Recognit.* **2001**, *34*, 2067–2070. [\[CrossRef\]](#)
33. Thomaz, C.E.; Kitani, E.C.; Gillies, D.F. A maximum uncertainty LDA-based approach for limited sample size problems—With application to face recognition. *J. Braz. Comput. Soc.* **2006**, *12*, 7–18.
34. Kuldip, K.; Sharma, P.A. Improved direct LDA and its application to DNA microarray gene expression data. *Pattern Recognit. Lett.* **2010**, *31*, 2489–2492.
35. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. [\[CrossRef\]](#)
36. Wu, X.H.; Fang, Y.H.; Wu, B.; Liu, M. Application of near-infrared spectroscopy and fuzzy improved null linear discriminant analysis for rapid discrimination of milk brands. *Foods* **2023**, *12*, 3929. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Zhang, T.; Wu, X.H.; Wu, B.; Dai, C.; Fu, H. Rapid authentication of the geographical origin of milk using portable near-infrared spectrometer and fuzzy uncorrelated discriminant transformation. *J. Food Process Eng.* **2022**, *45*, e14040. [\[CrossRef\]](#)
38. Huang, Z.; Ci, H.; Liu, Z.; Xue, Y.; Ren, X.; Xue, J.; Zhang, X. Comprehensive evaluation on yield and quality of medicinal chrysanthemum morifolium varieties based on principal component analysis and cluster analysis. *Sci. Technol. Food Ind.* **2024**, *45*, 271–280. (In Chinese)
39. Rao, G.; Shao, L.; Tong, J. Optimization of ultrasonic-assisted extraction of total flavonoids from Wuyuan imperial chrysanthemum and evaluation of their Antioxidant activity. *Food Res. Dev.* **2021**, *42*, 98–103. (In Chinese)
40. Jin, X.; Wang, L.L.; Zheng, W.J.; Zhang, X.D.; Liu, L.; Li, S.W.; Rao, Y.; Xuan, J.X. Predicting the nutrition deficiency of fresh pear leaves with a miniature near-infrared spectrometer in the laboratory. *Measurement* **2022**, *188*, 110553. [\[CrossRef\]](#)
41. Schafer, R.W. What is a Savitzky-Golay Filter. *IEEE Signal Proc. Mag.* **2011**, *28*, 111–117. [\[CrossRef\]](#)
42. Rima, H.; Arip, M.A.; Endang, L.; Nanda, E.; Taufiq, K.; Rinaldi, I. Combination of PCA with LDA and SVM classifiers: A model for determining the geographical origin of coconut in the coastal plantation, Aceh Province, Indonesia. *Case Stud. Chem. Environ. Eng.* **2024**, *9*, 100552.
43. Long, W.J.; Zhang, Q.; Wang, S.R.; Suo, Y.X.; Chen, H.Y.; Bai, X.Y.; Yang, X.L.; Zhou, Y.P.; Yang, J.; Fu, H.Y. Fast and non-destructive discriminating the geographical origin of Hangbaiju by hyperspectral imaging combined with chemometrics. *Spectrochim. Acta Part A* **2023**, *284*, 121786. [\[CrossRef\]](#) [\[PubMed\]](#)
44. He, C.Y.; Ji, Y.; Wu, B.; Wu, X.H.; Fu, H.J. Non-destructive classification of chrysanthemum tea using near-infrared spectroscopy (NIRS) and fuzzy improved pseudoinverse linear discriminant analysis (FIPLDA). *Anal. Lett.* **2023**, *57*, 905–919. [\[CrossRef\]](#)
45. Kim, N.; Jang, M.; Jo, J.; Park, J.; Kim, A.; Hwang, I. Application of energy dispersive X-ray fluorescence spectrometry and near-infrared reflectance spectroscopy combined with multivariate statistical analysis for discriminating the geographical origin of soybeans. *Food Control* **2022**, *140*, 109140. [\[CrossRef\]](#)
46. Hollingsworth, S.A.; Dror, R.O. Molecular dynamics simulation for all. *Neuron* **2018**, *99*, 1129–1143. [\[CrossRef\]](#)

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.