



Research article

Near-infrared spectroscopy combined with fuzzy fast pseudoinverse linear discriminant analysis to discriminate mee tea grades

Bin Wu^a, Wenbo Tang^b, Jin Zhou^b, Hongwen Jia^{a,*}, Hualei Shen^b, Zuxuan Qi^c^a Department of Information Engineering, Chuzhou Polytechnic, Chuzhou, 239000, China^b Institute of Talented Engineering Students, Jiangsu University, Zhenjiang, 212013, China^c School of Electrical and Information Engineering, Jiangsu University, Zhenjiang, 212013, China

ARTICLE INFO

Keywords:

Near-infrared spectroscopy

Mee tea

Pseudoinverse linear discriminant analysis

KNN

ABSTRACT

Mee tea, one of the major types of green tea in China, is often used for export because of its elegant appearance, high fragrance and strong taste. However, the quality of tea differs greatly due to the difference in raw material selection and production technology level. In order to accurately and quickly differentiate different grades of Mee tea, fuzzy fast pseudoinverse linear discriminant analysis (FFPLDA) was proposed based on fast pseudoinverse linear discriminant analysis (FPLDA) for extracting discriminant information from near-infrared (NIR) spectra. Firstly, NIR spectra of Mee tea samples were acquired, and then they were preprocessed by multiplicative scatter correlation (MSC). Secondly, the compression of data was achieved by principal component analysis (PCA). Thirdly, linear discriminant analysis (LDA), FPLDA, FFPLDA and fuzzy Foley-Sammon transformation (FFST) were respectively performed to retrieve discriminant information from NIR data. Finally, the K-nearest neighbor (KNN) was utilized to classify Mee tea grades. In this study, experimental results showed that the accuracy of FFPLDA was higher than that of LDA, FFST and FPLDA. Therefore, NIR spectroscopy coupled with FFPLDA and KNN has a good effect in discrimination of Mee tea grades and also a great application potential.

1. Introduction

Tea originated in China, with a long history and abundant production. In recent years, the tea industries have developed rapidly. In China, the area planted with tea has expanded from 154.7 km² per year in the 1950s to 3216.7 km² in 2020, and the yield has climbed from 41 kt to 3932 kt [1]. The development of the tea production is also happening in many other nations. Mee tea, also known as Tunxi green tea, is prized for its green leaves' bright color, delicate smell, flavorful richness, and distinctive eyebrow-like form [2,3]. Furthermore, Mee tea also has high nutritional value, such as lowering cholesterol and cardiovascular disease risk [4], uplifting the mood and lowering the levels of stress hormone [5], fighting cancer [6], treating diabetes [7], keeping mouth healthy [8], heat-clearing, detoxifying and weight loss [9]. For these reasons, it is frequently exported to nations including the United States, Morocco, Nigeria, Saudi Arabia, Australia, and Poland, and it is well-established and enjoys great popularity on the global markets

* Corresponding author.

E-mail address: jiahongwen@chzc.edu.cn (H. Jia).

<https://doi.org/10.1016/j.heliyon.2024.e27732>

Received 6 August 2023; Received in revised form 10 December 2023; Accepted 6 March 2024

Available online 9 March 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license

(<http://creativecommons.org/licenses/by-nc/4.0/>).

[10]. There is a higher demand for the high-quality tea products with the surge in tea consumption, the expansion of the tea market, and the emphasis on tea culture. However, the high quality of Mee tea available at the markets was sometimes mixed with the lower quality due to the commercial interests. Therefore, the creation of a Mee tea quality identification system will aid in determining tea quality as well as improving market order, production, and export of tea. At present, common methods of tea quality evaluation include sensory evaluation, spectral analysis, and chemical detection.

Sensory evaluation is a traditional method of evaluating food by means of taste, touch, sight, smell, and hearing. When the quality of tea is evaluated by sensory evaluation, it includes observing the evenness, purity, color, and tenderness of tea leaves; smelling the aroma of the tea to determine whether its aroma is pure or smells like smoke, coke, and mildew; tasting the taste of the tea. However, this method over-relies on human perception and it is susceptible to some influences, including the evaluators' past experiences and environmental circumstances, whose identification results are sometimes inaccurate. With the development of science and technology, researchers willingly adopt some new methods to evaluate the quality of tea. These new methods are more effective than the sensory evaluation method to classify the quality of tea more scientifically and reasonably. For example, Liu et al. examined the excellence and aging life of Pu-erh tea by HPLC fingerprint [11]. Tibetan tea grades could be classified by hyperspectral image and support vector machine [12]. Xu et al. tried an electronic nose and computer vision for the evaluation of the fragrance and visual characteristics of tea leaves to determine the tea grades [13]. Kanrar et al. used ICP-MS with chemometric software to discriminate the original place of tea [14]. In order to achieve the efficient identification of Mee tea quality, the discrimination model should be constructed by a non-destructive, rapid and accurate technology. Near-infrared (NIR) spectral analysis technology has such advantages in the detection of tea and other agricultural products [15–18]. NIR spectroscopy was combined with swarm intelligence methods for predicting the active constituents of green tea leaves [15]. NIR spectroscopy in conjunction with quadratic discriminant analysis shows higher accuracy rate than other single spectroscopic methods in classification of tea varieties [16]. NIR spectroscopy was applied as a non-invasive technology to identify four kinds of green tea, and the classification accuracy of support vector machine (SVM) was higher than partial least squares discriminant analysis [17]. Fourier transform near-infrared spectroscopy (FTNIR) was combined with possibilistic fuzzy discriminant c-means clustering (PFDCM) for classification of four varieties of green tea [18]. It is difficult to accurately analyze some samples with overlapped spectra or numerous interference factors by simply analyzing the peak position and intensity of the atlas, as the fact that the NIR spectra have overlapped broadband absorption due to the absorption peaks in the NIR wavelength range [19]. Nonetheless, the challenges posed by the conventional analytical graph have been significantly resolved thanks to the advancements in theory and the quick rise in computational power. The establishment of relevant models through spectral data can effectively meet the needs of the industry for food quality and authenticity analysis [20]. Anindya et al. discussed a method for speedy, exact and noninvasive classification of tea grades by principal component analysis (PCA) and NIR, showing that this technique might be a suitable substitute for tea grading in property management [21]. Ren et al. assessed the quality levels of black tea via vis-near infrared (Vis-NIR) spectrometer and SVM, and found that Vis-NIR spectroscopy could be a fast, cheap, nondestructive method to predict the black tea quality [22].

The dimension of high-dimensional spectra should be lowered by a feature extraction approach. For example, linear discriminant analysis (LDA) can create a function that connects the high-dimensional and low-dimensional spaces, making it easy to transform the data from the high-dimensional space into the low-dimensional one. Among feature extraction methods, PCA and LDA are popular methods used for dimension reduction [23], which can extract discriminant information to identify the varieties of tea including white tea [24], green tea [25] and black tea [26].

When the dimensionality of samples is much more than the number of samples, LDA may occasionally run into the small sample size (SSS) problem [25,27], which results in the within-class scatter matrix S_w singularity. This limits the extraction of discriminant information by LDA. Pseudoinverse linear discriminant analysis (PLDA) is an effective way to solve the SSS problem because it replaces S_w^{-1} with the pseudoinverse S_w^+ [27]. However, when PLDA computes the high-dimensional data, the data matrix will consume a large amount of resources. To reduce computation amount and improve calculation speed, Liu et al. presented fast pseudoinverse linear discriminant analysis (FPLDA) [27].

Fuzzy recognition is a pattern recognition method that applies fuzzy mathematical methods to address relevant issues. Fuzzy recognition offers the benefits of excellent stability and accurate description of sample data diversity compared with other pattern recognition methods. Nowadays, fuzzy set theory has been utilized in a variety of domains. Ganjeh-Alamdari et al. applied the multi-level fuzzy filter to remove the salt and pepper noise and Fuzzy IF-THEN rules to identify grainy pixels in the damaged images [28]. As the instrument can produce noise which is mixed with NIR spectra during the experiment, fuzzy recognition can reduce the influence of noise during classification. Qi et al. applied fuzzy improved linear discriminant analysis (FiLDA) and a portable NIR spectrometer for classification of red jujube varieties [29].

In this study, fuzzy fast pseudoinverse linear discriminant analysis (FFPLDA) was proposed by combining fuzzy recognition theory with FPLDA. Compared to FPLDA, FFPLDA solves the problem of sample information diversity by introducing fuzzy theory. Moreover, FFPLDA can process NIR spectra containing noise better than FPLDA because FFPLDA is a fuzzy feature extraction method. As a conclusion, FFPLDA can extract fuzzy features that increase the classification accuracy and outperforms FPLDA in classification of NIR spectra with noise.

2. Materials and methods

2.1. Sample preparation

Mee tea samples for the experiment were all acquired from the local supermarkets in Zhenjiang, China. According to the Chinese national standard (GB/T 14456.5–2016), tea samples can be divided into several grades. In this experiment, we selected six grades of Chun Mee tea as samples: special grades one and two; grades one, two, three, and four. According to the standard, the characteristics of these six grades of Mee tea are shown in Table 1 and Table 2. For each sample for detection, about 3.0 g sample was decanted into the beaker which was filled with 150 ml hot water. Then the tea residue was removed by filter paper once the tea soup reached indoor temperature. A pipette gun was applied to drop a little tea liquid in a quartz dish, and then the NIR spectrum of the liquid was acquired by Fourier transform near-infrared (FT-NIR) spectrometer. At a constant humidity, each sample was detected at the approximate 20 °C and 50% relative humidity. There are totally 360 Mee tea samples and sixty for each level grade of Mee tea samples.

2.2. Collection of NIR spectra

An Antaris II FT-NIR spectrometer (Thermo Fisher Scientific Co.) was used to measure each sample of Mee tea. After the spectrometer was preheated for 1 h, it was set in motion. Making choice of the right spectral bands can speed up the processing of spectral information. The wavelength range of the NIR spectrum was 10,000–4000 cm^{-1} . The sampling parameters were set: sampling frequency 4 cm^{-1} , scanning interval 3.857 cm^{-1} , and sampling times 32. The FT-NIR spectra acquired from Mee tea samples were the high-dimensional data with 1557 dimensions. Each sample was measured three times, and the mean value was used as one spectrum for the following experiments. The Matlab 2017b (The Mathworks, Inc.) software was used to program software of data analysis methods.

2.3. Spectral preprocessing method

The composition information of Mee tea can be analyzed by spectroscopic analysis. However, the sample's physical characteristics have a powerful influence on the original NIR spectrum. In addition to the necessary sample properties, spectral data may blend in with noises. Multivariate scattering correction (MSC), as one commonly employed method for preprocessing spectra, helps improve the spectrum contrast [30]. The scattering correction of the correlation spectrum can efficiently lower the influence of scattering, enhance the correlation spectrum's signal-to-noise proportion, and more precisely characterize the straight-line correlation between the optical density and molarity data of the component to be measured in the NIR full spectrum channel [31]. In this study, FT-NIR spectra of Mee tea samples were preprocessed by MSC using the Matlab function.

2.4. Data analysis methods

2.4.1. Principal component analysis

The dimension of FT-NIR spectra is 1557, and the initial FT-NIR spectra of these Mee tea samples contain a few extraneous features and noise data, making categorization more challenging and decreasing classification accuracy. Multiple eigenvectors must be extracted for analysis to retrieve the useful information from FT-NIR spectra. Nevertheless, too many eigenvectors will complicate computations and have an impact on the following spectral classification. Finding eigenvectors that directly reflect spectral differences is the goal of the compression of data. PCA is a commonly used feature extraction method for reducing redundant information [32]. Furthermore, by selecting the proper eigenvectors, PCA keeps the main features of NIR spectra [33].

2.4.2. Linear discriminant analysis

LDA is a classical spectral dimension reduction method [34]. During the dimension-reduction procedure, samples' prior knowledge and experience are utilized [35]. LDA aims to transform spectra data from the high-dimensional space to the low-dimensional space, maximizing inter-class distances and reducing intra-class distances. If the data are linearly separable, when the data are transformed by LDA, they can be classified correctly. However, when the dimensionality of data is high, and the dimension exceeds the number of samples, LDA has a small sample size problem.

Table 1
Shape characteristics of six grades of Mee tea in the experiment.

Grade	Strip	Whole pieces	Colour and lustre	Cleanliness
Grade S1	wiry and tip	evenly	green bloom and silvery	clean
Grade S2	tight	still evenly	green bloom	still clean
Grade 1	tight and heavy	Still evenly	green still bloom	still clean
Grade 2	Still tight	still homogeneous	yellowish green bloom	little immature stem
Grade 3	coarse and bold	still homogeneous	greenish yellow	slender stem
Grade 4	still coarse and loose	broken	yellow	stem

Table 2
Inner qualities of six grades of Mee tea in the experiment.

Grade	Aroma	Flavor	Soup hue	
Grade S1	high aroma	heavy mellow and thick	green and bright	tender and even
Grade S2	rich aroma	heavy and thick	yellowish green and bright	tender and even
Grade 1	still high aroma	heavy and mellow	yellowish green	still tender and even
Grade 2	pure and normal	mellow	greenish yellow	still soft and even
Grade 3	normal	neutral	greenish dull	still soft and greenish yellow
Grade 4	slightly harsh	slightly plain and harsh	slightly yellowish dull	still coarse and greenish yellow

2.4.3. Fast pseudoinverse linear discriminant analysis

Although PLDA can solve the SSS problem of LDA, it requires a great amount of calculation when dealing with a large matrix. To improve PLDA and reduce the calculation amount, FPLDA tries to compute the pseudoinverse S_w^+ by singular value decomposition (SVD) of $H_w^T H_w$. When the data matrix is large, FPLDA can extract the data information more effectively. This paper used FPLDA for feature extraction of Mee tea samples and dimensional-reduction processing of high-dimensional data.

2.4.4. Fuzzy fast pseudoinverse linear discriminant analysis

The steps of fuzzy fast pseudoinverse linear discriminant analysis are as follows.

1. Fuzz data to compute fuzzy membership u_{kj} and the cluster center v_i .
2. Calculate the matrices H_{fw}, H_{fb} .
3. SVD of H_{fw} in terms of $H_{fw} = Q_1 W_1 D_1^T$
4. Construct the matrix $H_{fb2} = Q_2^T H_{fb}$, and Q_2 consists of t ($t = \text{rank}(Q_1)$) columns of vectors of Q_1 .
5. Make the matrix $R = H_{fb2} H_{fb2}^T$, and the matrix $B = D_2^{-2} R$ where D_2 consists of t columns and t rows of vectors of D_1 .
6. Eigen decomposition of B in the form of $B = U_1 V U_1^T$. U_2 consists of r ($r = \text{rank}(U_1)$) columns of vectors of U_1 .
7. Construct the transformation matrix $G = Q_2 U_2$.

In Step 2, H_{fw} and H_{fb} are calculated as follows:

$$H_{fb} = \left[\sqrt{(U_1^f)^T} e (v_1 - \bar{x}), \sqrt{(U_2^f)^T} e (v_2 - \bar{x}), \dots, \sqrt{(U_c^f)^T} e (v_c - \bar{x}) \right] \tag{1}$$

$$H_{fw} = [A_1, A_2, \dots, A_c] \tag{2}$$

Here, U_j^f and A_j are defined as:

$$U_j^f = [u_{1j}^m, u_{2j}^m, \dots, u_{nj}^m]^T \tag{3}$$

$$A_j = \left[\sqrt{u_{1j}^m} (x_1 - v_j), \sqrt{u_{2j}^m} (x_2 - v_j), \dots, \sqrt{u_{nj}^m} (x_n - v_j) \right] \tag{4}$$

In above equations, \bar{x} is the mean of all samples $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} X e$; $e = [1, 1, \dots, 1]^T \in \mathbb{R}^n$; v_i is the i th cluster center; u_{kj} is fuzzy

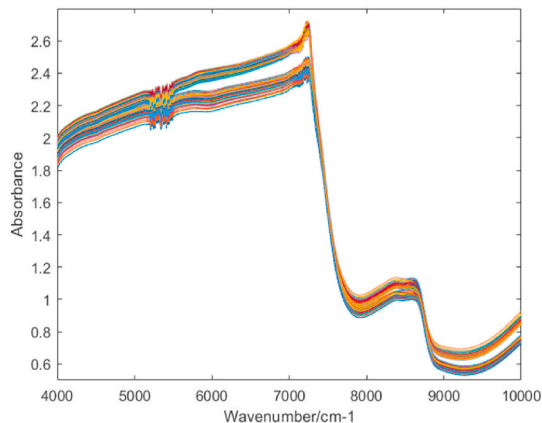


Fig. 1. The original FT-NIR spectra.

membership of the k th sample belonging to the j th class.

2.4.5. K -nearest neighbor

KNN algorithm is one of the commonly used classification algorithms. Its classification principle is to set the number of the nearest neighbors, i.e. the parameter K , to judge the category of prediction points according to most categories of adjacent points [36]. The features of FT-NIR spectra were extracted by PCA + LDA, PCA + FPLDA and PCA + FFPLDA, respectively, and then the KNN algorithm was applied to establish the classification model for Mee tea grades. During the creation of a classification model, both sample size and parameter K may affect the model's classification accuracy.

3. Results

3.1. Spectral data processing

The FT-NIR spectral data of Mee tea samples from six grade levels were used in this experiment, with a wave number range of $4000\text{--}10000\text{ cm}^{-1}$. Commonly, the most prominent absorption bands are associated with molecular overtone and combination vibrations of some hydrogen-based functional groups, such as O–H, C–H, C–O, N–H etc. Fig. 1 displays the initial NIR spectra of Mee tea samples, and Fig. 2 shows the spectra after MSC treatment. Due to the large amounts of organic compounds in tea, which contain many functional groups, The FT-NIR spectra reflected the existence of some functional groups. Fig. 1 shows that the spectral waveform appears burrs at wavenumber around 5300 cm^{-1} and 7100 cm^{-1} . This is because the tea sample contains water, and the absorption bands corresponding to hydroxyl O–H stretching and O–H deformation in water affect the spectral analysis results, and the absorption peaks near 7250 cm^{-1} is formed by the tensile vibration absorption of hydroxyl (O–H) and nitrogen-hydrogen (N–H) bond in tea polysaccharides and hydrogen containing compounds. The last absorption peaks near 8500 cm^{-1} are related to the vibrational stretching of the carbon hydrogen (C–H) bond in tea polyphenols and protein [37,38]. Comparing Figs. 1 and 2, it could be seen that in the initial NIR spectral analysis image, the spectral information was relatively scattered, and had a slight baseline drift, while the spectral lines after MSC preprocessing were more concentrated, which was helpful for the next steps of data analysis. This is because MSC can correct the baseline translation and deviation of FT-NIR spectral data by means of spectral data. The variance reduction within the class is much less than the original spectra, especially at the absorption peak (7247 cm^{-1}). This increases the spectrum's rate of recognition.

3.2. Data processing with PCA

The spectral data after MSC processing are still the high-dimensional data, and they contain the redundant information, which is not conducive to the extraction of NIR features. Therefore, PCA was performed to remove redundant information and conduct spectral dimension reduction. By running PCA, the eigenvalues of PCA were 146.742, 5.398, 3.209, 0.343, 0.142 and 0.052, respectively. PCA removed the superfluous information including the eigenvector associated with the lower eigenvalue and the linearly dependent vectors that can be represented by other vectors during the process of dimension reduction. Meanwhile, it still kept the NIR spectral data's feature information. As the training samples were processed by PCA to generate six principal components (PC1 to PC6), we chose PC1, PC2 and PC3 to make a projection, and the first three principal components (PC1, PC2, and PC3) made up 99.655% of the overall variance. Fig. 3 shows PCA score plots for vectors with PC1, PC2, and PC3. As can be seen from Fig. 3, the clustering position of each type of Mee tea is different, demonstrating that the feature extraction approach can be utilized for the classification and recognition of various grades of Mee tea samples.

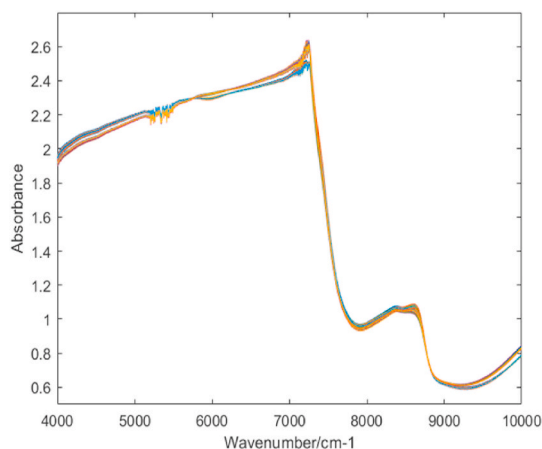


Fig. 2. FT-NIR spectra by MSC.

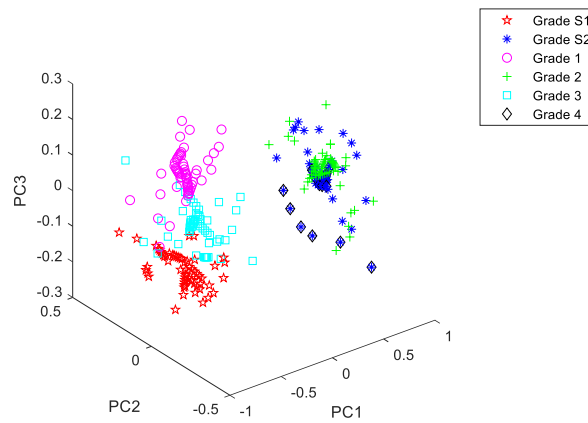


Fig. 3. PCA scores plot of vectors with PC1, PC2 and PC3.

3.3. Data processing with LDA

After PCA processing, the 1557-dimensional spectral data were reduced to 6-dimensional data. The 360 samples of Mee tea were classified into test samples (30% of the samples in each category, 108 in total) and training samples (70% of the samples in each category, 252 in total). Similar to the processing method of PCA, the linear discriminant vectors LDV1, LDV2 and LDV3 were used for data analysis, and the accuracy of the LDA method was illustrated in Fig. 7.

3.4. Data processing with FPLDA

The operation procedure of FPLDA is similar to that of LDA. FPLDA processed the training set to generate three ideal discriminant vectors (DV1, DV2 and DV3), and test samples were made projection onto DV1, DV2 and DV3. The FPLDA scores plot of three vectors is shown in Fig. 4. As shown in Fig. 4, among six grades of samples, there are some overlapped data and each kind of Mee tea is close to each other, which will make the classification difficult.

3.5. Data processing with FFPLDA

After PCA dimension reduction, FFPLDA was used to extract the feature information from the FT-NIR spectral data. Fig. 5 displays the initial fuzzy membership values of training samples. In Fig. 5, the abscissa represents the k th training sample, and the ordinate represents the fuzzy membership value that varies between 0 and 1. The six subgraphs represent the six varieties in the experiment, respectively, representing the fuzzy membership value of the k th training sample. When the fuzzy membership value of the k th training sample is the largest in the interval of the i th class, we can confirm that the k th training sample belongs to the i th class. Compared with the six subgraphs, the abscissa intervals of the six varieties of samples with the highest fuzzy membership value are independent and non-overlapping, which indicates that most of the data in each kind of training samples can be classified clearly by fuzzy membership values. The FFPLDA scores plot of vectors is shown in Fig. 6. From the data distribution of Fig. 6, the correct classification of grades of samples is not easy because data in different grades of samples are distributed in the cross.

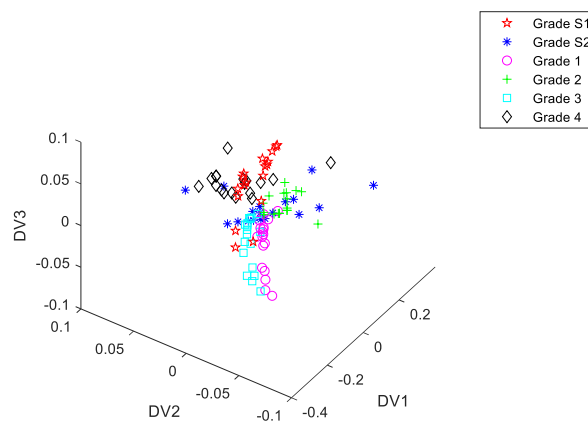


Fig. 4. FPLDA scores plot of three vectors.

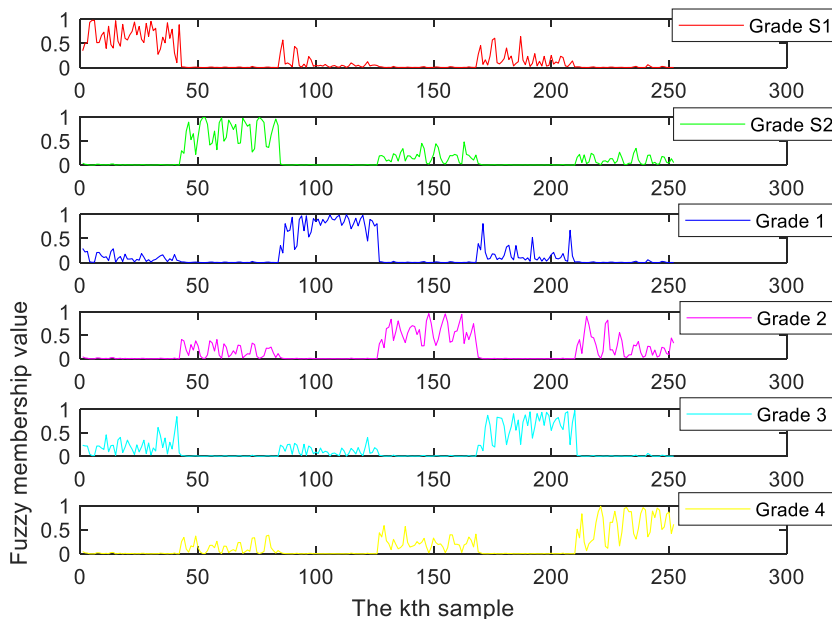


Fig. 5. Initial fuzzy membership values.

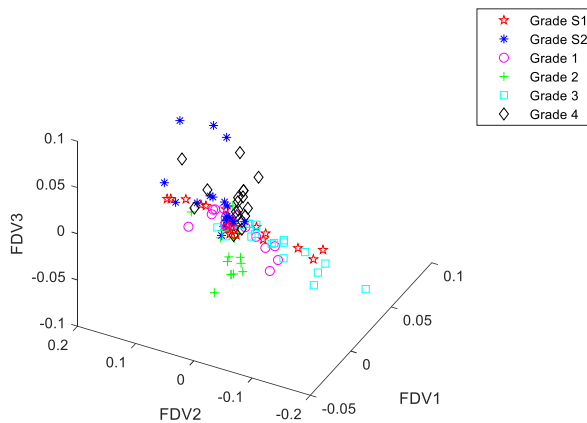


Fig. 6. FFPLDA scores plot of vectors with FDV1, FDV2 and FDV3.

3.6. Classification with KNN

Although six-dimensional test samples were transformed into three dimensional data with LDA, FPLDA and FFPLDA, respectively, three dimensions are not the optimal dimensions for classification of data accurately, and the optimal number of dimensions are the number of variety minus one, i.e. five [35]. In summary, after FT-NIR spectra of the samples were preprocessed by MSC and compressed to six-dimensional data by PCA, they were transformed into five-dimensional data by LDA, FPLDA and FFPLDA, respectively. Then KNN was conducted on the data processed by LDA, FPLDA and FFPLDA, respectively. The K parameter of KNN was set as 1, 3, 5, 7 and 9, respectively, and their corresponding classification accuracies are shown in Fig. 7. When K was 1 and 3, FFPLDA achieved the highest classification accuracy 94.44%.

4. Discussion

In the experiment, the NIR spectral data were the high dimensional data, and it was difficult to process a large number of variables. To solve this problem, dimensionality reduction methods including PCA, LDA, FPLDA, and FFPLDA were used for the extraction of discriminant information from NIR spectra. At first, the Antaris II NIR spectrometer was used to obtain the FT-NIR spectral data of Mee tea samples, and then the spectral data were preprocessed by MSC. Next, we respectively adopted three methods PCA + LDA, PCA + FPLDA and PCA + FFPLDA to reduce the dimension and extract discriminant vectors from data. Finally, data were classified by KNN

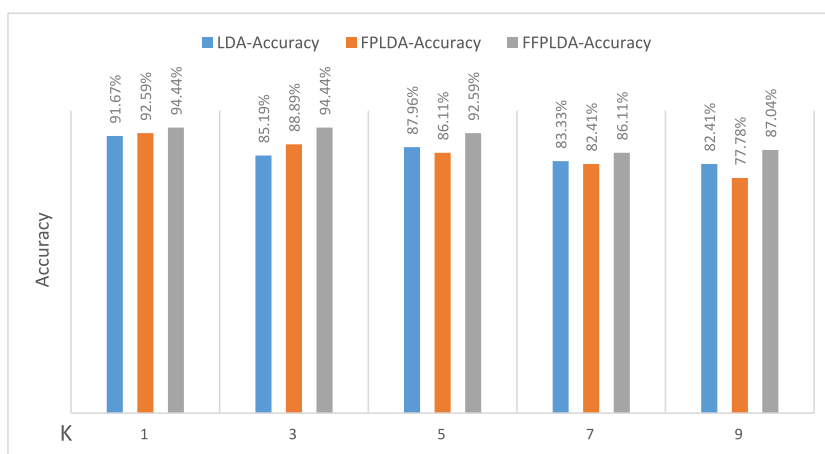


Fig. 7. Classification accuracy of LDA, FPLDA and FFPLDA.

algorithm. Although the five-dimensional data were difficult to be shown in three-dimensional images (such as Figs. 4 and 6), we calculated the classification accuracies of LDA, FPLDA and FFPLDA. As we can see in Fig. 7, all three methods had high classification accuracies, and the classification accuracies varied slightly under different methods and different K values. When K was 1 and 3, FPLDA had the higher accuracies than those of LDA; when K was 5, 7, or 9, FPLDA had the lower accuracies than those of LDA. But the classification accuracies of FFPLDA were obviously higher than those of FPLDA and LDA in all the cases, which reflects that FFPLDA has a significant improvement in classification accuracy by adding fuzzy membership.

In addition, we also discussed the accuracies of the FFPLDA under different m values. The m parameter in FFPLDA, also called the fuzziness parameter, is a fuzzification index. Adjusting the value of m directly affects the fuzziness level and classification results of the FFPLDA method. A larger m value makes the assignment in the membership matrix more ambiguous, resulting in a more uniform membership distribution for each data point across different classes, thus blurring the classification boundary. And a smaller m value makes the allocation in the membership matrix more concentrated, with each data point having a stronger inclination towards a specific class, closer to either 0 or 1 membership values. By trying different m values and comparing the classification accuracy under different settings, the most suitable m value can be selected to achieve the best classification performance. As shown in Fig. 8, when m was equal to 1.2, the accuracy of FFPLDA was the highest, reaching 95.37%. And regardless of how m varied, the accuracy of FFPLDA always remained above 93.52%, maintaining a high level of accuracy. At last, we chose the value of m to be 2 based on traditional experience.

Meanwhile, FFPLDA was compared with fuzzy Foley-Sammon transformation (FFST), which was also a fuzzy feature extraction method for food quality identification [39], and NIR spectral data of Mee tea were processed by FFST and classified with KNN. Then its classification accuracies were calculated under different values of m . The classification results were shown in Fig. 9. It can be seen that despite both using the fuzzy membership, FFPLDA still showed a better performance than FFST in classification accuracy in the case of the same m .

The experimental results provided valuable insights into how dimensionality reduction techniques can be used to improve the classification of spectral data and highlighted the potential benefits of using more advanced techniques such as FFPLDA, which can be predicted to excel in the food industry. Compared with other methods, FFPLDA has similar complexity but higher classification accuracy. Together with the superiority of NIR spectral analysis technology, this makes people have better choice when tea grade classification is required. For example, although it is expensive for individuals to purchase equipment such as spectrometers, when applied to the tea processing plants, the measured products can be classified with low cost and high efficiency through the NIR spectra collected by the spectrometer and processed by the computer. In this paper, there are six grades of Mee tea samples, S1, S2, grades 1, 2, 3 and 4, respectively, which represent the quality from high to low. In practical applications, the products can also be classified by the samples' grades to meet the requirements.

5. Conclusions

As one of the three major drinks in the world, tea has broad market prospects. However, due to the non-standard industrial chain and the tea market being quite a mixed bag, the quality of tea products cannot be guaranteed. And this has led to some fakes and substandard products in the tea product market. In this study, Mee tea samples were used as an example to identify the quality of Mee tea non-destructively and efficiently by FT-NIR spectroscopy. FPLDA was combined with fuzzy set theory and FFPLDA was proposed for feature extraction of FT-NIR data. FT-NIR spectral data of 360 Mee tea samples were collected by Antaris II NIR spectrometer. After the NIR spectra were preprocessed by MCS, the data dimension was reduced by PCA. Then LDA, FPLDA, FFST and FFPLDA were used for feature extraction. In the end, KNN was used to identify the grades of Mee tea. After experiments, we found that compared with other feature extraction algorithms, FFPLDA has top classification accuracy, reaching 94.44%. It was convincing that FFPLDA coupled

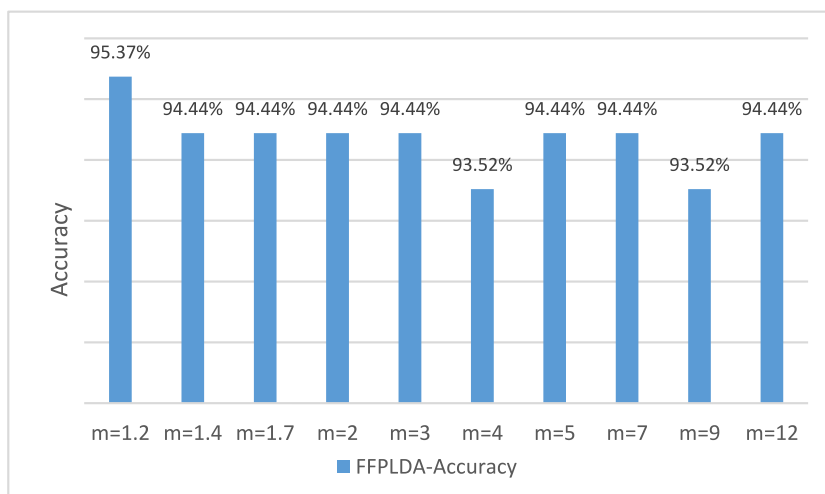


Fig. 8. Accuracy of FFPLDA with different values of m .

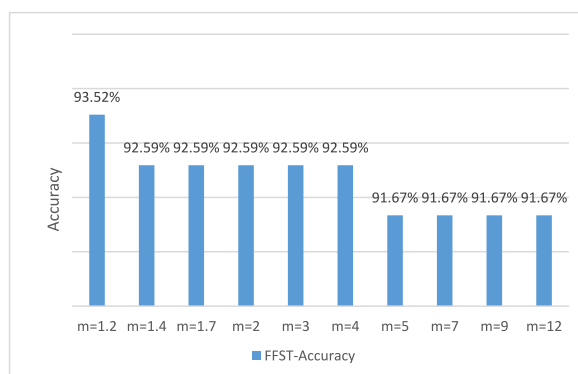


Fig. 9. Accuracy of FFST with different values of m .

with FT-NIR technology will have the great application potential in the grade recognition of Mee tea and other teas. The classification model for Mee tea grades can be applied on a large scale, it will be designed to conform to the standard of the Mee tea industry.

Data availability statement

Not applicable.

Funding

This research was funded by the Major Natural Science Research Projects of Colleges and Universities in Anhui Province (2022AH040333), the Undergraduate Innovation and Entrepreneurship Training Program of Jiangsu Province (202210299258Y), the Talent Program of Chuzhou Polytechnic (YG2019024), and the Key Science Research Project of Chuzhou Polytechnic (YJZ-2020-12).

CRedit authorship contribution statement

Bin Wu: Conceptualization, Funding acquisition, Methodology, Software. **Wenbo Tang:** Funding acquisition, Validation, Writing – original draft. **Jin Zhou:** Data curation, Formal analysis, Investigation. **Hongwen Jia:** Conceptualization, Project administration, Supervision, Validation, Writing – review & editing. **Hualei Shen:** Formal analysis, Resources, Visualization. **Zuxuan Qi:** Methodology, Software.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to

influence the work reported in this paper.

References

- [1] Y. Wang, Y. Yang, X. Zhang, Z. Mao, Y. Yuan, G. Du, Comparative advantage and regional change trend analysis of tea production in China, *J. Tea Commun.* 1 (2023) 1–7.
- [2] Y. Wang, W. Li, J. Ning, R. Hong, H. Wu, Major flavonoid constituents and short-term effects of Chun Mee tea in rats, *J. Food Drug Anal.* 23 (2015) 93–98.
- [3] W. Zhou, Z. Mao, R. Yu, C. Wei, L. Xu, J. Xu, Z. Liu, D. Li, H. Sha, Chemical analysis of exported green tea-mee tea, *J. China Tea Process* 1 (2008) 40–44.
- [4] N. Jochmann, G. Baumann, V. Stangl, Green tea and cardiovascular disease: from molecular targets towards human health, *Curr. Opin. Clin. Nutr. Metab. Care* 11 (2008) 758–765.
- [5] N. Gilbert, The science of tea's mood-altering magic, *Nature* 566 (2019) S8–S9.
- [6] M.F. McCarty, The Japanese experience suggests that lethal prostate cancer is almost wholly preventable with a quasi-vegan diet, soy products, and green tea, *Med. Hypotheses* 164 (2022) 110839.
- [7] O. Asbaghi, F. Fouladvand, M.J. Gonzalez, V. Aghamohammadi, R. Choghakhor, A. Abbasnezhad, The effect of green tea on C-reactive protein and biomarkers of oxidative stress in patients with type 2 diabetes mellitus: a systematic review and meta-analysis, *Complement Ther. Med.* 46 (2019) 210–216.
- [8] M. Mohan, G. Jeevanandan, R.M. S, The role of green tea in oral health - a review, *Asian J. Pharm.* 11 (2018) 1–3.
- [9] P. Mason, T. Bond, Tea and wellness throughout life, *Food Sci. Nutr.* 7 (2021) 1.
- [10] M. Zhou, Problem on the quality of hunan mee tea, *J. Tea Commun.* 3 (1990) 49–50.
- [11] X. Liu, J. Shao, X. Chen, T. Lin, L. Wang, Q. Li, H. Liu, Study on the HPLC fingerprint of different fermentation years Pu-erh tea and identification of aging years, *Chin. J. Anal. Lab.* 34 (2015) 1159–1163.
- [12] Y. Hu, P. Huang, Y. Wang, J. Sun, Y. Wu, Z. Kang, Determination of Tibetan tea quality by hyperspectral imaging technology and multivariate analysis, *J. Food Compos. Anal.* 117 (2023) 105136.
- [13] M. Xu, J. Wang, S. Gu, Rapid identification of tea quality by e-nose and computer vision combining with a synergetic data fusion strategy, *J. Food Eng.* 241 (2019) 10–17.
- [14] B. Kanrar, S. Kundu, P. Khan, V. Jain, Elemental profiling for discrimination of geographical origin of tea (*Camellia sinensis*) in north-east region of India by ICP-MS coupled with chemometric techniques, *Food Chem. Adv.* 1 (2022) 100073.
- [15] Z. Guo, A.O. Barimah, A. Shujat, Z. Zhang, Q. Chen, Simultaneous quantification of active constituents and antioxidant capability of green tea using NIR spectroscopy coupled with swarm intelligence algorithm, *LWT-Food Sci. Technol.* 129 (2020) 109510.
- [16] A. Dankowska, W. Kowalewski, Tea types classification with data fusion of UV-Vis, synchronous fluorescence and NIR spectroscopies and chemometric analysis, *Spectrochim. Acta* 211 (2019) 195–202.
- [17] V.G. Kelis Cardoso, R.J. Ronei, Non-invasive identification of commercial green tea blends using NIR spectroscopy and support vector machine, *Microchem. J.* 164 (2021) 106052.
- [18] B. Wu, H.J. Fu, X.H. Wu, Y. Chen, H.W. Jia, Classification of FTNIR spectra of tea via possibilistic fuzzy discriminant C-means clustering, *Spectrosc. Spect. Anal.* 40 (2020) 512–516.
- [19] Y. Tang, R. Liu, L. Wang, H. Lü, Z. Tang, H. Xiao, S. Guo, W. Fan, Application of one-class classification combined with spectral analysis in food authenticity identification, *Spectrosc. Spect. Anal.* 42 (2022) 3336–3344.
- [20] S. Lohumi, S. Lee, H. Lee, B.K. Cho, A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration, *Trends Food Sci. Technol.* 46 (2015) 85–98.
- [21] R.O. Anindya, J. Muninggar, F.S. Rondonuwu, Indonesian black tea classification using Fourier-Transform near-infrared spectroscopy and a principal component analysis, *J. Phys. Conf. Ser.* 1093 (2018) 1742–6596.
- [22] G. Ren, Y. Liu, J. Ning, Z. Zhang, Assessing black tea quality based on visible–near infrared spectra and kernel-based methods, *J. Food Compos. Anal.* 98 (2021) 103810.
- [23] Y. Huang, Y. Guan, On the linear discriminant analysis for large number of classes, *Eng. Appl. Artif. Intell.* 43 (2015) 15–26.
- [24] L. Zhang, H. Dai, J. Zhang, Z. Zheng, B. Song, J. Chen, G. Lin, L. Chen, W. Sun, Y. Huang, A study on origin traceability of white tea (White Peony) based on near-infrared spectroscopy and machine learning algorithms, *Foods* 12 (2023) 499.
- [25] J. Wang, X. Wu, J. Zheng, B. Wu, Rapid identification of green tea varieties based on FT-NIR spectroscopy and LDA/QR, *Food Sci. Tech.* 42 (2022) e73022.
- [26] M. Palit, B. Tudu, P.K. Dutta, A. Dutta, A. Jana, J.K. Roy, N. Bhattacharyya, R. Bandyopadhyay, A. Chatterjee, Classification of black tea taste and correlation with tea taster's mark using voltammetric electronic tongue, *IEEE Trans. Instrum. Meas.* 59 (2009) 2230–2239.
- [27] J. Liu, S. Chen, X. Tan, D. Zhang, Efficient pseudoinverse linear discriminant analysis and its nonlinear form for face recognition, *Int. J. Pattern Recogn.* 21 (2007) 1265–1278.
- [28] M. Ganjeh-Alamdari, R. Alikhani, I. Perfilieva, Fuzzy logic approach in salt and pepper noise, *Comput. Electr. Eng.* 102 (2022) 108264.
- [29] Z. Qi, X. Wu, Y. Yang, B. Wu, H. Fu, Discrimination of the red jujube varieties using a portable NIR spectrometer and fuzzy improved linear discriminant analysis, *Foods* 11 (2022) 763.
- [30] C. Yang, H. Wu, Y. Yang, L. Su, Y. Yuan, H. Liu, A. Zhang, Z. Song, Identification model of fake and adulterated quinceum antler hat powder by mid-infrared spectroscopy and support vector machine, *Spectrosc. Spect. Anal.* 42 (2022) 2359–2365.
- [31] Y. Lu, Y. Qu, M. Song, Research on the correlation chart of near infrared spectra by using multiple scatter correction technique, *Spectrosc. Spect. Anal.* 27 (2007) 877–880.
- [32] S. Marukatat, Tutorial on PCA and approximate PCA and approximate kernel PCA, *Artif. Intell. Rev.* 56 (2023) 5445–5477.
- [33] S.J. Dixon, R.G. Brereton, Comparison of performance of five common classifiers represented as boundary methods: euclidean distance to centroids, linear discriminant analysis, quadratic discriminant analysis, learning vector quantization and support vector machines, as dependent on data structure, *Chemometr. Intell. Lab. Syst.* 95 (2009) 1–17.
- [34] E. Dogantekin, A. Dogantekin, D. Avci, An automatic diagnosis system based on thyroid gland: ADSTG, *Expert Syst. Appl.* 37 (2010) 6368–6372.
- [35] S.J. Dixon, Application of classification methods when group sizes are unequal by incorporation of prior probabilities to three common approaches: application to simulations and mouse urinary chemosignals, *Chemometr. Intell. Lab. Syst.* 99 (2009) 111–120.
- [36] C. Maione, R.M. Barbosa, Recent applications of multivariate data analysis methods in the authentication of rice and the most analyzed parameters: a review, *Crit. Rev. Food Sci. Nutr.* 59 (2019) 1868.
- [37] F. He, X. Wu, B. Wu, S. Zeng, X. Zhu, Green tea grades identification via Fourier transform near-infrared spectroscopy and weighted global fuzzy uncorrelated discriminant transform, *J. Food Process. Eng.* 45 (2022) e14109.
- [38] Q. Chen, J. Zhao, C. Fang, D. Wang, Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine(SVM), *Spectrochim. Acta* 66 (2007) 568–574.
- [39] Y. Shen, X. Wu, B. Wu, Y. Tan, J. Liu, Qualitative analysis of lambda-cyhalothrin on Chinese cabbage using mid-infrared spectroscopy combined with fuzzy feature extraction algorithms, *Agriculture* 11 (2021) 275.