



## OPEN ACCESS

## EDITED BY

Huiling Chen,  
Wenzhou University, China

## REVIEWED BY

Alireza Sanaeifar,  
University of Minnesota Twin Cities,  
United States  
Lei Zhou,  
Nanjing Forestry University, China

## \*CORRESPONDENCE

Guijun Yang

✉ yanggj@nercita.org.cn  
Dandan Duan

✉ duandd@nercita.org.cn

RECEIVED 03 July 2023

ACCEPTED 14 August 2023

PUBLISHED 29 August 2023

## CITATION

Cao Q, Zhao C, Bai B, Cai J, Chen L,  
Wang F, Xu B, Duan D, Jiang P, Meng X  
and Yang G (2023) Oolong tea  
cultivars categorization and  
germination period classification  
based on multispectral information.  
*Front. Plant Sci.* 14:1251418.  
doi: 10.3389/fpls.2023.1251418

## COPYRIGHT

© 2023 Cao, Zhao, Bai, Cai, Chen, Wang, Xu,  
Duan, Jiang, Meng and Yang. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that  
the original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Oolong tea cultivars categorization and germination period classification based on multispectral information

Qiong Cao<sup>1,2</sup>, Chunjiang Zhao<sup>1,2</sup>, Bingnan Bai<sup>1</sup>, Jie Cai<sup>1</sup>,  
Longyue Chen<sup>1</sup>, Fan Wang<sup>1</sup>, Bo Xu<sup>1</sup>, Dandan Duan<sup>1\*</sup>,  
Ping Jiang<sup>2</sup>, Xiangyu Meng<sup>1</sup> and Guijun Yang<sup>1\*</sup>

<sup>1</sup>Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China,

<sup>2</sup>Hunan Agricultural University College of Mechanical and Electronical Engineering, Changsha,  
Hunan, China

Recognizing and identifying tea plant (*Camellia sinensis*) cultivar plays a significant role in tea planting and germplasm resource management, particularly for oolong tea. There is a wide range of high-quality oolong tea with diverse varieties of tea plants that are suitable for oolong tea production. The conventional method for identifying and confirming tea cultivars involves visual assessment. Machine learning and computer vision-based automatic classification methods offer efficient and non-invasive alternatives for rapid categorization. Despite advancements in technology, the identification and classification of tea cultivars still pose a complex challenge. This paper utilized machine learning approaches for classifying 18 oolong tea cultivars based on 27 multispectral characteristics. Then the SVM classification model was executed using three optimization algorithms, namely genetic algorithm (GA), particle swarm optimization (PSO), and grey wolf optimizer (GWO). The results revealed that the SVM model optimized by GWO achieved the best performance, with an average discrimination rate of 99.91%, 93.30% and 92.63% for the training set, test set and validation set, respectively. In addition, based on the multispectral information (h, s, r, b, L, Asm, Var, Hom, Dis,  $\sigma$ , S, G, RVI, DVI, VOG), the germination period of oolong tea cultivars can be completely evaluated by Fisher discriminant analysis. The study indicated that the practical protection of tea plants through automated and precise classification of oolong tea cultivars and germination periods is feasible by utilizing multispectral imaging system.

## KEYWORDS

oolong tea cultivar, multispectral characteristics, SVM, identification, and germination

## 1 Introduction

In most nations, tea plant (*Camellia sinensis*) is a significant economic crop, with potential medicinal application (Pan et al., 2022). Most tea plantation management activities are conducted without the utilization of intelligent technology. The procedure of growing and producing tea heavily depends on skilled human workers, resulting in a labor-intensive and inefficient process (Zhang et al., 2023). Identifying tea cultivars plays a crucial role in the tea industry as it directly impacts the commercial tea's yield and quality. There is great variation in taste and quality among different cultivars of the tea plant (*Camellia sinensis*) (Yue et al., 2023; Zaman et al., 2022). Improving the identification and assessment of tea cultivars is essential in tea processing to ensure the production of tea with high quality and yield. The identification of tea plant cultivars is typically reliant on human observation, and it requires expert judgement and time to manually determine the cultivar type. It becomes challenging to differentiate between a series of tea plant cultivars when they appear highly similar, making manual identification difficult. Biochemical and molecular methods used for laboratory analysis are expensive, complex, and can cause damage to the sample. Oolong tea is a famous Chinese tea category, which quality is significantly affected by the cultivar of tea plant (Lin et al., 2022). Most oolong tea products are named after their cultivar names, like Benshan, Dahongpao, Tieluohan. Accurate identification and classification of oolong tea cultivar contribute to the protection and preservation of different genetic resources, preventing the loss and reduction of genetic diversity. Cultivar identification enable targeted quality control measures to ensure oolong tea product consistency and quality stability. Oolong tea cultivars identification provides guidance for cultivation management, including suitable growing environments, fertilization, and pest control measures. Therefore, there is an urgent need for a fast and non-invasive method to identify oolong tea plant cultivars.

The extensive use of remote sensing techniques for plant monitoring in the field has made it feasible to quickly identify tea cultivars. While it is possible to classify various types of vegetation using remotely sensed images, the identification of different cultivars is relatively uncommon. Several studies have differentiated between tea plantations and extract tea plantations using multispectral satellite image, as demonstrated by Chen et al. (Chen et al., 2022) and Zhu et al. (Zhu et al., 2019). Bao et al. (Bao et al., 2023) utilized a UAV platform to detect tea leaf blight, while Tu et al. (Tu et al., 2018) employed hyperspectral information obtained from a UAV to classify tea plant cultivars. The use of thermal images holds promise for applications in plant research. For example, thermal cameras were utilized by Batchuluun et al. (Batchuluun et al., 2022) for the purpose of classifying plants and identifying diseases. Furthermore, the integration of spectral information and image information through spectral imaging analysis, such as hyperspectral (Zhao et al., 2022), near-infrared spectral, and multispectral, has demonstrated significant benefits in nondestructive detection, identification, quality evaluation, and safety control of agricultural products. Zou et al. (Zou et al., 2023) evaluated Mengding mountain green tea varieties using

hyperspectral image. Wang et al. (Wang et al., 2021) applied near-infrared hyperspectral imaging to analyze spatial distribution of total polyphenols in tea. Cao et al. (Cao et al., 2022a) combined hyperspectral and multispectral information to monitor tea plant growth.

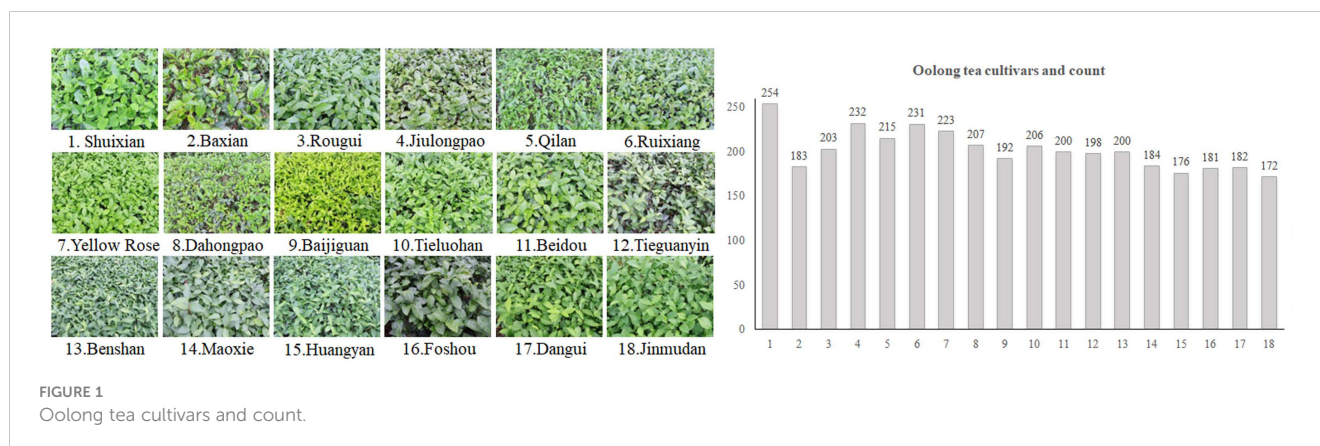
Typical spectral images obtained through multispectral imaging can also offer plentiful information regarding the object of detection. Multispectral imaging (MSI) technology has been demonstrated in numerous studies to allow for non-invasive and unbiased identification of plant phenotyping (Chen et al., 2021), including but not limited to assessing fruit quality (Liu et al., 2015) and distinguishing between different crop varieties (Liu et al., 2016). By utilizing MSI technology, Cao et al. (Cao et al., 2022b) successfully developed a model for distinguishing 16 different types of tea cultivars, achieving highly accurate classification results. The use of multispectral imaging in conjunction with machine learning algorithms in these applications has established a foundation of knowledge and expertise for utilizing MSI technology to accurately identify tea cultivars. Currently, there are few studies on techniques for on-site identification of specific tea varieties, particularly for oolong tea varieties. Therefore, it is urgent to investigate a quick and efficient alternative to the traditional laborious and subjective methods used for categorizing oolong tea cultivars. The aim of this study was to utilize MSI to enhance the identification of tea cultivars for effective tea plant management and to facilitate phenotyping for cultivars with high yield. In this study, multispectral images of the canopies of different oolong tea cultivars were captured using a multispectral camera (RedEdge-MX, Micasense, Seattle, WA, USA). Then, oolong tea cultivars were classified based on the color indicators, spectral characteristics, and texture features of the tea canopies.

## 2 Data and methods

### 2.1 Data acquisition

The experiment was done in a tea farm located in Anxi County, Quanzhou City, Fujian Province, China. Fujian is one of the major producing areas of oolong tea. Anxi has been producing tea for thousands of years and is renowned as the birthplace of Anxi Tieguanyin and “the hometown of Chinese Oolong Tea (Famous Tea)”. It is considered the top tea-producing county in China and is widely recognized as “China's Tea Capital”. The research site possesses a marine monsoon climate of mid-subtropical region in the southern hemisphere, with an average yearly temperature ranging from 19 to 21°C and an annual precipitation of 1600 mm, which renders it an appropriate environment for the growth of tea plants. Figure 1 displays information regarding the oolong tea cultivars examined in this experiment.

The examined set consisted of 18 types of plant cultivars that are suitable for producing oolong tea. These cultivars originated in Fujian, and were subsequently introduced to Guangdong, Zhejiang, and Jiangnan tea regions. Based on their germination period, they can be divided into three stages: early species (Baxian, Yellow Rose, Huangyan, Dangui, Jinmudan), mesophytic species (Baijiguan,



Beidou, Benshan, Maoxie, Foshou), late species (Shuixian, Rougui, Jiulongpao, Qilan, Ruixiang, Dahongpao, Tieluohan, Tieguanyin).

In this experiment, a multispectral camera with five bands, including blue (B), green (G), red (R), near infrared (NIR), and red edge (RED) was used. The center wavelengths of these bands were 475nm, 560nm, 668nm, 840nm, and 717nm. The images were captured on a sunny day between 10:00 am and 3:00 pm to minimize the impacts of illumination changes. The camera was positioned 120cm above tea plant canopy. A barium sulfate ( $\text{BaSO}_4$ ) reference board was placed on the tea plant canopy to calibrate the spectral reflectance. The calibration equation was calculated using:

$$R_i = \frac{DN_i * R_s}{DN_s} \quad (1)$$

where,  $DN_s$  and  $DN_i$  represented the digital number of the stand reference board and multispectral images, respectively, and  $R_s$  and  $R_i$  indicated the reflectance of the standard reference board and multispectral images, respectively.

Over the course of two years, in May 2020 and October 2021, a total of 3639 multispectral images were captured for 18 different oolong tea cultivars.

## 2.2 Multispectral data processing

The initial stage of processing multispectral images involves the registration of images and fusion of bands to correct for any spatial deviation or misalignment that may have occurred during image acquisition due to sensor layout and movement. The SIFT algorithm has proven to be effective in both multispectral image band fusion and information sampling, as it facilitates the automatic selection and matching of feature points within each band image. To eliminate the impact of soil, shadows, and other backgrounds, the raw images were used to isolate the fresh leaves of the tea plant canopy. Previous research has shown that ExGR is effective in extracting plant and crop images. The ExGR value was determined using the following formula:

$$ExGR = (2 * G - R - B) - (1.4 * R - G) \quad (2)$$

The ExGR can be utilized to improve the distinctive features of tea leaves in multispectral images of the tea plant canopy and

enhance their contrast with other objects on the ground, facilitating the differentiation of tea leaves from the background. The Otsu method was employed to segment the images, followed by background masking to isolate the tea leaves of the tea plant canopy. For a comprehensive understanding of the multispectral image processing procedures, please refer to Cao et al. (Cao et al., 2022b).

## 2.3 Multispectral feature extraction

The multispectral images provide abundant information that can be analyzed. Color, texture, and spectral behaviors are the key monitoring indicators of tea plant growth and markedly differ among tea cultivars. For this study, data on color values, texture, and reflectance were obtained from five different bands of tea canopy images. Additionally, various color and vegetation indices were calculated as the primary characteristics. A sum of 86 indicators were chosen for this study, encompassing color value, color indicators, texture data, single band reflectance, and vegetation indices.

### 2.3.1 Color information

For describing color information, we have selected three color models: RGB, HSV, and CIE  $L^*a^*b^*$ . We obtained basic color indicators for each sample by calculating the mean of 9 color values ( $r$ ,  $g$ ,  $b$ ,  $h$ ,  $s$ ,  $v$ ,  $L^*$ ,  $a^*$ ,  $b^*$ ) from masked images. Furthermore, advanced color indicators were derived by computing 7 color indices., as listed in Table 1.

### 2.3.2 Texture information

To quantitatively describe the texture features of the tea plant canopy and differentiate among the 18 oolong tea cultivars, texture analysis methods such as grey-level co-occurrence matrix (GLCM) and local binary patterns (LBP) were utilized. The GLCM was proposed by Haralick et al. (Haralick et al., 1973) in 1970s. The principle of GLCM involves analyzing the relationship between the gray values of pixels in a grayscale image. This is achieved by calculating the probability ( $P_{ij}$ ) of a particular gray value occurring in a fixed pixel, alongside the gray value of another pixel located at a certain distance ( $d$ ) and direction ( $\theta$ ) within the image area. The

TABLE 1 Color indicators compiled from the literature.

CIs	Formula	Reference	CIs	Formula	Reference
LI	$L^* \cdot b^*$	(Wahono et al., 2019)	AI	$b^* \cdot a^*$	(Wahono et al., 2019)
AL	$a^*/L^*$	(Wahono et al., 2019)	AB	$a^*/(L^* \cdot b^*)$	(Wahono et al., 2019)
NDLBI	$(L^* - b^*) / (L^* + b^*)$	(Wahono et al., 2019)	NDALI	$(a^* - L^*) / (a^* + L^*)$	(Wahono et al., 2019)
NDABI	$(a^* - b^*) / (a^* + b^*)$	(Wahono et al., 2019)			

values of  $\theta$  include  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ . By analyzing these relationships, GLCM can provide useful statistical information about the texture of an image, such as contrast, homogeneity, and entropy. The LBP operator is frequently employed to characterize the texture properties of grayscale images due to its ability to maintain gray and rotational invariance. The principle of LBP operator is to compare the intensity values of a central pixel with its surrounding pixels in a circular neighborhood. The pixels are then classified as 1 or 0 based on whether their intensity values are greater than or less than that of the central pixel. This process generates a binary code that represents the local texture of the image. The LBP operator has the advantage of being able to maintain gray invariance, which means that it is not affected by changes in illumination. It is also rotation invariant, meaning that the texture features can be extracted regardless of the orientation of the image. This texture information can then be used for image processing tasks such as feature extraction, segmentation, and classification. This study involved the calculation of 8 features using GLCM and 7 features using LBP, as shown in Table 2.

### 2.3.3 Spectral information

The pigments present in plants mainly influence their reflectance in the visible range, while the reflectance in the NIR range is primarily determined by the cellular structure and canopy

morphology (Yuan et al., 2019). The reflectance values in the Red and NIR bands provide valuable information that can be used to assess the biophysical condition of plants. For the original dataset, we utilized both the reflectance values of individual bands and various vegetation indices. We evaluated a set of 50 traditional vegetation indices that are linked to plant pigments, water levels, plant stress, and other biochemical characteristics including cellulose and lignin. Table 3 lists the formula for calculations.

## 2.4 Feature selection

The primary purpose of feature screening is to decrease the data dimensionality, simplify post-processing, and eliminate irrelevant or incorrect information that may affect the final classification outcomes. Due to the high dimensionality of the extracted data, feature selection was performed prior to modeling. The feature selection technique was utilized to identify a subset of indicators that had minimal collinearity, negligible redundancy, and valuable information to effectively represent all the multispectral data.

Uninformation variable elimination (UVE) has a notable advantage in selecting wavelengths, as it combines both noise and spectral information to extract the characteristic wavelengths of the spectrum. The outcomes of UVE's selection process are more

TABLE 2 Texture indices (TI) calculated by GLCM and LBP.

TI	Formula	Method		TI	Formula	Method	
Asm	$\sum_i \sum_j P(i, j)^2$	GLCM	(3)	$\mu$	$\sum_{g=0}^{L-1} gP(g)$	LBP	(11)
Ent	$-\sum_i \sum_j p(i, j) \log(p(i, j))$	GLCM	(4)	$\sigma$	$\sqrt{\sum_{g=0}^{L-1} (g - \mu)^2 P(g)}$	LBP	(12)
Con	$\sum_{n=0}^{N_x-1} n^2 \{ \sum_{i=1}^{N_x} \sum_{j=1}^{N_x} p(i, j) \}_{ i-j =n}$	GLCM	(5)	S	$\sum_{g=0}^{L-1} (g - \mu)^3 P(g)$	LBP	(13)
Cor	$\frac{\sum_i \sum_j (ij) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$	GLCM	(6)	K	$\frac{1}{\sigma^4} \sum_{g=0}^{L-1} (g - \mu)^4 P(g)$	LBP	(14)
Mea	$\sum_{i,j=0}^{N-1} i(P_{ij})$	GLCM	(7)	G	$\sum_{g=0}^{L-1} P(g)^2$	LBP	(15)
Var	$\sigma_i^2 = \sum_{i,j=0}^{N-1} P_{ij} (i - \mu_i)^2$ $\sigma_j^2 = \sum_{i,j=0}^{N-1} P_{ij} (j - \mu_j)^2$	GLCM	(8)	E	$-\sum_{g=0}^{L-1} P(g) \log_2[P(g)]$	LBP	(16)
Hom	$\sum_{i,j=0}^{N-1} \frac{P_{ij}}{1 + (i - j)^2}$	GLCM	(9)	R	$\frac{1}{1 + \sigma^2}$	LBP	(17)
Dis	$\sum_{i,j=0}^{N-1} P_{ij}  i - j $	GLCM	(10)				

TABLE 3 Vegetation indices compiled from the literature.

VIs	Formula	Reference	VIs	Formula	Reference
NDVI	$(\text{NIR}-\text{R})/(\text{NIR}+\text{R})$	(Main et al., 2011)	RVI	$\text{NIR}/\text{R}$	(ELeblanc, 2000)
DVI	$\text{NIR}-\text{R}$	(ELeblanc, 2000)	EVI	$2.5(\text{B}-\text{G})/(\text{B}+6\text{G}-7.5\text{R}+1)$	(Ahamed et al., 2011)
VOG	$(\text{B}-\text{G})/(\text{R}+\text{RED})$	(VOGELMANN et al., 1993)	MTCI	$(\text{B}-\text{G})/(\text{R}-\text{RED})$	(Main et al., 2011)
GNDVI	$(\text{NIR}-\text{G})/(\text{NIR}+\text{G})$	(Merzlyak, 1996)	RDVI	$(\text{NIR}-\text{RED})/\sqrt{(\text{NIR}+\text{RED})}$	(ELeblanc, 2000)
OSAVI	$1.16(\text{NIR}-\text{RED})/(\text{NIR}+\text{RED}+0.16)$	(Kramer, 2001)	NLI	$(\text{NIR}^2-\text{RED})/(\text{NIR}2+\text{RED})$	(Kramer, 2001)
TGI	$\text{G}-0.39\text{R}-0.61\text{B}$	(Bannari et al., 2011)	ExG	$2\text{G}-\text{R}-\text{B}$	(Mortensen, 1995)
VARI	$(\text{R}-\text{G})/(\text{G}+\text{R}-\text{B})$	(Merzlyak et al., 1999)	NDRE	$(\text{NIR}-\text{RED})/(\text{NIR}+\text{RED})$	(Merzlyak et al., 1999)
WDRVI	$(0.1\text{NIR}-\text{R})/(0.1\text{NIR}+\text{R})$	(Ahamed et al., 2011)	GRVI	$(\text{G}-\text{R})/(\text{G}+\text{R})$	(Ahamed et al., 2011)
PSRI	$(\text{R}-\text{G})/\text{RED}$	(Merzlyak et al., 1999)	PGR	$\text{R}/\text{G}$	(Haboudane et al., 2004)
CCCI	$((\text{NIR}-\text{RED})/(\text{NIR}+\text{RED}))/((\text{NIR}-\text{R})/(\text{NIR}+\text{R}))$	(Haboudane et al., 2004)	MCARI	$((\text{RED}-\text{R}-0.2(\text{RED}-\text{G}))*(\text{RED}/\text{R}))$	(Haboudane et al., 2004)
BGI	$\text{B}/\text{G}$	(Frutos, 2005)	BI	$(\text{NIR}+\text{R}+\text{G})/\sqrt{3}$	(Frutos, 2005)
GI	$\text{G}/\text{R}$	(Frutos, 2005)	SIPI	$(\text{NIR}-\text{B})/(\text{NIR}+\text{R})$	(Main et al., 2011)
PVI	$(1/\sqrt{1.17^2+1})(\text{NIR}-1.17\text{R}-3.37)$	(ELeblanc, 2000)	SAVI	$(\text{NIR}-\text{R})/(\text{NIR}+\text{R}+0.5)*1.5$	(Ahamed et al., 2011)
SR	$\text{NIR}/\text{R}$	(Ahamed et al., 2011)	GDVI	$\text{NIR}-\text{G}$	(Merzlyak et al., 1999)
RI	$\text{RED}/\text{R}$	(Kramer, 2001)	RGI	$\text{R}/\text{G}$	(Ceccato et al., 2001)
BRI	$\text{B}/\text{R}$	(Major et al., 1990)	GMR	$\text{G}-\text{R}$	(Wang et al., 2013)
NRI	$\text{R}/(\text{R}+\text{G}+\text{B})$	(Ansari et al., 2021)	NGI	$\text{G}/(\text{R}+\text{G}+\text{B})$	(Ansari et al., 2021)
INT	$(\text{R}+\text{G}+\text{B})/3$	(Kawashima and N., 1998)	NBI	$\text{B}/(\text{R}+\text{G}+\text{B})$	(Ansari et al., 2021)
NDI	$128(\text{G}-\text{R})/(\text{R}+\text{G})$	(Woebbecke et al., 1993)	WI	$(\text{G}-\text{B})/( \text{R}-\text{G} )$	(Mortensen, 1995)
ExR	$1.4*\text{R}-\text{G}$	(Castillo-Martinez et al., 2020)	ExGR	$\text{ExG}-\text{ExR}$	(Castillo-Martinez et al., 2020)
CIVE	$0.441\text{R}-0.811\text{G}+0.385\text{B}+18.78745$	(Castillo-Martinez et al., 2020)	NGRDI	$(\text{G}-\text{R})/(\text{G}+\text{R})$	(Castillo-Martinez et al., 2020)
VEG	$\text{G}/\text{R}^{2/3}*\text{B}^{1/3}$	(Hague et al., 2006)	COM1	$\text{ExG}+\text{CIVE}+\text{ExGR}+\text{VEG}$	(Lu et al., 2022)
COM2	$0.36\text{ExG}+0.47\text{CIVE}+0.17\text{VEG}$	(Lu et al., 2022)	RGBVI	$(\text{G}^2-\text{R}^2)/(\text{G}^2+\text{R}^2)$	(Bendig et al., 2015)
MGRVI	$(\text{G}^2-\text{R}^2)/(\text{G}^2+\text{R}^2)$	(Bendig et al., 2015)	MExG	$1.262\text{G}-0.884\text{R}-0.311\text{B}$	(Burgos-Artizzu et al., 2011)
NDYI	$(\text{G}-\text{B})/(\text{G}+\text{B})$	(Sulik and Long, 2016)	GLI	$(2\text{G}-\text{R}-\text{B})/(2\text{G}+\text{R}+\text{B})$	(Louhaichi et al., 2001)

straightforward and easier to interpret. The UVE approach utilizes the partial least squares regression coefficient as the primary criterion for wavelength selection (Wang et al., 2020), enabling the identification of valuable wavelengths while removing irrelevant or redundant data. This method ensures that only useful wavelengths are retained for further analysis. Wang et al. (Wang et al., 2022) utilized the UVE method to enhance the predictive accuracy of the nitrogen and carbon content assessment model for maize canopy by utilizing NIR spectra. Shen et al. (Shen et al., 2022) applied UVE algorithm to select variables from THz spectra, for determining the origin of wheat. Previous research has shown that the parameters obtained by UVE method of the prediction mode are superior to the full-spectrum modelling model.

Least Absolute Shrinkage and Selection Operator (LASSO) is a shrinkage estimation algorithm proposed by Tibshirani (Robert, 1996).

It is a regression analysis method that performs both variable selection and regularization to improve the accuracy and interpretability of the model. The principle of LASSO is to minimize the sum of the squared errors between the predicted values and the actual values, subject to a constraint that the sum of the absolute values of the coefficients is less than or equal to a specified constant. This constraint forces some of the coefficients to be exactly zero, effectively performing variable selection and removing irrelevant or redundant features from the model. The LASSO algorithm uses  $\lambda$  to control the strength of the constraint and balance the trade-off between model complexity and predictive accuracy. By adjusting the value of  $\lambda$ , the LASSO algorithm can produce a sequence of models with different numbers of non-zero coefficients. The use of LASSO variable selection has become more popular in the fields of bioinformatics and stoichiometry (Massaro et al., 2023). Based on the advantages of UVE and LASSO algorithms,

this paper introduced them into the feature screening of tea canopy spectral image information.

## 2.5 Classification models of oolong tea cultivar

Previous research has suggested that the SVM algorithm exhibits certain benefits in categorizing different types of tea plants, achieving a relatively high level of precision (Cao et al., 2022b). The SVM algorithm was utilized as the primary classification technique for identifying oolong tea cultivars in this study. One major benefit of the SVM algorithm as a supervised classification approach is that it does not necessitate prior knowledge. The fundamental idea of SVM learning is to determine the separation hyperplane that can effectively divide the training dataset with the greatest margin (Phillips and Abdulla, 2021). This hyperplane is selected to have the largest possible margin, representing the distance between the hyperplane and the nearest data points from each class. By maximizing the margin, SVM aims to achieve improved generalization performance and robustness. The penalty factor “c” and kernel parameter “g” have a great influence on the performance of SVM in tea cultivars classification, directly affecting the accuracy and reliability of classifier model. The optimization problem of SVM involves finding the optimal hyperplane parameters that minimize the classification error while maximizing the margin. This is formulated as a convex optimization problem, which can be solved efficiently using optimization algorithms. Therefore, it is necessary to find an intelligent algorithm that can be employed for optimizing SVM parameters to improve classification accuracy. The current study utilizes and explores GA, PSO, and GWO algorithms to acquire the optimal c and g values, thereby enhancing the accuracy of SVM-based tea cultivar identification.

Genetic algorithm (GA) is a powerful optimization technique inspired by biological evolution and natural selection, which employs adaptive probabilities to search for the optimal solution globally. Initially introduced by Professor Holland as a heuristic search method based on Darwin’s theory of evolution, GA produces the succeeding generation of solutions using heredity operations such as selection, crossover, and mutation. This process ensures that only individuals with high fitness function values are preserved, while those with low values are gradually eliminated. By repeating this cycle, optimal solutions that meet the specified constraints can be obtained, enabling optimization objectives to be achieved. GA has global search capabilities. It effectively searches the parameter space of SVM through genetic operations, making it suitable for fine-tuning SVM model parameters.

Particle swarm optimization (PSO) is derived from the behavior of bird predation patterns. In PSO, each bird is represented as a random particle, characterized by its position and speed. The particle’s initial position and speed are randomly assigned, and through iterative updates, the particle moves towards an optimal solution. During each iteration, the algorithm generates the global

extreme value, which is the optimal solution across the entire population, as well as the individual extreme value, which is the optimal solution for each individual particle. PSO has both global and local search capabilities. It effectively addresses the parameter optimization of SVM by updating particle positions and velocities to explore the parameter space and find the optimal solution. PSO converges quickly and produces good results for SVM parameter optimization. The detail algorithm described SVM optimized by PSO can be referenced in Liu et al. (Liu et al., 2019).

Grey Wolf Optimization Algorithm (GWO) is a nature-inspired algorithm that mimics the hunting behavior of wolves to search for optimal prey. The hunting process involves three main stages: tracking, encircling, and attacking. In the context of parameter optimization, GWO can efficiently search for the optimal penalty parameter c and kernel parameter g for tea tree variety classification model system, thereby achieving the objective of parameter optimization. GWO algorithm demonstrates good convergence and speed and strong global search capabilities. It effectively addresses the parameter optimization of SVM by updating parameter values based on the search behavior of grey wolves, aiming to find the optimal solution, yielding favorable results for SVM parameter optimization. It has been found to be a promising technique for improving accuracy in classification problems by optimizing SVM parameters by GWO. (Barman and Choudhury, 2020).

The datasets were split into training, testing, and validation sets in a 6:2:2 ratio, resulting in a total of 3639 datasets. The model’s accuracy was determined by calculating the ratio of correct samples to total samples for each tea cultivar. All data analysis was performed using Matlab 2017b ((MathWorks Inc., Natick, MA, USA)).

## 3 Results

### 3.1 Characteristic indicators analysis

Initially, a total of 86 characteristics were extracted as input data. However, after applying the UVE algorithm, only 57 indicators remained. Subsequently, the LASSO algorithm was employed following UVE. In the first step, cross-validation was used to select the optimal  $\lambda$  for the model. This ensured minimal fluctuation and stability of the cross-validation error. Ultimately, 27 indicators were retained after UVE-LASSO algorithms. These indicators are h, s, r, b, L\*, Asm, Var, Hom, Dis,  $\sigma$ , S, G, RVI, DVI, VOG, MTCl, NLI, VARI, MCARI, BGI, SR, RI, GMR, ExR, AL, NDLBI, and NDALI. Figure 2 depicts the process of indicators selected by LASSO.

### 3.2 Classification results based on different parameter optimization algorithms

The multispectral imaging system has shown to be a valid method for qualitatively and quantitatively monitoring tea quality

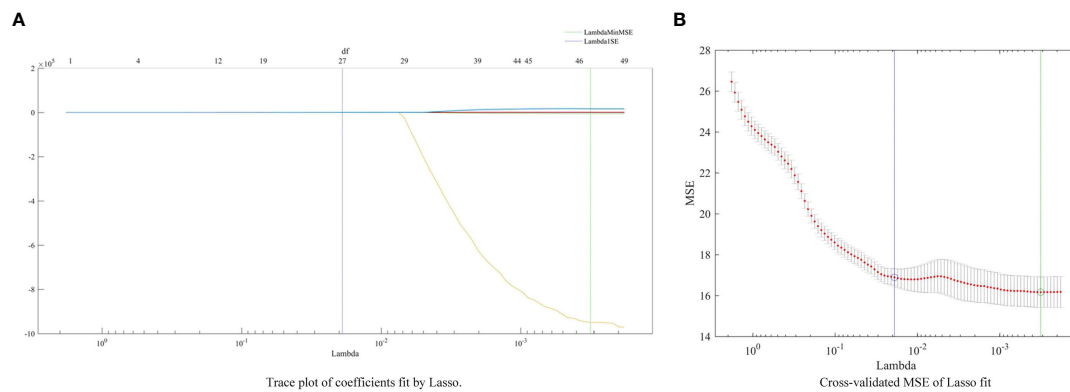


FIGURE 2 The process of indicators selected by LASSO.

(Chen and Yan, 2020; Chen et al., 2021), and can assist in identifying tea plant varieties using the SVM method (Cao et al., 2022b). Initially, the SVM algorithm was applied to classify oolong tea cultivars, and achieved average accuracies of 99.79%, 91.31% and 90.62% for the training, test, and validation sets. However, the identification rate for the Huangyan cultivar was less than 80%. To improve the efficiency of oolong tea cultivars, optimized algorithms, including GA, PSO, and GWO, were proposed to optimize the parameters *c* and *g* of the SVM model. In the GA-SVM model, satisfactory results were obtained with best *c* and *g* values of 31.2540 and 8.3673, respectively, and the average accuracies for the training, testing, and validation sets were 99.96%, 92.43%, and 91.76%, respectively. The cross-validation accuracy was 87.4027%. With the GA-SVM model, the identification accuracy for each oolong tea plant cultivar was above 80%. Results of oolong tea cultivars' identification by the GA-SVM model are shown in the Figure 3.

In the PSO-SVM, the average accuracies for the training, testing, and validation sets were 99.91%, 93.05%, and 92.37%,

respectively, with best *c* and *g* values of 20.7432 and 8.0080. The identification accuracies for the Ruixiang, Benshan, and Foshou cultivars in the testing and validation sets were improved compared to the GA-SVM model. The detailed results of each cultivar identification rate by PSO-SVM are shown in Figure 4.

Lastly, when the GWO algorithm was used to optimize the SVM classification model for identifying oolong tea cultivars, the average accuracies were higher than the PSO-SVM model, with 99.91%, 93.30%, and 92.63% for the training, testing, and validation sets, respectively. In the GWO-SVM, the best *c* and *g* values were 23.7723 and 7.4001, respectively. The identification rates for Shuixian, Ruixiang, and Dahongpao cultivars in the testing and validation sets further improved compared to the PSO-SVM model. However, the identification rates of other cultivars remained the same as the PSO-SVM model. Figure 5 displays the identification results of 18 oolong tea cultivars by GWO-SVM. GWO, based on gray wolf behavior, has advantages over GA and PSO in optimizing SVM models. GWO mimics the behavior of gray wolves, which allows for

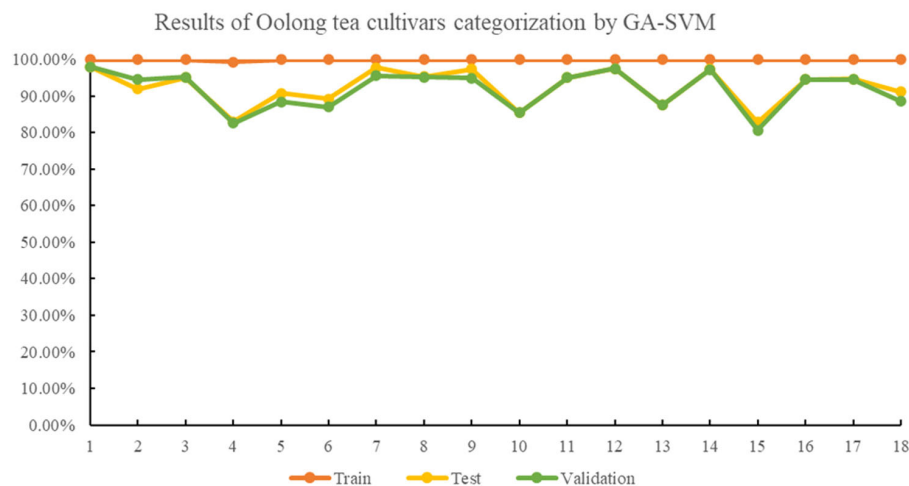


FIGURE 3 Results of oolong tea cultivar categorization by GA-SVM.

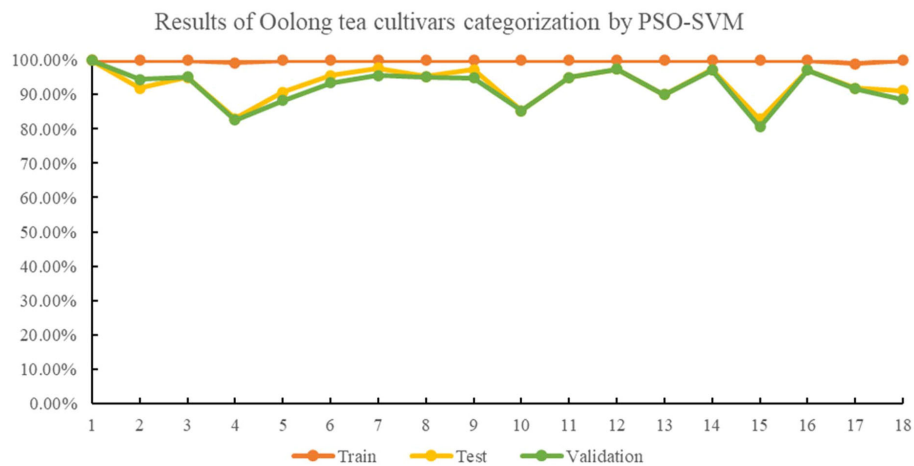


FIGURE 4 Results of oolong tea cultivar categorization by PSO-SVM.

better exploration and utilization of the search space. The algorithm’s simplicity in parameter settings and fast convergence rate also contribute to its potential superiority. Therefore, SVM recognition results optimized by GWO are better than GA and PSO in this study.

### 3.3 Germination period classification of 18 oolong tea cultivars

The 18 selected oolong tea cultivars can be classified into three types, early, mesophytic and late species. Additionally, 27 indicators have been selected for evaluating oolong tea cultivars. The average values of these indicators are used as characteristics for each cultivar. Color, texture, and spectral behaviors of the oolong tea cultivars vary among three germination periods, and Fisher discriminant analysis was used to evaluate the germination period of the cultivars. The mean value of 27 indicators of 18 oolong tea cultivars as input was Fisher discriminated, and the discriminant

functions were as follows:

$$Y_1 = -4.876 \cdot h + 9.114 \cdot s - 18.636 \cdot r + 12.217 \cdot b - 3.143 \cdot L - 5.471 \cdot \text{Asm} - 3.089 \cdot \text{Var} - 16.540 \cdot \text{Hom} + 14.565 \cdot \text{Dis} - 2.819 \cdot \sigma + 5.969 \cdot S + 15.693 \cdot G + 0.101 \cdot \text{RVI} - 0.244 \cdot \text{DVI} + 3.937 \cdot \text{VOG} \quad (3)$$

$$Y_2 = 3.941 \cdot h + 7.469 \cdot s + 12.911 \cdot r - 8.468 \cdot b - 2.090 \cdot L - 4.088 \cdot \text{Asm} - 3.011 \cdot \text{Var} + 7.450 \cdot \text{Hom} - 5.551 \cdot \text{Dis} + 4.832 \cdot \sigma - 3.818 \cdot S - 1.803 \cdot G + 3.083 \cdot \text{RVI} - 6.684 \cdot \text{DVI} + 6.749 \cdot \text{VOG} \quad (4)$$

The class mean projection matrix represents the class center position of the three species. As can be seen in Table 4, tea cultivars of early, mesophytic and late species were correctly identified. The scatterplot, generated by plotting the two discriminant scores of three

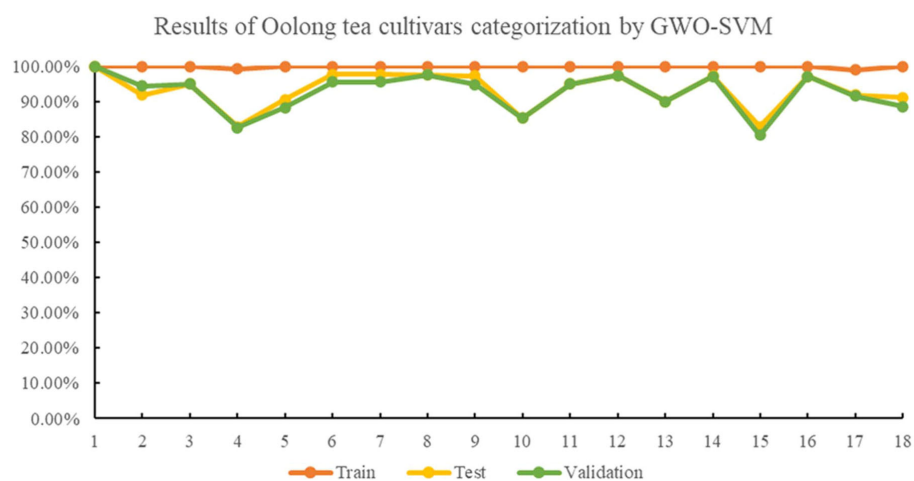


FIGURE 5 Results of oolong tea cultivar categorization by GWO-SVM.



TABLE 4 Discriminant analysis of the corresponding confusion matrix and projection matrix.

Species	Confusion matrix			Class mean projection matrix		
	From/To	Early species	Mesophytic species	Late species	Function 1	Function 2
Early species		5	0	0	0.353	-2.845
Mesophytic species		0	5	0	-7.907	0.984
Late species		0	0	8	4.721	1.163

species, as shown in Figure 6, indicates a significant separation effect among different cultivars. This suggests that indicators (h, s, r, b, L, Asm, Var, Hom, Dis,  $\sigma$ , S, G, RVI, DVI, VOG) distinguishes the early, mesophytic and late species based on Fisher discriminant analysis.

## 4 Discussion

### 4.1 Effects of different scales of datasets on oolong tea cultivar identification

The identification accuracies of the 18 different oolong tea cultivars were found to be above 80% with GA-, PSO-, and GWO-optimized SVM classification models, as seen in Figures 3–5. The GWO-SVM achieved high accuracies for both average identification rate and each species, with the identification rates of Shuixian, Baxian, Rougui, Dahongpao, Tieluohan, Benshan, Maoxie, Huangyan, Dangui, and Jinmudan improving in both test and validation sets. This research involved capturing and analyzing 3639 sets of images to explore how dataset size affects identification accuracy. The 18 oolong tea cultivars dataset was randomly divided into five parts (1/5, 2/5, 3/5, 4/5, and 5/5), and each dataset was then used for GWO-SVM modeling. These datasets were further divided into training, test, and validation sets (at a ratio of 3:1:1). As shown in Figure 7, the accuracies in the test and validation sets were less than 80% when the dataset was only 1, 2, or 3 folds. However, the accuracies of the test and validation sets improved with the increase in data scales, while the identification rate of the training set was not significantly affected by the scale of the dataset.

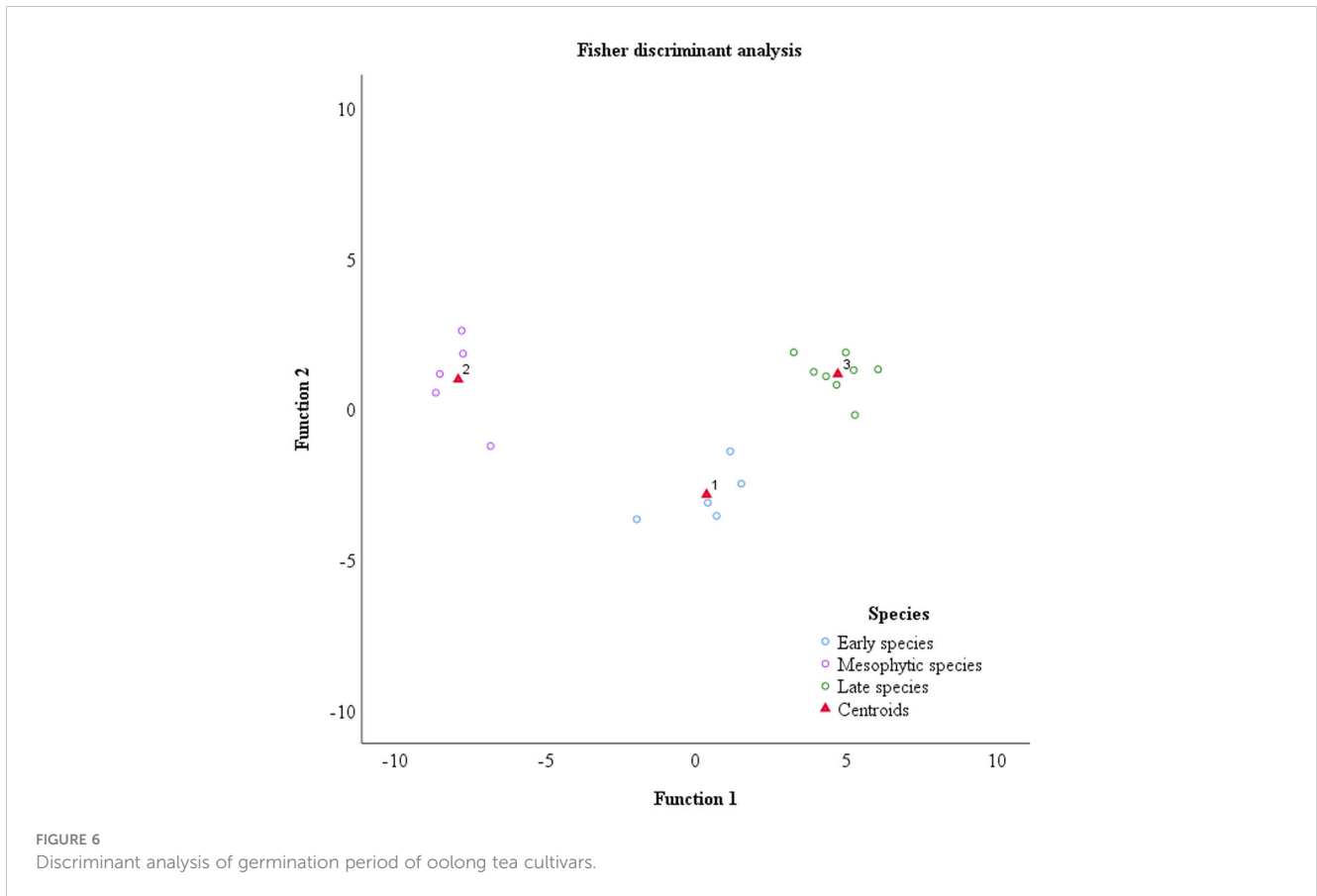
Furthermore, the number of datasets used for each tea variety was reduced by the same amount, and the GWO-SVM was employed to create a tea cultivar recognition model. Figure 8 illustrates that 10 sets of images were randomly selected and reduced for each variety during each trial. Despite minimal changes in the accuracy of the training set, the recognition accuracy of the test and validation sets decreased as the dataset size decreased. When the number of images obtained from each oolong tea cultivar decreased by 60 sets, the recognition accuracy of the model decreased significantly, with the accuracy of the validation set falling below 85%. Similarly, when 90 data sets of each oolong tea cultivar were reduced, the accuracy of the verification set decreased to less than 80%.

### 4.2 Variable importance analysis of selected indicators

This study used UVE-LASSO to select 27 indicators that can be used for oolong tea cultivar identification. To investigate the contribution of these selected indicators, the variable importance in the projection algorithm was employed. Figure 9 shows that all 27 factors are significantly important, with VIP scores  $\geq 0.5$ . Among these,  $\sigma$ , AL, NDALI, s, b, r, G, MTCl, and L contributed the most (VIP > 1.0). Compared to Cao (Cao et al., 2022b), it is evident that  $\sigma$  plays a crucial role in identifying tea plant varieties, including oolong tea cultivars. Additionally, color information plays a more important role in the classification of oolong tea cultivars than other tea plant cultivars.

### 4.3 Analysis of factors influencing the discriminant results

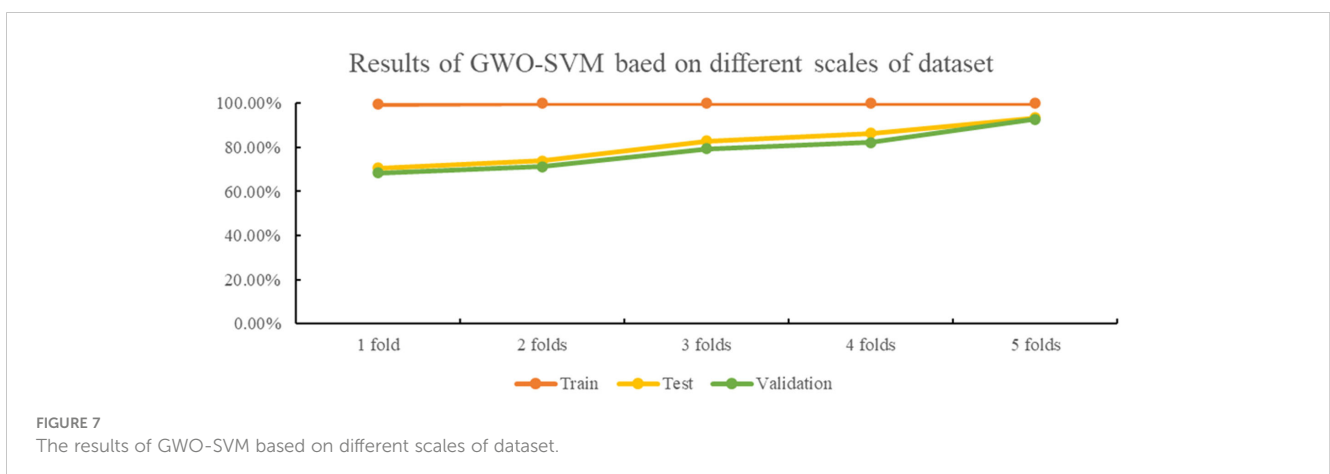
The Baijiguan, Shuixian, Yellow Rose, and Benshan differ greatly in color. Additionally, the texture information of 18 oolong tea cultivars is distinct. The spectral information of five bands is mainly related to quality components such as moisture, polyphenols, chlorophyll, and amino acids. This information varies between different oolong tea cultivars, providing the basis for tea plant identification using a multispectral imaging system. The confusion matrix of the test and validation sets is shown in Figure 10. It can be observed that Baxian was sometimes misjudged as Dahongpao in both the test and validation sets, likely due to their similar color. Figure 9 also indicates that color characteristics play an important role in oolong tea cultivar classification. Jiulongpao cultivar was wrongly discriminated as Baxian, Rougui, Qilan, Ruixiang, Dahongpao, and Tieguan Yin. The majority of wrongly discriminated cultivars were late species. Qilan was consistently discriminated as Baxian, Jiulongpao, Ruixiang, and Huangyan in both the test and validation sets. Ruixiang was misjudged as Qilan and Rougui. Tieluohan was misclassified as Yellow Rose and Dangui, likely due to their similar leaf texture and phenotypic appearance. Benshan was wrongly identified as Huangyan, likely due to their similar canopy structure and appearance. Jinmudan was wrongly discriminated as Foshou, possibly due to the similar texture and appearance of their leaves. Overall, the optimization of the GWO-



SVM model improved the recognition rate of oolong tea cultivars. The factors influencing misjudgment mainly included similar canopy structure, leaf color and texture, and germination period of tea plants. Oolong tea cultivars with the same germination period can be misjudged. This model has demonstrated a high recognition rate specifically for the 18 oolong tea varieties. For other unmentioned cultivars, it is essential to collect data, optimize the model, and readjust the parameters accordingly. Moreover, when expanding the dataset, it is necessary to consider the variations in traits of tea varieties cultivated in different seasons and ecological environments across various regions.

## 5 Conclusion

Both spectral analysis and machine learning methods are valuable for processing multispectral data in tea classification and are effective in extracting useful information. This paper proposes, for the first time, the combination of UVE and LASSO to select proper multispectral indicators for tea cultivar identification and evaluation. The selected 27 factors combined with the GWO-optimized SVM classification model showed good identification. Tea cultivar recognition with computer technology, instead of traditional manual identification, is an inevitable



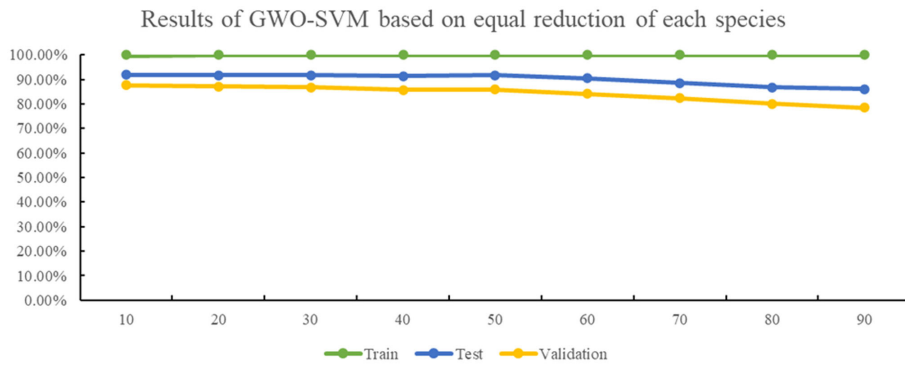


FIGURE 8 The results of GWO-SVM based on equal reduction of each cultivar.

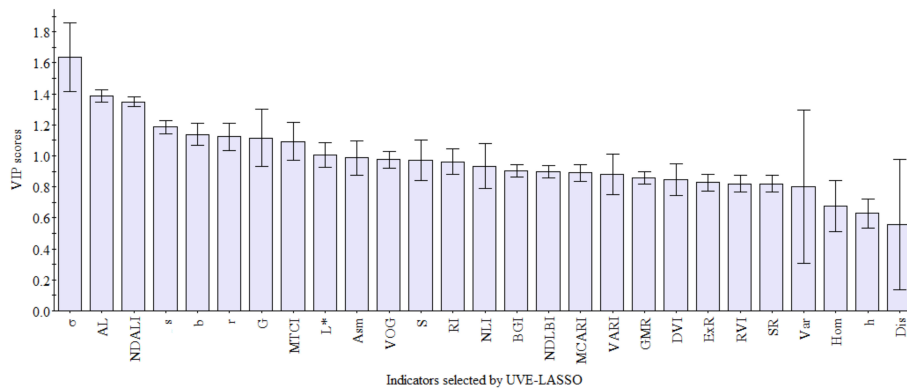


FIGURE 9 VIP scores plot of indicators selected by UVE-LASSO.

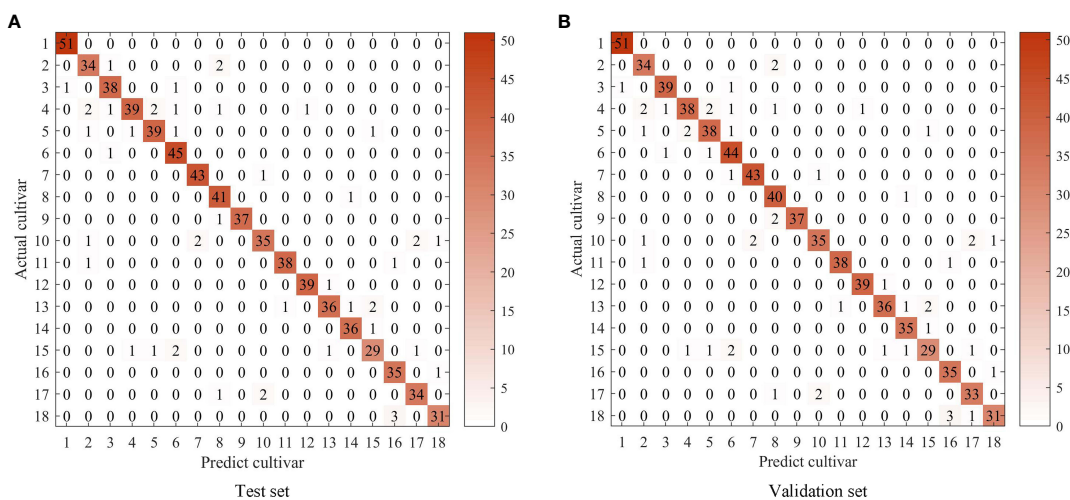


FIGURE 10 Confusion matrix of predict cultivar and actual cultivar in test and validation sets.

attempt. Furthermore, based on the multispectral information, the Fisher discriminate analysis completely divided these oolong tea cultivars into early, mesophytic, and late species, indicating the feasibility of oolong tea cultivar germination classification by effective multispectral information. The rapid and accurate classification and identification of tea cultivars and germination period have practical implications across various areas of the tea industry, including variety conservation and management, tea quality control, market competitiveness, fraud prevention, and scientific research and genetic resource conservation.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

## Author contributions

QC and GY: Conceptualization. BB, JC, LC, and FW: Resources, data curation. QC and BX: Methodology. QC: Writing—original draft preparation. GY and XM: Writing—review and editing. GY and PJ: Supervision. CZ and DD: Funding acquisition. All authors contributed to the article and approved the submitted version.

## Funding

This research was supported by the National Key R&D Program (2021YFD1601103) of China, Chongqing Technology Innovation

and Application Development Special Project (cstc2021jcsxbX0064), the National Key Research and Development Program of China (No.2021YFD2000100, No.2022YFF1003500, No.2021YFD1201601), Guangdong Provincial Science and Technology Innovation Strategy Special Fund (“Big Project + Task List”)—Shaoguan Smart Ecological Tea Garden Construction and Carbon Storage Monitoring Technology Research (DZX202109).

## Acknowledgments

The authors thank Bama Tea Co.,Ltd for providing support for the experiment.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

This study received support such as experimental sites and sampling from Bama Tea Co.,Ltd. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Ahamed, T., Tian, L., Zhang, Y., and Ting, K. C. (2011). A review of remote sensing methods for biomass feedstock production. *Biomass Bioenergy* 35 (7), 2455–2469. doi: 10.1016/j.biombioe.2011.02.028
- Ansari, N., R., S. S., Jahan, A., Ashik-E-Rabbani, M., and Rahman, A. (2021). Inspection of paddy seed varietal purity using machine vision and multivariate analysis. *J. Agric. Food Res.* 3, 10019. doi: 10.1016/j.jagr.2021.100109
- Bannari, A. M., Bonn, F., and Huete, A. R. (2011). Remote sensing leaf chlorophyll content using a visible band index. *Agron. J.* 103 (1-2), 1090–1099. doi: 10.2134/ agronj2010.0395
- Bao, W., Z., Z., Hu, G., Zhou, X., Zhang, D., and Yang, X. (2023). UAV remote sensing detection of tea leaf blight based on DDMA-YOLO. *Comput. Electron. Agric.* 205 (107637). doi: 10.1016/j.compag.2023.107637
- Barman, M., and Choudhury, N. B. D. (2020). A similarity based hybrid GWO-SVM method of power system load forecasting for regional special event days in anomalous load situations in Assam, India. *Sustain. Cities Soc.* 61 (3), 102311. doi: 10.1016/j.scs.2020.102311
- Batchuluun, G., Nam, S. H., and Park, K. R. (2022). Deep learning-based plant classification and crop disease classification by thermal camera. *J. King Saud University-Computer Inf. Sci.* 34 (10), 10474–10486. doi: 10.1016/j.jksuci.2022.11.003
- Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., et al. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Observation Geoinformation* 39, 79–87. doi: 10.1016/j.jag.2015.02.012
- Burgos-Artiz, X. P., Ribeiro, A., Guijarro, M., and Pajares, G. (2011). Real-time image processing for crop/weed discrimination in maize fields. *Comput. Electron. Agric.* 75 (2), 337–346. doi: 10.1016/j.compag.2010.12.011
- Cao, Q., Yang, G. J., Duan, D. D., Chen, L. Y., Wang, F., Xu, B., et al. (2022a). Combining multispectral and hyperspectral data to estimate nitrogen status of tea plants (*Camellia sinensis* (L.) O. Kuntze) under field conditions. *Comput. Electron. Agric.* 198, 107084. doi: 10.1016/j.compag.2022.107084
- Cao, Q., Yang, G. J., Wang, F., Chen, L. Y., Xu, B., Zhao, C. J., et al. (2022b). Discrimination of tea plant variety using in-situ multispectral imaging system and multi-feature analysis. *Comput. Electron. Agric.* 202, 107360. doi: 10.1016/j.compag.2022.107360
- Castillo-Martinez, M. A., Gallegos-Funes, F. J., Carvajal-Gamez, B. E., Urriolagoitia-Sosa, G., and Rosales-Silva, A. J. (2020). Color index based thresholding method for background and foreground segmentation of plant images. *Comput. Electron. Agric.* 178, 105783. doi: 10.1016/j.compag.2020.105783
- Ceccato, P. F., Flasse, Stéphane, Tarantola, Stefano, Jacquemoud, Stéphane, and Grégoire, Jean-Marie (2001). Detecting vegetation leaf water content using reflectance in the optical domain. *Remote Sens. Environ.* 77 (1), 22–23. doi: 10.1016/S0034-4257(01)00191-2
- Chen, L. Y., Xu, B., Zhao, C. J., Duan, D. D., Cao, Q., and Wang, F. (2021). Application of multispectral camera in monitoring the quality parameters of fresh tea leaves. *Remote Sens.* 13 (18). doi: 10.3390/rs13183719
- Chen, B., and Yan, J. L. (2020). Fresh tea shoot maturity estimation via multispectral imaging and deep label distribution learning. *Ice Trans. Inf. Syst.* E103d (9), 2019–2022. doi: 10.1587/transinf.2020EDL8038

- Chen, P. P., Zhao, C. J., Duan, D. D., and Wang, F. (2022). Extracting tea plantations in complex landscapes using Sentinel-2 imagery and machine learning algorithms. *Community Ecol.* 23 (2), 163–172. doi: 10.1007/s42974-022-00077-8
- LEblanc, N. H. (2000). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sens. Environ.* 76 (2), 156–172. doi: 10.1016/S0034-4257(00)00197-8
- Frutos, P. Z.-T. A. B. R. L.-L. J. M. P. M. V. C. M. G. A. (2005). Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sens.* 99 (3), 271–287. doi: 10.1016/j.rse.2005.09.002
- Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., and Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* 90 (3), 337–352. doi: 10.1016/j.rse.2003.12.013
- Hague, T., Tillett, N. D., and Wheeler, H. (2006). Automated crop and weed monitoring in widely spaced cereals. *Precis. Agric.* 7 (1), 21–32. doi: 10.1007/s11119-005-6787-1
- Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *Stud. Media Communication* 6, 610–621. doi: 10.1109/TSMC.1973.4309314
- Kawashima, S., and N., M. (1998). An algorithm for estimating chlorophyll content in leaves using a video camera. *Ann. Bot.* 1, 49–54. doi: 10.1006/anbo.1997.0544
- Kramer, E. (2001). “Assessing the spatial distribution of plant litter,” in *Proceedings of the Tenth JPL Airborne Earth Science Workshop*. 105–114.
- Lin, Y., Yu, W. T., Cai, C. P., Wang, P. J., Gao, S. L., Zhang, J. M., et al. (2022). Rapid varietal authentication of oolong tea products by microfluidic-based SNP genotyping. *Food Res. Int.* 162, 11970. doi: 10.1016/j.foodres.2022.111970
- Liu, W., Guo, G., Chen, F. J., and Chen, Y. H. (2019). Meteorological pattern analysis assisted daily PM2.5 grades prediction using SVM optimized by PSO algorithm. *Atmospheric pollut. Res.* 10 (5), 1482–1491. doi: 10.1016/j.apr.2019.04.005
- Liu, C. H., Liu, W., Chen, W., Yang, J. B., and Zheng, L. (2015). Feasibility in multispectral imaging for predicting the content of bioactive compounds in intact tomato fruit. *Food Chem.* 173, 482–488. doi: 10.1016/j.foodchem.2014.10.052
- Liu, W., Liu, C., Ma, F., Lu, X., Yang, J., and Zheng, L. (2016). Online variety discrimination of rice seeds using multispectral imaging and chemometric methods. *J. Appl. Spectrosc.* 82 (6), 993–999. doi: 10.1007/s10812-016-0217-1
- Louhaichi, M. B., M. M., and Johnson, D. E. (2001). Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto Int.* 16 (1), 65–70. doi: 10.1080/10106040108542184
- Lu, Y. Z., Young, S., Wang, H. F., and Wijewardane, N. (2022). Robust plant segmentation of color images based on image contrast optimization. *Comput. Electron. Agric.* 193, 106711. doi: 10.1016/j.compag.2022.106711
- Main, R., Cho, M. A., Mathieu, R., O’Kennedy, M. M., Ramoelo, A., and Koch, S. (2011). An investigation into robust spectral indices for leaf chlorophyll estimation. *ISPRS J. Photogrammetry Remote Sens.* 66 (6), 751–761. doi: 10.1016/j.isprsjprs.2011.08.001
- Major, D. J., B., F., and Guyot, G. (1990). A ratio vegetation index adjusted for soil brightness. *Int. J. Remote Sens.* 11 (5), 727–740. doi: 10.1080/01431169008955053
- Massaro, A., Bragolusi, M., Tata, A., Zacometti, C., Lefevre, S., Fregiere-Salomon, A., et al. (2023). Non-targeted authentication of black pepper using a local web platform: Development, validation and post-analytical challenges of a combined NIR spectroscopy and LASSO method. *Food Control* 145, 109477. doi: 10.1016/j.foodcont.2022.109477
- Merzlyak, A. A. G. Y. J. K. M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* 58 (3), 289–298. doi: 10.1016/S0034-4257(96)00072-7
- Merzlyak, M. N., G., A. A., Chivkunova, O. B., and Rakitin, V. Y.U. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia Plantarum* 106 (1), 135–141. doi: 10.1034/j.1399-3054.1999.106119.x
- Mortensen, D. M. W. G. E. M. K. V. B. D. A. (1995). Color indices for weed identification under various soil, residue, and lighting conditions. *Trans. Asae* 38 (1), 259–269. doi: 10.13031/2013.27838
- Pan, H., L., M., He, C., Yang, C. S., and Ling, T. (2022). Dark tea: A popular beverage with possible medicinal application. *Chin. Herbal Medicines* 15 (1), 33–36. doi: 10.1016/j.chmed.2022.08.005
- Phillips, T., and Abdulla, W. (2021). Developing a new ensemble approach with multi-class SVMs for Manuka honey quality classification. *Appl. Soft Computing* 111, 107710. doi: 10.1016/j.asoc.2021.107710
- Robert, T. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Society. Ser. B (Methodological)* 58 (1), 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x
- Shen, Y., Li, B., Li, G. L., Lang, C. C., Wang, H. F., Zhu, J., et al. (2022). Rapid identification of producing area of wheat using terahertz spectroscopy combined with chemometrics. *Spectrochimica Acta Part a-Molecular Biomolecular Spectrosc.* 269, 120694. doi: 10.1016/j.saa.2021.120694
- Sulik, J. J., and Long, D. S. (2016). Spectral considerations for modeling yield of canola. *Remote Sens. Environ.* 184, 161–174. doi: 10.1016/j.rse.2016.06.016
- Tu, Y. X., Bian, M., Wan, Y. K., and Fei, T. (2018). Tea cultivar classification and biochemical parameter estimation from hyperspectral imagery obtained by UAV. *PeerJ* 6, e4858. doi: 10.7717/peerj.4858
- VOGELMANN, J. E., R., B. N., and MOSS, D. M. (1993). Red edge spectral measurements from sugar maple leaves. *Int. J. Remote Sens.* 14, 1563–1575. doi: 10.1080/01431169308953986
- Wahono, W., L. D., Sunarminto, B. H., Haryono, E., and Prajitno, D. (2019). CIE L\*a\*b\* Color space based vegetation indices derived from unmanned aerial vehicle captured images for chlorophyll and nitrogen content estimation of tea (*Camellia sinensis* L. Kuntze) leaves. *Ilmu Pertanian (Agricultural Science)* 4, 46–51. doi: 10.22146/ipas.40693
- Wang, Z. L., Chen, J. X., Fan, Y. F., Cheng, Y. J., Wu, X. L., Zhang, J. W., et al. (2020). Evaluating photosynthetic pigment contents of maize using UVE-PLS based on continuous wavelet transform. *Comput. Electron. Agric.* 169, 10516. doi: 10.1016/j.compag.2019.105160
- Wang, Z. L., Chen, J. X., Zhang, J. W., Tan, X. M., Raza, M. A., Ma, J., et al. (2022). Assessing canopy nitrogen and carbon content in maize by canopy spectral reflectance and uninformative variable elimination. *Crop J.* 10 (5), 1224–1238. doi: 10.1016/j.cj.2021.12.005
- Wang, Y. J., Liu, Y., Chen, Y. Y., Cui, Q. Q., Li, L. Q., Ning, J. M., et al. (2021). Spatial distribution of total polyphenols in multi-type of tea using near-infrared hyperspectral imaging. *Lwt-Food Sci. Technol.* 148, 111737. doi: 10.1016/j.lwt.2021.111737
- Wang, Y., Wang, D. J., Zhang, G., and Wang, J. (2013). Estimating nitrogen status of rice using the image segmentation of G-R thresholding method. *Field Crops Res.* 149, 33–39. doi: 10.1016/j.fcr.2013.04.007
- Woebbecke, D. M., M., G. E., Von Bargaen, K., and Mortensen, D. A. (1993). Plant species identification, size, and enumeration using machine vision techniques on near-binary images. *Optics Agric. Forestry*, 1836. doi: 10.1117/12.144030
- Yuan, L., Yan, P., Han, W. Y., Huang, Y. B., Wang, B., Zhang, J. C., et al. (2019). Detection of anthracnose in tea plants based on hyperspectral imaging. *Comput. Electron. Agric.* 167, 105039. doi: 10.1016/j.compag.2019.105039
- Yue, C., C., H., Zhang, S., Hao, Z., Wu, Z., Luo, L., et al. (2023). Aroma characteristics of Wuyi rock tea prepared from 16 different tea plant varieties. *Food chemistry: X17*, 100586. doi: 10.1016/j.fochx.2023.100586
- Zaman, F., Zhang, E., Khattak, W. A., Li, J., Ilyas, M., Deng, X. L., et al. (2022). Natural variations and dynamics of macronutrients for 87 tea plant (*Camellia sinensis*) varieties throughout the growing seasons in Wuhan. *Scientia Hort.* 306. doi: 10.1016/j.scienta.2022.111425
- Zhang, X., W., Z., Cao, C., Luo, K., Qin, K., Huang, Y., et al. (2023). Design and operation of a deep-learning-based fresh tea-leaf sorting robot. *Comput. Electron. Agric.* 206, 107664. doi: 10.1016/j.compag.2023.107664
- Zhao, X. H., Zhang, J. C., Huang, Y. B., Tian, Y. Y., and Yuan, L. (2022). Detection and discrimination of disease and insect stress of tea plants using hyperspectral imaging combined with wavelet analysis. *Comput. Electron. Agric.* 193, 106717. doi: 10.1016/j.compag.2022.106717
- Zhu, J., Pan, Z. W., Wang, H., Huang, P. J., Sun, J. L., Qin, F., et al. (2019). An improved multi-temporal and multi-feature tea plantation identification method using sentinel-2 imagery. *Sensors* 19 (9), 2087. doi: 10.3390/s19092087
- Zou, Z., W., Q., Long, T., Zou, B., Zhou, M., Wang, Y., et al. (2023). Classification and adulteration of mending mountain green tea varieties based on fluorescence hyperspectral image method. *J. Food Composition Anal.* 117 (105141). doi: 10.1016/j.jfca.2023.105141