

Research Article

Qualitative Discrimination of Intact Tobacco Leaves Based on Near-Infrared Technology

Mengyao Lu ¹, Qiang Zhou,² Tian'en Chen ^{1,3}, Junhui Li,⁴ Shuwen Jiang,^{1,3} Qin Gao,² Cong Wang,¹ and Dong Chen^{1,3}

¹National Engineering Research Center for Information Technology in Agriculture, Beijing 100019, China

²Anhui Wannan Tobacco Co., Ltd., Xuancheng 242000, China

³NONGXIN(Nanjing) Smart Agriculture Research Institute, Nanjing 211808, China

⁴China Agricultural University, Beijing 100083, China

Correspondence should be addressed to Tian'en Chen; chente@nercita.org.cn

Received 6 September 2021; Revised 1 November 2021; Accepted 2 November 2021; Published 2 December 2021

Academic Editor: Arnaud Cuisset

Copyright © 2021 Mengyao Lu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To explore the application of near-infrared (NIR) technology to the quality analysis of raw intact tobacco leaves, a nondestructive discrimination method based on NIR spectroscopy is proposed. A “multiregion + multipoint” NIR spectrum acquisition method is developed, allowing 18 NIR diffuse reflectance spectra to be collected from an intact tobacco leaf. The spectral characteristics and spectral preprocessing methods of intact tobacco leaves are analyzed, and then different spectra (independent or average spectra) and different algorithms (discriminant partial least-squares (DPLS) and Fisher's discriminant algorithms) are used to construct discriminant models for verifying the feasibility of intact leaf modeling and determining the optimal model conditions. Qualitative discrimination models based on the position, green-variegated (GV), and the grade of intact tobacco leaves are then constructed using the NIR spectra. In the application and verification stage, a multiclassification voting mechanism is used to fuse the results of multiple spectra from a single tobacco leaf to obtain the final discrimination result for that leaf. The results show that the position-GV discrimination model constructed using independent spectra and the DPLS algorithm and the grade discrimination model constructed using independent spectra and Fisher's algorithm achieve optimal results with intact leaf NIR wavenumbers from 5006–8988 cm^{-1} and the first-derivative and standard normal variate transformation preprocessing method. Finally, when applied to new tobacco leaves, the position-GV model and the grade model achieve discrimination accuracies of 95.18% and 92.77%, respectively. This demonstrates that the two models have satisfactory qualitative discrimination ability for intact tobacco leaves. This study has established a feasible method for the nondestructive qualitative discrimination of the position, GV, and grade of intact tobacco leaves based on NIR technology.

1. Introduction

Enhanced requirements in the Chinese cigarette industry regarding the quality of flue-cured tobacco raw materials mean that improving and maintaining the quality of tobacco leaves in the production area is increasingly important in the tobacco processing industry [1, 2]. In the management of tobacco quality, the position and grade of tobacco are the most important key factors [3]. According to the different growing positions of the tobacco on the tobacco stalk, tobacco positions are divided into three major positions (i.e., upper(B), middle(C), and lower(X)). The chemical

components of tobacco leaves in different positions are obviously different, showing different aromas, tastes, and irritations [3, 4]. In accordance with the Chinese national standard for “Flue-cured tobacco” (GB2635-1992) [5], based on the differences of tobacco characteristics, such as tissue structure, oil content, thickness, maturity, and injury degree, tobacco leaves in the same position can be further divided into 1, 2, 3, and 4 grades. Therefore, the final grade of tobacco leaves is a combination of tobacco position, grade in each position, and tobacco color (i.e., lemon-yellow(L), orange(F), and reddish-brown(R)). For example, B2F refers to the tobacco leaf of upper, 2 grades, and orange. Different

grades of tobacco with different physical and chemical properties play different roles in the formulation of Chinese cigarettes [2, 6]. However, there are also some bad tobacco leaves with no industrial availability (i.e., green tobacco(G) and variegated tobacco(V)) that need to be recognized and discarded. In general, the accurate discrimination of the position, green-variegated (GV), and the grade of raw tobacco leaves in the production area is an important means of ensuring the quality of raw tobacco leaves [7]. At present, there are two main methods for discriminating the position, GV, and grade of intact tobacco leaves: manual discrimination and image recognition technology. The former method mainly relies on experienced grading workers to discriminate tobacco leaves through sensory feelings such as visual observation and touch based on GB2635-1992. This traditional manual operation method is easily affected by the subjective-cognitive differences of grading workers, resulting in poor accuracy and consistency of the grading results [8]. Image recognition involves collecting the images of tobacco leaves in the visible light band and then extracting the image features and discriminating them through some machine learning algorithm. Yao et al. proposed an image feature classification method for tobacco leaves based on principal component analysis (PCA), a genetic algorithm (GA), and a support vector machine (SVM). They reduced the dimensionality of 15 extracted features (e.g., color, leaf state, and texture) and realized tobacco leaf classification using a hybrid GA-SVM [9]. Dasari et al. used a convolution neural network to extract the RGB image features of tobacco leaves and classify the leaves into three grades [10]. Tobacco leaf classification based on image recognition technology can achieve good recognition results. However, image-based methods depend on the appearance of the tobacco leaves, such as the leaf shape and color, and ignore important factors that affect the grade of tobacco leaves, such as identity, maturity, oil content, and other features that cannot be extracted from the images. Therefore, it is essential to develop an accurate, reasonable, and rapid discrimination method for the position, GV, and grade of intact tobacco leaves.

Near-infrared (NIR) spectroscopy is used in tobacco quality evaluation [11], routine tobacco chemical composition analysis [12], and Chinese cigarette formula design [13]. NIR light covers electromagnetic waves between visible light and midinfrared light, with a wavelength range of 750–2500 nm. NIR technology has the advantages of being environmentally sustainable, fast, and nondestructive [14, 15] and is widely used in many aspects of the tobacco industry. The traditional NIR spectrum detection of tobacco leaves mainly uses a Fourier-type [16] or grating-type NIR spectrometer [17]. During the spectrum detection process, the tobacco leaves need to be grounded into powder [18] or cut into shreds [19]. When the sample is placed in the test sample cup, shaking and compaction operations need to be carried out to ensure the uniformity of the sample and the accuracy of NIR spectrum detection. The absorption signals of C-H, N-H, and O-H in the NIR spectra are able to show the contents of carbohydrates and the nitrogenous substances of tobacco leaves [20–22]. The chemical information

contained in the NIR spectra of tobacco has a certain correlation with the grades/positions of a tobacco leaf [3, 6]. This correlation is the basis of the tobacco grading/classification model using the NIR spectra. However, there is still no national standard to show the definite value of chemical contents corresponding to tobacco's grade/position. Currently, the methods and standards of tobacco grade discrimination are qualitative rather than quantitative in tobacco industrial production [5]. Also, there have been some applications of NIR technology in the grade and position qualitative discriminations of tobacco. Wang et al. proposed a principal component accumulation method to solve the tobacco position classification based on the NIR spectroscopy [23]. Luan et al. applied NIR and multiple classifier (SVM, PLS, and PPF) fusion models to classify the position of the tobacco leaves [24]. Bin et al. proposed a modified random forest approach to improve the multiclass classification performance of tobacco leaf grades coupled with the NIR spectroscopy [25]. Previous studies have achieved good results in the qualitative discrimination of a tobacco leaf's position and grade, but their research materials are tobacco powder rather than the intact tobacco leaf. This detection method destroys the integrity of the tobacco leaves and has a long sample preparation time and thus obviously cannot meet the quality analysis requirements of raw intact tobacco leaves.

In terms of the NIR detection of intact tobacco leaves, Dong et al. [26], Bin et al. [27], and He et al. [28] collected the NIR spectra of intact tobacco leaves using 10-point, 8-point, and 3-point methods, respectively, for the quantitative analysis of the chemical components of tobacco leaves. Ying et al. discussed the application of the NIR spectroscopy for nondestructive detection of the position and color of intact flue-cured tobacco [29]. The above research provides a certain theoretical basis and technical route for the collection and analysis of the NIR spectra of intact tobacco leaves. However, previous research has not analyzed the preprocessing methods required in the NIR spectroscopy of intact tobacco leaves. Additionally, there have been no reports on qualitative discrimination methods for GV and grade, two important evaluation criteria of tobacco leaf quality in the Chinese cigarette industry.

To realize the discrimination of the position, GV, and grade of intact flue-cured tobacco based on NIR technology, this study took intact tobacco leaves from Southern Anhui as the research object, explored NIR spectrum acquisition and the spectrum preprocessing methods, and constructed qualitative discrimination models of the tobacco position, GV, and grade. The resulting models can realize the non-destructive discrimination of flue-cured tobacco.

2. Materials and Methods

2.1. Materials and Instruments. We obtained 349 flue-cured tobacco leaves of the Yunyan-87 variety grown in Southern Anhui in 2018, including three tobacco positions (i.e., upper(B), middle(C), and lower(X)), green tobacco, and variegated tobacco, with a total of ten grades (i.e., B2F (upper, 2 grade, and orange), B3F (upper, 3 grade, and

orange), C2F (middle, 2 grade, and orange), C3F (middle, 3 grade, and orange), C4F (middle, 4 grade, and orange), X2F (lower, 2 grade, and orange), X3F (lower, 3 grade, and orange), GY2 (green tobacco), B2K (upper and variegated tobacco), and CX1K (middle-lower and variegated tobacco)). The qualitative discrimination information (position, GV, and grade) of these tobacco leaves was determined by two professional grading staff with more than 5 years of experience based on GB2635-1992. Table 1 summarizes some statistics pertaining to the tobacco leaves.

An Armor711 NIR spectrometer (Carl Zeiss Co., Ltd., Germany) was used to obtain the NIR spectra. This spectrometer uses optical fiber diffuse reflection with an InGaAs detector. The light source is a halogen lamp with a voltage of 12 V and a power of 50 W. The diameter of the field of view is 30 mm. The spectral range is 910–2200 nm with a spectral resolution of 8 nm. The wavelength accuracy is less than 0.5 nm, and the wavelength repeatability is less than 0.05 nm [30, 31].

2.2. NIR Spectra Acquisition. An intact tobacco leaf is wide and long, and so the chemical composition, oil content, color, and surface flatness may vary across its area [32, 33]. The internal factors of tobacco and the external measurement conditions both affect the accuracy of the NIR spectra acquisition. Therefore, a “multiregion + multipoint” NIR spectra acquisition method is proposed. First, the multiple regions are determined, that is, the tobacco leaf is equally divided into four regions (leaf tip, upper-middle leaf, lower-middle leaf, and leaf base) from leaf tip to leaf base along the main vein direction. As shown in Figure 1, A1 is the leaf tip region; the upper-middle and lower-middle leaf regions are divided along the main vein to give regions A2, A3, A4, and A5; finally, A6 is the leaf base region. The NIR spectrometer used in this study has a light spot with an effective area of 30 mm diameter, which can expand the spectra scanning range to a certain extent. However, to acquire the representative NIR spectra and improve the quality of regional spectral data, a “multipoint” spectra acquisition method is proposed. In each acquisition area, three acquisition points are randomly selected within 2 cm of the geometric center of the area (if there is a leaf damage at the desired point, the nearest available point is used).

The NIR spectra of the intact tobacco leaves were collected at $25 \pm 1^\circ\text{C}$ (room temperature) and 80% relative humidity. 32 scans were collected for each spectrum with 256 data points. During spectra acquisition, the optical fiber detector head was positioned at 90° (perpendicular) to the tobacco sample, and the distance between the lower end of the detector head and the tobacco leaf surface was maintained at 100 mm [30, 31]. A total of 18 spectra were acquired in the six regions of a single tobacco leaf.

2.3. Spectra Preprocessing. Instrument noise, sample conditions, environmental factors, and personnel operation may cause some errors in the acquired NIR spectra of tobacco leaves. To improve the accuracy of the spectral data and the subsequent modeling accuracy, the original NIR spectra of

the intact tobacco leaves were preprocessed by using first-derivative (1st derivative) and standard normal variate transformation (SNV) to remove systematic noise and random errors [34, 35]. The accuracy of the spectral modeling using the 1st derivative, SNV, and 1st derivative + SNV preprocessing is compared in this study.

2.4. Qualitative Discrimination Method. PCA [36], discriminant partial least squares (DPLS) [37], and Fisher's discrimination [38], are widely used for the qualitative discrimination of the NIR spectra. PCA is used to reduce the dimension of the NIR spectra, which is beneficial for subsequent data analysis. In DPLS, the output value from regular PLS is used as a classification label for the regression of the sample about the class. In Fisher's discrimination, projection technology is used to reduce the dimension of the NIR spectra from multiple samples so that the different samples have the largest interclass distance and the smallest intraclass distance in the new projection space. This enables the correct classification of samples.

When constructing the discrimination model for the intact tobacco NIR spectra, 349 tobacco samples were randomly divided into a training set and a validation set at a ratio of 4 : 1. The method of randomly dividing is as follows:

- (1) The randperm function in MATLAB is used to generate a row vector containing 298 (i.e., rounding 349×0.8 to 298) unique integers selected randomly from 1 to 349.
- (2) According to the integers in the vector, the corresponding numbered samples from all samples are taken out in order. These samples form the training set.
- (3) The remaining samples form the validation set.

The tobacco spectra in the different sets were then used for modeling. In addition, a total of 18 independent spectra were collected at different acquisition points from each intact tobacco leaf, allowing the results of models constructed using two types of spectra to be compared: the independent spectra at different acquisition points (hereafter called the independent spectra) and the average spectra. The average spectra are the average values of 18 spectra from an intact tobacco leaf.

2.5. Model Evaluation Indexes. In this study, we use a discriminant accuracy to evaluate the model performance. The discriminant accuracy is defined as follows:

$$\text{accuracy} = \frac{\text{No. samples accurately discriminated by the model}}{\text{No. all samples}} \times 100\%. \quad (1)$$

3. Results and Discussion

3.1. Characteristics of the NIR Spectra. As can be seen from Figures 2(a)–2(d), the outline of the spectrum is smooth and clear, and the positions of the absorption peak and absorption valley are obvious. The absorption peaks are mainly

TABLE 1: Summary statistics of tobacco leaves.

Positions-GV	Upper (B)		Middle (C)			Lower (X)		Green (G)	Variegated (V)	
Grade	B2F	B3F	C2F	C3F	C4F	X2F	X3F	GY2	B2K	CX1K
Number	23	62	20	46	33	22	37	43	35	28
	85		99			59		43	63	

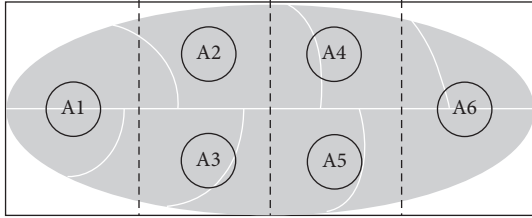


FIGURE 1: Spectra acquisition area of intact tobacco leaf.

near the wavenumbers of 6846 cm^{-1} and 5173 cm^{-1} , and the absorption valleys are near the wavenumbers of 6050 cm^{-1} and 5407 cm^{-1} , which is consistent with the characteristics and positions of the peaks and valleys of the NIR spectra collected by tobacco powders and prepared by traditional methods [24, 39]. These results show that the spectra collected from intact tobacco leaves using the multipoint method contain better information about the different functional groups in tobacco. Figure 2 also shows that there is some background interference and baseline shift in the spectra. This may be because the surface of the intact tobacco leaves is uneven, resulting in different distances between the detector head and the tobacco. Moreover, the random fluctuations that appear on the NIR spectra are noise, which can originate from the instrument or environmental laboratory conditions [40, 41].

3.2. Optimal Conditions of the Qualitative Discrimination Model. The original NIR spectra display strong noise in the wavenumber range from $4528\text{--}4995\text{ cm}^{-1}$ and contain less information in the wavenumber range from $9036\text{--}11057\text{ cm}^{-1}$. Therefore, these two bands should be deleted during modeling. The spectra from $5006\text{--}8988\text{ cm}^{-1}$ were selected for modeling. PCA was applied to reduce the dimension of the spectra. The cumulative contribution rate of the nine principal components of the independent spectra is 99.15%, and the six principal components of the average spectra contribute 99.66% of the total. Therefore, the nine principal components from the independent spectra and six principal components from the average spectra were used in the modeling.

3.2.1. Selection of Preprocessing Methods. The 1st derivative, SNV, and 1st derivative + SNV methods were used to preprocess the independent spectra and the average spectra. The position-GV and grade discrimination models were constructed using the DPLS algorithm and preprocessed spectra. The modeling results are presented in Tables 2 and 3.

As can be seen from Table 2, when the independent spectra were preprocessed with either the 1st derivative or the SNV method, the validation set achieved higher accuracy

with both the position-GV discrimination model and the grade discrimination model than when no preprocessing was applied. When the independent spectra were preprocessed by 1st derivative + SNV, the two discriminant models reached accuracy levels of 96.22% and 86.25%, respectively, with the validation set, representing a further improvement over the use of a single preprocessing method (1st derivative or SNV). Table 3 shows that the results from the models constructed using the average spectra are similar, that is, the 1st derivative + SNV preprocessing method is better than either the single preprocessing method, and preprocessing is always better than no preprocessing. The discrimination models constructed using the average spectra reached the accuracy levels of 92.71% and 81.75%, respectively, with the validation set. Analyzing the above results, it appears that the 1st derivative preprocessing method eliminates the background interference and baseline shift in the spectrum, while the SNV preprocessing method reduces the scattering interference in the spectra acquisition process from the uneven surface of the tobacco leaves, which effectively improves the data quality and the modeling results. Thus, the combined preprocessing method of 1st derivative + SNV is suitable for both independent spectra and average spectra of the intact tobacco leaves.

3.2.2. Modeling Results Using Fisher's Algorithm. Fisher's algorithm combined with the independent spectra and average spectra preprocessed with 1st derivative + SNV was used to construct tobacco discrimination models. Figures 3(a)–3(c) and 3(d)–3(f) show the Fisher's two-dimensional projection of position-GV discrimination modeling using the independent and average spectra, respectively. The training set and validation set accuracies of the position-GV discrimination model constructed by the independent spectra are 97.67% and 94.27%, respectively, whereas those of the model constructed using the average spectra are 100.00% and 91.43%, respectively. As can be seen from Figures 3(a)–3(c), there is no cross between the upper tobacco (B) and lower tobacco (X), whereas there are crosses between middle tobacco (C) and upper tobacco (B), lower tobacco (X) and variegated tobacco (V), which is in accord with the basic order of the tobacco's natural growth. The samples of different classes in the validation set are mostly within the confidence ellipse of the training set, while the misjudged samples are mainly concentrated around the intersection of different classes, and the overall discrimination results are good. It can be seen from Figures 3(d)–3(f) that the samples from different classes in the validation set exhibit a cluster distribution. In general, the discrimination results for the model using the average spectra are slightly worse than those for the model using the independent spectra.

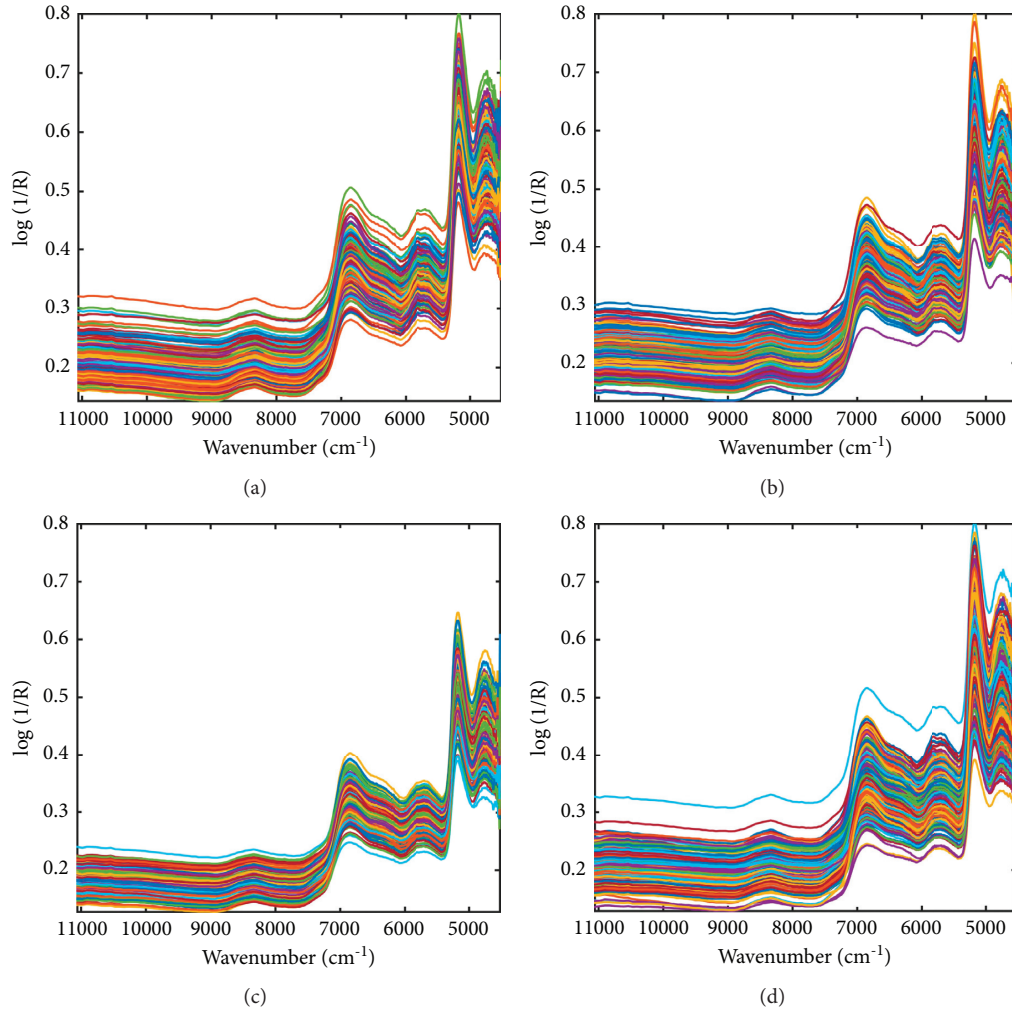


FIGURE 2: NIR spectra of intact tobacco leaves. (a) B, upper. (b) C, middle. (c) X, lower. (d) GV, green-variegated.

TABLE 2: Accuracy of the DPLS model using different preprocessing methods (independent spectra).

Preprocessing method	Position-GV discrimination model		Grade discrimination model	
	Training set	Validation set	Training set	Validation set
None	92.98	90.75	85.76	80.37
1st derivative	98.31	94.22	91.27	84.46
SNV	100.00	94.22	89.39	82.33
1st derivative + SNV	100.00	96.22	90.51	86.25

TABLE 3: Accuracy of the DPLS model using different preprocessing methods (average spectra).

Preprocessing method	Position-GV discrimination model		Grade discrimination model	
	Training set	Validation set	Training set	Validation set
None	96.24	86.81	83.88	77.96
1st derivative	98.49	90.69	86.10	80.02
SNV	92.08	90.41	85.30	80.02
1st derivative + SNV	98.12	92.71	86.84	81.75

Figures 4(a)–4(c) and 4(d)–4(f) show the Fisher's two-dimensional projection of the grade discrimination modeling using the independent spectra and average spectra, respectively. The training set and validation set accuracies of the grade discrimination model constructed by the independent spectra

are 99.64% and 97.49%, respectively, whereas those for the model constructed using the average spectra are 100.00% and 91.43%, respectively. As can be seen from Figures 4(a)–4(c), the upper tobacco (i.e., B2F, B3F) is located on the far left of the overall distribution, intersecting with the middle tobacco (i.e.,

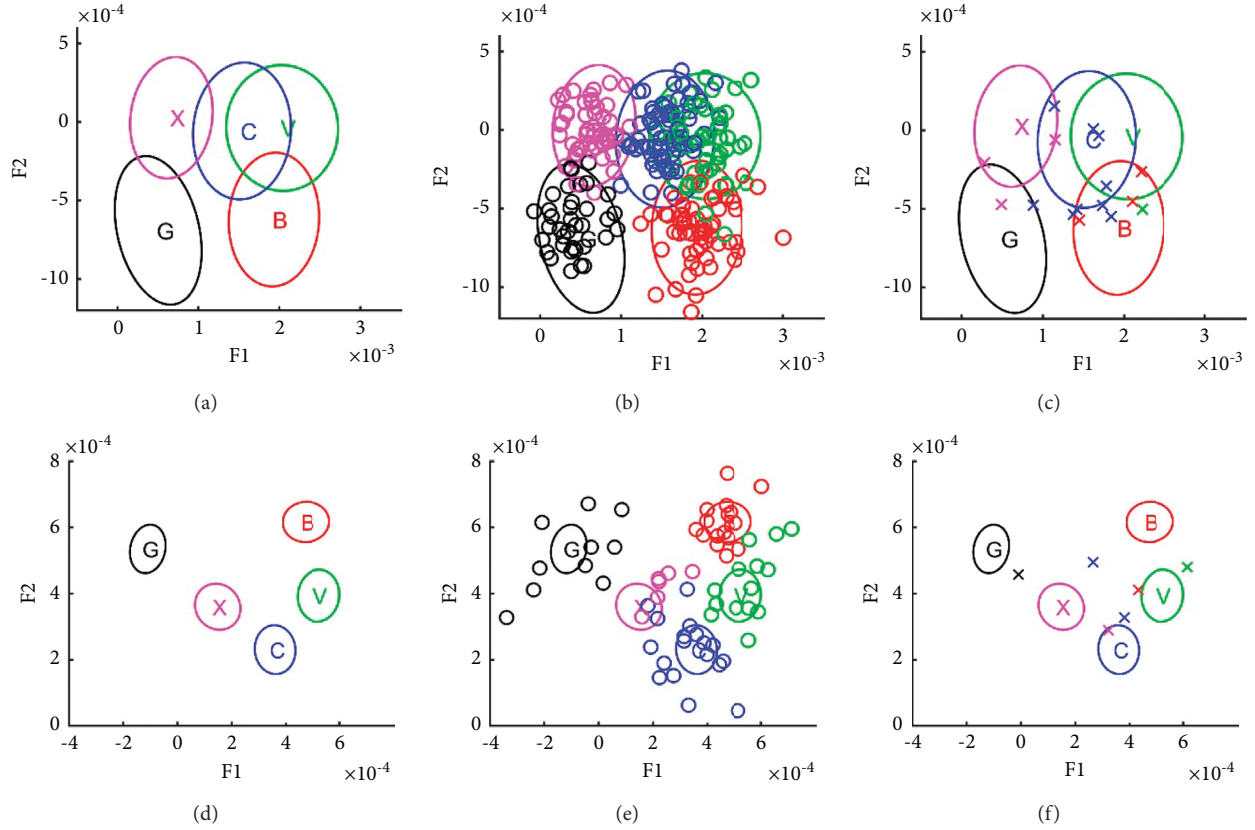


FIGURE 3: Two-dimensional projection distribution diagrams of the training set (a, d), sample scatter diagrams of the correct discrimination of the validation set (b, e), and sample scatter diagrams of the wrong discrimination of the validation set (c, f) for the independent spectra and average spectra modeling (position-GV discrimination).

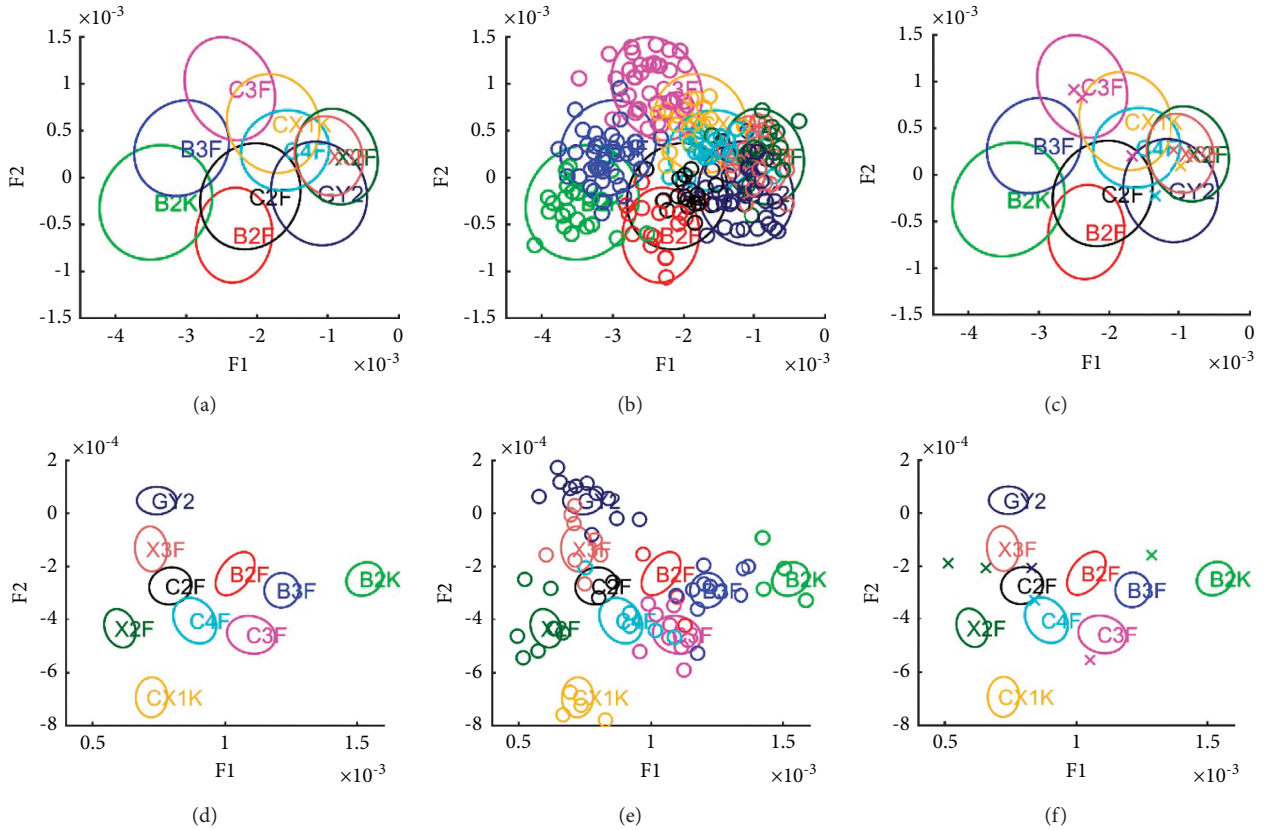


FIGURE 4: Two-dimensional projection distribution diagrams of the training set (a, d), sample scatter diagrams of the correct discrimination of the validation set (b, e), and sample scatter diagrams of the wrong discrimination of the validation set (c, f) for independent spectra and average spectra modeling (grade discrimination).

TABLE 4: Results of the qualitative discrimination model of intact tobacco leaves (spectra wavenumbers from 5006–8988 cm^{-1} ; preprocessing method: 1st derivative + SNV).

Modeling algorithm	Modeling spectra	Position-GV discrimination		Grade discrimination	
		Training set (%)	Validation set (%)	Training set (%)	Validation set (%)
DPLS	Independent spectra	100.00	96.22	90.51	86.25
	Average spectra	98.12	92.17	86.84	81.75
Fisher	Independent spectra	97.67	94.27	99.64	97.49
	Average spectra	100.00	91.43	100.00	91.43

TABLE 5: Confusion matrix of position-GV discrimination.

True class	Predict class				
	Upper(B)(%)	Middle(C)(%)	Lower(X)(%)	Green(G)(%)	Variegated(V)(%)
Upper(B)	100.00	0.00	0.00	0.00	0.00
Middle(C)	0.00	92.59	7.41	0.00	0.00
Lower(X)	0.00	5.88	94.12	0.00	0.00
Green(G)	0.00	7.69	0.00	92.31	0.00
Variegated(V)	0.00	0.00	0.00	0.00	100.00

C2F, C3F), and the lower tobacco (i.e., X2F, X3F) is located on the far right of the overall distribution. The overall distribution shows the clustering of tobacco positions. As shown in Figures 4(d)–4(f), the sample scatter of the validation set is close to the confidence ellipse of the training set, and the samples of the same class are clustered together.

3.2.3. Comparison of Modeling Conditions. Table 4 presents the modeling results under different conditions. The following conclusions can be drawn:

- (1) The discrimination accuracy on the validation set is better when using the independent spectra than when using the average spectra. Because of the uneven distribution of physical and chemical properties in different regions of intact tobacco leaves, the average spectra eliminate errors in different regions. However, the average spectra also offset the contribution of the differences between the independent spectra in different regions to the qualitative discrimination results. Moreover, the discrimination results of the 18 independent spectra of a single tobacco leaf may be correct or incorrect, but the average spectra reduce the fault tolerance of a single tobacco leaf, which reduces the overall discrimination accuracy.
- (2) In terms of position-GV discrimination, the validation set accuracy is similar when using the Fisher's algorithm and DPLS. However, for grade discrimination, the validation set accuracy modeled by the Fisher's algorithm is significantly higher than when using DPLS. The experimental results show that the DPLS algorithm is more suitable for discriminant tasks with a small number of classification classes, such as the discrimination of tobacco position-GV, which has a total of five classes. When the number of classification classes is large and the amount of data in each class is unbalanced, the DPLS algorithm gives poor discrimination results, such as in the

discrimination of the tobacco grade, which has a total of 10 classes. In this case, the Fisher's algorithm is recommended.

- (3) The optimal model condition for the position-GV discrimination of intact tobacco leaves is DPLS with the independent spectra. The optimal model condition for the grade discrimination is the Fisher's algorithm with the independent spectra.

3.3. Application and Verification of the Model. The above modeling methods are based on the NIR spectra. We now describe the qualitative discrimination of intact tobacco leaves in a practical application. For this, another 83 Southern Anhui tobacco leaves from the same year were collected, and 18 NIR spectra were obtained from each tobacco leaf for model verification. During verification, 5 (or 10) spectra were randomly selected from the 18 spectra of each tobacco leaf and input into the position-GV discrimination model (or grade discrimination model). The multiclassification voting mechanism was then used to fuse the discrimination results of these 5 (or 10) spectra, and the class with the most occurrences in the statistical result gave the final discrimination class.

The overall discrimination accuracy of the position-GV model is 95.18% and that of the grade discrimination model is 92.77%. The detailed discrimination results are presented in Tables 5 and 6. Table 5 indicates that the overall discrimination result of the position-GV model is good, and there is no misjudgment between the upper and lower tobacco leaves. From Table 6, it is clear that the misjudgments made by the grade discrimination model are mainly concentrated in the adjacent grades or in the same position, and the accuracy of adjacent grade discrimination reaches 98.33%. The results of this application and verification show that the intact tobacco leaf qualitative discrimination method based on NIR technology proposed in this paper can achieve good discrimination ability for the position-GV and grade of Southern Anhui tobacco leaves.

TABLE 6: Confusion matrix of grade discrimination.

True class	Predict class									
	B2F (%)	B3F (%)	C2F (%)	C3F (%)	C4F (%)	X2F (%)	X3F (%)	B2K (%)	CX1K (%)	GY2 (%)
B2F	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B3F	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C2F	0.00	0.00	90.91	9.09	0.00	0.00	0.00	0.00	0.00	0.00
C3F	0.00	0.00	8.33	91.67	0.00	0.00	0.00	0.00	0.00	0.00
C4F	0.00	0.00	25.00	0.00	75.00	0.00	0.00	0.00	0.00	0.00
X2F	0.00	0.00	0.00	0.00	0.00	90.00	10.00	0.00	0.00	0.00
X3F	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00
B2K	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
CX1K	0.00	0.00	0.00	0.00	12.50	0.00	0.00	0.00	87.50	0.00
GY2	0.00	10.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	90.00

4. Conclusions

In this paper, we described the development of a qualitative discrimination method for the position-GV and grade of intact tobacco leaves using NIR technology. A “multi-region + multipoint” method for collecting the NIR spectra of intact tobacco leaves was proposed. The qualitative discrimination models of tobacco leaf position-GV and grade were constructed and were applied to classify the new tobacco leaves so as to verify their performance. We showed that 1st derivative + SNV was the best preprocessing method for the NIR spectra of intact tobacco leaves and using independent spectra gave higher discriminate accuracy than average spectra. We also indicated that the DPLS algorithm is suitable for the analysis with a small number of classes, whereas Fisher’s algorithm achieves better results when the number of classes is large, and the amount of data in each class is unbalanced. It is concluded that the NIR spectra and the proposed method can realize the effective nondestructive qualitative discrimination of intact tobacco leaves and can be applied in the tobacco industry to assist manual grading and classification.

Data Availability

The raw data in our research cannot be shared at this time as the data also form part of an ongoing study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Key Science and Technology Project of Anhui Tobacco Company under Grant of China (No: 20180563006).

References

- [1] X. L. Li, J. S. Sun, and W. Liang, “A study on building quality assurance system for industrial tobacco raw materials,” *Guangdong Agricultural Sciences*, vol. 11, pp. 142–144, 2008, in Chinese.
- [2] S. L. Wang and G. Y. Wang, “Strategies and effects of upgrading tobacco grades quality,” *Journal of Anhui Agricultural Sciences*, vol. 45, no. 31, pp. 223–225, 2017, in Chinese.
- [3] J. Y. Cai, M. Liang, Y. D. Wen et al., “Analysis of tobacco color and location features using visible-near infrared hyperspectral data,” *Spectroscopy and Spectral Analysis*, vol. 34, no. 10, pp. 2758–2763, 2014.
- [4] K. Yang, J. Y. Cai, C. P. Zhang et al., “Analysis of tobacco site features using near-infrared spectroscopy and projection model,” *Spectroscopy and Spectral Analysis*, vol. 34, no. 12, pp. 2764–2768, 2014.
- [5] General Administration of Quality Supervision, *Inspection and Quarantine of the People’s Republic of China, Standardization Administration of China. Flue-cured tobacco, GB 2635-1992*, Chinese Standard Publishing house, Beijing, China, 1992, in Chinese.
- [6] Y. L. Liu, H. Y. Zhang, H. G. Peng, L. L. Zhao, X. Q. Tao, and J. H. Li, “Reliability and chemical composition analysis of tobacco leaf grade model by near-infrared spectroscopy,” *Spectroscopy and Spectral Analysis*, vol. 40, no. 10, pp. 3260–3264, 2020.
- [7] J. H. Zhou and J. P. Zhang, “On the construction of leaf tobacco safeguarding system for Chinese style cigarette,” *Acta Tabacaria Sinica*, vol. 014, no. 001, pp. 42–46, 2008.
- [8] Z. Y. Pan, Z. Z. Bao, Q. Liu, Y. Chen, and J. Wang, “Current situation and countermeasures of specialization flue-cured tobacco grading team construction in mountainous area, Guizhou,” *Guizhou Agricultural Sciences*, vol. 45, no. 09, pp. 166–170, 2017, in Chinese.
- [9] X. L. Yao, F. Q. He, A. Ping, H. Luo, and Q. M. Guan, “Leaf tobacco grading method based on PCA-GA-SVM,” *Tobacco Science and Technology*, vol. 51, no. 12, pp. 98–105, 2018.
- [10] S. K. Dasari and V. Prasad, “A novel and proposed comprehensive methodology using deep convolutional neural networks for flue cured tobacco leaves classification,” *International Journal of Information Technology*, vol. 11, no. 1, pp. 107–117, 2019.
- [11] Y. H. Lai, Y. Lin, H. Tao, and Y. Wang, “Rapid identification of tobacco mildew based on near infrared spectroscopy and random forest algorithm,” *Tobacco Science and Technology*, vol. 26, no. 02, pp. 40–47, 2020.
- [12] C. Wang, Y. L. Liu, and K. Yang, “Relationship between grade quality and key chemical indexes of flue-cured tobacco by near infrared rapid detection technology,” *China Measurement & Testing Technology*, vol. 47, no. 02, pp. 81–86, 2021.
- [13] S. T. Li, F. Liao, and W. M. He, “Tobacco substitution and cigarette blend maintenance based on near infrared spectral

- similarity," *Tobacco Science and Technology*, vol. 53, no. 2, pp. 88–93, 2020.
- [14] G. T. Xu, H. F. Yuan, and W. Z. Lu, "Progress in modern near infrared spectroscopy and its application," *Spectroscopy and Spectral Analysis*, vol. 02, pp. 134–142, 2000.
 - [15] D. F. Barbin, R. P. Sobottka, W. E. Risso, C. Zucareli, and E. Y. Hirooka, "Influence of plant densities and fertilization on maize grains by near-infrared spectroscopy," *Spectroscopy Letters*, vol. 49, no. 2, pp. 73–79, 2016.
 - [16] Y. Zhang, L. H. Tan, and Y. He, "Determination of water content in watered anhydrous coolant liquid and brake fluid of automobile using fourier transform near-infrared spectroscopy," *Spectroscopy and Spectral Analysis*, vol. 39, no. 07, pp. 2128–2135, 2019.
 - [17] X. Li, J. Bin, W. Fan, J. H. Zhou, and W. R. Chen, "Study on quantitative analysis of samples with different physical states by near infrared spectroscopy hybrid model," *Chinese Journal of Analytical Chemistry*, vol. 45, no. 07, pp. 958–964, 2017.
 - [18] W. Huang, Y. Q. Li, and J. Ge, "Quantitative model transfer between different instruments for total alkaloids in tobacco by near-infrared spectroscopy with consistent and stable wavelengths," *Tobacco Science and Technology*, vol. 53, no. 6, pp. 35–40, 2020.
 - [19] R. L. Li, Y. Y. Liu, W. W. Li, T. C. Li, and J. M. Wang, "Study on rapid determination of tobacco blending uniformity by near infrared spectroscopy," *Food & Machinery*, vol. 35, no. 05, pp. 83–87, 2019.
 - [20] Y. L. Yan, B. Chen, and D. Z. Zhu, *Qualitative Analysis Methods of Near Infrared Spectroscopy*, Chinese Light Industry Press, Beijing, China, 2013.
 - [21] J. J. Workman, P. R. Mobley, B. R. Kowalski, and R. Bro, "Review of chemometrics applied to spectroscopy," *Applied Spectroscopy Reviews*, vol. 31, no. 1-2, pp. 73–124, 1996.
 - [22] Y. Ge, A. Atefi, H. Zhang et al., "High-throughput analysis of leaf physiological and chemical traits with vis-nir-swir spectroscopy: a case study with a maize diversity panel," *Plant Methods*, vol. 15, no. 66, 2019.
 - [23] Y. Wang, X. Ma, Y. Wen, J. Liu, W. Cai, and X. Shao, "Discrimination of plant samples using near-infrared spectroscopy with a principal component accumulation method," *Analytical Methods*, vol. 4, no. 9, pp. 2893–2899, 2012.
 - [24] L. Luan, Y. Wang, X. Li et al., "Application of multiple classifier fusion in the discriminant analysis of near infrared spectroscopy for agricultural products," *Journal of Near Infrared Spectroscopy*, vol. 24, no. 4, pp. 363–372, 2016.
 - [25] J. Bin, F. F. Ai, W. Fan, J. H. Zhou, Y. H. Yun, and Y. Z. Liang, "A modified random forest approach to improve multi-class classification performance of tobacco leaf grades coupled with NIR spectroscopy," *RSC Advances*, vol. 6, no. 36, pp. 30353–30361, 2016.
 - [26] X. Dong, M. A. Qiang, and C. K. Li, "Determination of chemical components of handed tobacco leaves by NIRS," *Chinese Tobacco Science*, vol. 29, no. 4, pp. 10–14, 2008.
 - [27] J. Bin, W. Fan, and R. Liu, "Study on spectral acquisition modes in determining chemical components of intact tobacco leaf by NIRS," *Chinese Tobacco Science*, vol. 39, no. 03, pp. 89–97, 2018.
 - [28] C. R. He, Y. H. Yang, J. H. Li, and C. L. Lao, "Study on measurement method of total alkaloids in freshly roasted intact tobacco leaves by near infrared spectroscopy," *Spectroscopy and Spectral Analysis*, vol. 37, no. 10, pp. 3088–3093, 2017.
 - [29] Y. Zhang and L. Y. He, "Auto-grouping method of flue-cured tobacco leaves based on near infrared spectra technology," *Transactions of the CSAE*, vol. 27, no. 4, pp. 350–354, 2011.
 - [30] P. Liu, X. J. Sui, S. Q. Yang, P. G. Guan, C. Shi, and Y. F. Peng, "Study on the method for judging the anomaly of the online near infrared spectroscopy based on the method of the tobacco leaf," *Journal of Qingdao Agricultural University (Natural Science)*, vol. 34, no. 1, pp. 018–021, 2017.
 - [31] Z. Q. Tang, Y. Liu, R. X. Shu et al., "Quality analysis of the before redrying raw tobacco & after redrying sheet tobacco by using online near infrared spectroscopy," *Spectroscopy and Spectral Analysis*, vol. 34, no. 12, pp. 3273–3276, 2014.
 - [32] W. Jian, G. Liu, and H. Shen, "Distribution of tobacco constituents in tobacco leaf tissue: 4 types general chemical composition," *Chinese Agricultural Science Bulletin*, vol. 27, no. 7, pp. 418–422, 2011.
 - [33] X. D. Wang, H. Z. Gu, D. F. Guo, X. C. Cui, L. Liu, and Y. T. Wei, "Contents and harmony of chemical components in different parts of flue-cured tobacco leaves," *Tobacco Science & Technology*, vol. 54, no. 6, pp. 22–29, 2021.
 - [34] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
 - [35] R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra," *Applied Spectroscopy*, vol. 43, no. 5, pp. 772–777, 1989.
 - [36] I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, NY, USA, Springer Series in Statistics, 2nd edition, 2002.
 - [37] O. Galtier, O. Abbas, Y. Le Dréau et al., "Comparison of PLS1-DA, PLS2-DA and SIMCA for classification by origin of crude petroleum oils by MIR and virgin olive oils by NIR for different spectral regions," *Vibrational Spectroscopy*, vol. 55, no. 32, pp. 132–140, 2011.
 - [38] O. D. Richard and E. H. Peter, *Pattern Classification and Scene Analysis*, John Wiley, New York, NY, USA, 1973.
 - [39] M. Y. Lu, K. Yang, and P. F. Song, "Study on modeling method of tobacco near infrared spectrum classification based on convolution neural network," *Spectroscopy and Spectral Analysis*, vol. 38, no. 12, pp. 3724–3728, 2018.
 - [40] H. H. Zhao and Y. L. Yan, "The effects of noise on NIR analysis and related mathematic pretreatments and models," *Spectroscopy and Spectral Analysis*, vol. 26, no. 5, pp. 842–845, 2006.
 - [41] P. A. Henríquez and G. A. Ruz, "Noise reduction for near-infrared spectroscopy data using extreme learning machines," *Engineering Applications of Artificial Intelligence*, vol. 79, pp. 13–22, 2019.