

Classification of black tea leaf water content based on hyperspectral imaging

Peiye Li

Guangdong Experimental High School, Guangdong Guangzhou 510055, China.

Abstract: Withering is an important process step in black tea production, and the degree of withering directly determines the quality and flavor of black tea. Presently, the judgment of the degree of the withering of black tea at home and abroad mainly depends on the experience of tea masters, which is highly subjective and quickly leads to the inconsistent quality of finished tea. Based on these reasons, this thesis takes Ying Hong No. 9 black tea as the research object, extracts the spectral information of 50 tea leaves in four time periods using hyperspectral imaging technology and an image processing algorithm, and classifies them using a support vector machine, K-value proximity, random forest algorithm, and BP neural network, and the experimental results show that the spectral information of tea leaves in different periods has some variability. The BP neural network in The classification accuracy on the test set reached 73.3%, which was significantly better than the other three algorithms. Therefore, it is feasible to use hyperspectral imaging technology and related classification algorithms to nondestructively identify and classify the water content of black tea leaves during the tail-withering process.

Keywords: Hyperspectral imaging technology; Black tea tail withering; Ying Hong IX; BP neural network.

1. Introduction

Tea is significant to the spread of Chinese culture and economic trade [1]. Foreign countries developed nondestructive testing technology earlier. However, China still has certain defects in tea quality testing and lacks an efficient testing method. Thus, there is a need to develop an immediate and nondestructive testing method. As a necessary substance for plant growth, water participates in chemical reactions such as photosynthesis and respiration as raw material and acts as a solvent for material absorption and transportation and is an important component for maintaining plant cell morphology [2]. In tea, the water content of fresh leaves is 63%~80%, affecting the concentration of other components within the tea, thus affecting the quality of tea [3]. The moisture content of black tea is an important factor in controlling its quality during production. The control of moisture content is one of the main evaluation indicators of black tea tailing. Therefore, measuring the moisture content of tea leaves can not only understand the degree of end fading to a certain extent but also control the processing process of tea leaves to ensure the quality of tea leaves and avoid mold in the storage process [4].

The traditional method of tea testing involves picking tea leaves and bringing them back to the lab for chemical testing. The process is time-consuming and complicated [5]. Zhao et al. proposed a new method for the rapid detection of significant catechins in tea by applying Fourier near-infrared diffuse reflectance spectroscopy analysis using infrared spectroscopy for the rapid detection of catechins in green tea [6]. Zhou et al. conducted a study on the quality evaluation of Dafo Longjing tea based on NIR spectroscopy. Using Dafo Longjing as the analysis object, seven quantitative analysis models were established for the single factor scores of dry tea color, soup color, aroma, leaf base, and the total scores of physical silvers and six factors, respectively, using the partial least squares method of NIR spectroscopy. The results showed that it is feasible to apply NIR spectroscopy for the quality evaluation of Dafo Longjing tea [7]. Chen Xiaojing et

al. proposed a new method for tea classification using color features of multispectral images to classify two kinds of tea leaves that the naked eye could barely distinguish. Wu et al. conducted a study on tea water content and established a regression detection model for tea water content with a correlation coefficient of 0.79 in the experimental test [8]. However, in terms of nondestructive detection of water content in the tail withering process of black tea, less research has been carried out in China, and the use of hyperspectral detection of water content has not been carried out [9-10].

In this study, on the basis of another nondestructive testing, we conducted a survey of the nondestructive testing of leaf water content during the tailing process of black tea, demonstrating the feasibility of hyperspectral testing of tea water content, and proposed a classification testing model. The classification modeling of the water content of black tea during the end-withering process is carried out by hyperspectral technology, and the method of tea water content classification detection based on hyperspectral technology is explored to provide a reference basis for nondestructive testing of tea water content.

2. Materials and methods

2.1. Experimental materials

Hyperspectral imaging-based water content classification study of black tea leaves Ying Hong IX was collected at Ying Hong Farmer Eco-Farm, Ying Hong Town, Yingde City, Guangdong, China. The material for this experiment was compiled on 2022-06-21, 5:30-7:30 a.m., in a breezy and cloudy climate with 85% relative humidity and 33° temperature. Tea trees with good growth conditions and luxuriant branches were randomly selected during the collection process, and the collection standard was one bud and one leaf or one bud and two leaves. 2 kg of samples were collected, preserved in a thermal box and ice bag, and immediately brought back to the laboratory for relevant experiments. Before the experiment, 50 fresh leaves with uniform leaf size and basically the same morphology were

selected and scanned spectrally and subsequently allowed to undergo withering tail treatment in a constant temperature and humidity chamber with a temperature setting of 25° and relative humidity of 90%, and the spectral information was subsequently collected at an interval of one hour. Thus, a total of 200 spectral data were obtained for four-time intervals.

2.2. Experimental equipment

The experiment uses a GaiaField-V10E hyperspectral device, which contains three parts: image acquisition unit, light source, and sample delivery platform, among which the image acquisition unit consists of an image spectrometer CCD camera, and lens, etc. The spectral range of the device's light source is 400-1000 nm, with a spectral resolution of 2.8 nm, containing a total of 360 bands.

2.3. Sample Acquisition

Based on the size of the images that can be acquired by the system, it was decided to acquire ten samples per image, and a total of 5 images were acquired. The initial values of the system were set to 150 ms exposure time and 25 cm object distance without distortion, and the samples were placed on a 30 cm × 30 cm black background according to the principle that different colors absorb light differently to reduce the effect of the background. In order to reduce the interference of the light source caused by the temperature during the experiment, one all-black calibration image and one all-white calibration image were acquired for every image in the experiment.

3. Data processing and classification algorithm

3.1. Spectral image correction

In order to reduce the measurement errors caused by fluctuations of light sources and the quantum efficiency variability of CCD cameras at different wavelengths, the hyperspectral images must be corrected. The image obtained by scanning the standard whiteboard is R. Then, the image is captured by the camera with the lens cover, and the all-black image is scanned without light transmission D. The original image is calibrated using the whiteboard image and the all-black image, and the calibration formula can be defined as follow:

$$I = \frac{I_0 - D}{R - D} \quad (1)$$

where I is the corrected image; I₀ is the original image; R is the whiteboard image; D is the full black image.

3.2. Spectral feature extraction

Due to the huge amount of data contained in hyperspectral images and the huge workload when using professional software for data extraction, this thesis uses the maximum interclass variance algorithm to separate the foreground and background to obtain a binary image and then uses the mask extraction to obtain the ROI region, and uses the algorithm to calculate the average spectrum of the ROI region, the image processed by the algorithm is shown in Fig.1. Fig.2 shows the average spectral curves of Ying Hong IX black tea leaves for four-time intervals. Fig. 2 shows that the differences of the 4 types of curves are relatively obvious and possess differentiability.

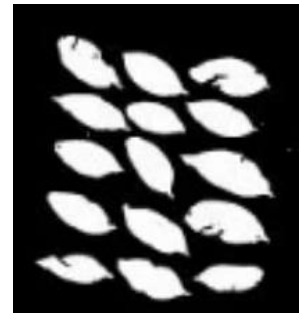


Fig.1 The spectral image processed by the algorithm

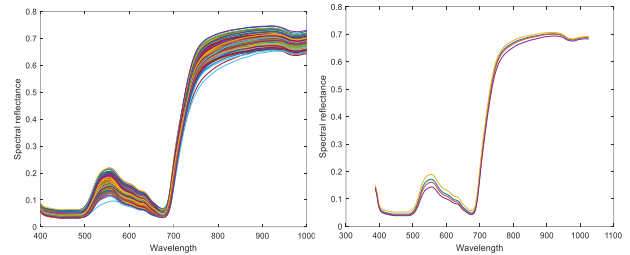


Fig. 2 Primordial spectrogram of red tea

3.3. Spectral pre-processing

When the hyperspectral camera acquires spectral data, the acquired spectral data will have certain errors due to the instrumentation and other factors. Facing this situation, it is generally impossible to build a curve model by traditional methods, so this thesis adopts multiple scattering correction (MSC) for spectral preprocessing, and the results are shown in Fig. 3, and it is obvious from the graph analysis that the preprocessed curve is smoother.

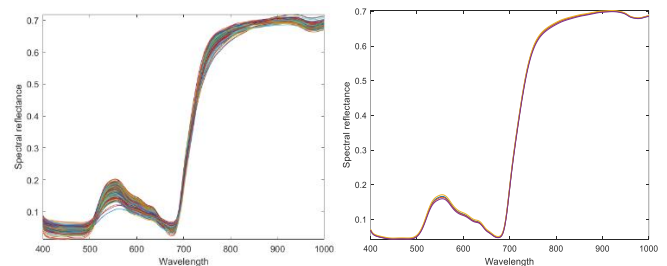


Fig. 3 Primordial spectrogram of red tea

3.4. Classification Algorithms

Support Vector Machine (SVM), also known as support vector machine, is a binary classification model. It can also be used for classification of multi-class problems if modified. Support vector machines can be divided into two categories: linear kernel nonlinear. The main idea is to find a hyperplane in the space that is more than enough to delimit all data samples and to make the distance from all data in this set to this hyperplane the shortest.

The K-Nearest Neighbor (KNN) algorithm is a simple and practical machine learning classification algorithm that can be used for both classification and regression, and the main idea of KNN is to perform classification by measuring the distance between different feature values.

Random Forest (RF) is an algorithm that integrates multiple trees through the idea of integrated learning, and its basic unit is a decision tree, and its essence belongs to a major branch of machine learning - Ensemble Learning method, which is mainly used to solve classification. It is mainly used to solve classification problems.

Back Propagation (BP) neural network is a multi-layer feedforward network trained by error back propagation

algorithm, which is widely used in the present time.

4. Result analysis and discussion

In the experiment, a 5-fold cross-validation method was used to randomly select 70% of the samples in each time period as the training set samples (a total of 140 tea leaves) and the rest as the validation set samples (a total of 60 tea leaves). The average reflected light intensity of 200 tea leaves in 360 bands was used as the feature variable, and the classification recognition model based on four classification learning algorithms, namely SVM, KNN, RF and BP, was established after pre-processing with MSC algorithm.

Table 1. Classification identification results

Algorithms	Time				Total
	0h	1h	2h	3h	
SVM	66.7%	53.3%	86.7%	73.3%	70.0%
KNN	50.0%	33.3%	80.0%	60.0%	57.5%
RF	46.2%	30.0%	86.7%	45.5%	47.5%
BP	46.7%	80.0%	86.7%	80.0%	73.3%

From the experimental results, it can be seen that both BP and SVM algorithms can effectively classify and identify the degree of loss of water content of black tea leaves, and their classification accuracy is 73.3% and 70.0%, respectively, and the classification effect of BP neural network algorithm is slightly better than that of SVM algorithm, and the classification effect of traditional KNN and RF algorithms is not very satisfactory, neither of which exceeds 60%, with RF algorithm performing the worst Fig. 4 shows the confusion matrix of the algorithm classification more visually, and the results are consistent with the experimental statistics. The overall analysis shows that it is difficult to distinguish the degree of water content loss of black tea leaves after 1 h. It is possible that during the experiment, there is a certain amount of error caused by the experimental instrument or operation, resulting in the poor classification effect of the traditional machine learning algorithm, but the BP neural network performance is not good at 0h.



Fig. 4 Confusion matrix of 4 algorithms

5. Conclusion

This thesis investigated the discrimination of the degree of water loss of Ying Hong No. 9 black tea leaves during the tailing process using hyperspectral imaging equipment with wavelength range of 400 -1000nm, image processing methods and relevant classification recognition algorithms. The classification discrimination was achieved by extracting the average relative reflection spectra of the whole area of black tea leaves in the full-band region of interest as classification features, combined with models such as BP neural network, and the classification accuracy of the final BP neural network in the prediction set was 73.3%, which was higher than the other three traditional machine learning algorithms of the algorithm, but the classification accuracy of the whole test set needed further improvement. Therefore, hyperspectral technology can be used to effectively classify and discriminate the degree of water loss in black tea leaves and provide a reference for the application of hyperspectral imaging technology in the black tea tail withering process.

References

- [1] Corretti M C , Deleon Y W , Mangano C , et al. The acute effect of tea consumption with a high fat meal on flow-mediated brachial artery vasodilation in healthy individuals[J]. Journal of the American College of Cardiology, 2002, 39(02):218-218.
- [2] Xiaoli, Li, Chuanqi, et al. Characterizing the Moisture Content of Tea with Diffuse Reflectance Spectroscopy Using Wavelet Transform and Multivariate Analysis[J]. Sensors, 2012, 12(7):9847-9861.
- [3] Wei Y, Wu F, Xu J, et al. Visual detection of the moisture content of tea leaves with hyperspectral imaging technology[J]. Journal of Food Engineering, 2019, 248: 89-96.
- [4] Jabeen S , Alam S , Saleem M , et al. Withering timings affect the total free amino acids and mineral contents of tea leaves during black tea manufacturing[J]. Arabian Journal of Chemistry, 2019, 12(8):s 2411-2417.
- [5] Qu S , Li X , Gan Z . A Review of Hyperspectral Image Classification Based on Joint Spatial-spectral Features[J]. Journal of Physics: Conference Series, 2022, 2203(1):012040.
- [6] Zhao JW, Guo ZM, Chen JS, et al. Rapid determination of catechin in green tea by near-infrared spectroscopy green tea [J]. Journal of Optics, 2008, 28(12) :2302-2306.
- [7] Zhou Xiaofen, Ye Yang, Zhou Zhuding, et al. A study on the quality evaluation of Dafo tea based on near-infrared spectroscopy Longjing tea quality evaluation based on NIR spectroscopy [J]. Spectroscopy and Spectral Analysis, 2012, 32(11): 2971 - 2975.
- [8] Wu Weibin, Liu Wenchao, Li Zeyi, et al. Establishment and experimental study of hyperspectral-based water content detection model of tea leaves[J]. Journal of Henan Agricultural University, 2018, 52(5):7.
- [9] Yw A ,Xi B , Yong H B . Generalisation of tea moisture content models based on VNIR spectra subjected to fractional differential treatment[J]. Biosystems Engineering, 2021, 205:174-186.
- [10] Wei Y, He Y , Li X . Tea moisture content detection with multispectral and depth images[J]. Computers and Electronics in Agriculture, 2021, 183:106082.