IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 22, 2019, accepted December 3, 2019, date of publication December 10, 2019,
date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2958882

A Model for Yellow Tea Polyphenols Content Estimation Based on Multi-Feature Fusion

BAOHUA YANG[1,2], YUE ZHU[1], MENGXUAN WANG[1], AND JINGMING NING[3]

[1]School of Information and Computer, Anhui Agricultural University, Hefei 230036, China
[2]National Engineering and Technology Center for Information Agriculture (NETCIA), Nanjing Agricultural University, Nanjing 210095, China
[3]State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei 230036, China

Corresponding authors: Baohua Yang (ybh@ahau.edu.cn) and Jingming Ning (ningjm@ahau.edu.cn)

This work was supported in part by the Natural Science Foundation of Anhui Province under Grant 1808085MF195, in part by the National Key Research and Development Program under Grant 2016YFD0300608, in part by the Natural Science Research Project of Anhui Province under Grant KJ2016A837, in part by the Science and Technology Major Special Project of Anhui Province under Grant 1803071149, in part by the Open Fund of the Key Laboratory of Technology Integration and Application in Agricultural Internet of Things, and the Ministry of Agriculture under Grant 2016KL02, and in part by the National Natural Science Foundation of China under Grant 331772057.

ABSTRACT Polyphenols are one of the important ingredients determining the quality of tea, which play an important role in affecting tea quality standards and quality control. At present, NIR spectroscopy technology has been widely used in tea quality detection to achieve good results. However, due to the lack of spatial features, it is difficult to meet the accuracy requirements to comprehensively judge the quality of both inside and outside of tea. With the development of hyperspectral imaging system (HIS), although progress has been made in the estimation of tea texture based on gray level co-occurrence matrix (GLCM) extraction, there are still some obstacles in practical applications. When the resolution is low, the texture features of the image are not significantly different, and a small number of features cannot more fully represent the image features, resulting in lower model accuracy. When the resolution is higher, the increase in the dimension of the feature leads to a more complex model. Therefore, it is necessary to extract multi-scale features from the hyperspectral image while preserving the original information of the hyperspectral image. In this study, a multi-feature fusion method is proposed to improve the accuracy of tea polyphenols content estimation. The process includes multi-scale wavelet decomposition, and feature fusion based on wavelet coefficient features, GLCM texture features and wavelet texture features. Experiments were carried out on five kinds of yellow tea by comparing different models based on different features, including partial least squares regression (PLSR) and support vector regression (SVR). Results indicated that the model based on multi-feature fusion is more accurate than the individual indicators, and models based on SVR perform better than PLSR.

INDEX TERMS Polyphenols, multi-feature fusion, hyperspectral, wavelet.

I. INTRODUCTION

Yellow tea is special tea in China, which is a micro-fermented tea. Yellow tea is popular among consumers because of its taste and freshness. Tea polyphenol is a general term for polyphenolic compounds in tea. The tea polyphenols content has a great influence on the quality of tea [1]. It could compound with more than 20 kinds of ions, which would produce obvious taste and color change [2]. Moreover, tea polyphenols have various pharmacological effects such as anti-oxidation, anti-inflammatory, anti-tumor, anti-radiation and anti-virus [3]. It is well known that quantitative extraction and detection of tea polyphenols content is important for rational use and physiological effects [4]. Therefore, the establishment of an efficient, accurate and simple tea polyphenols detection method is of great significance in tea production and quality control.

With the rapid development of sensors, non-destructive monitoring technologies such as electronic nose, electronic tongue, machine vision technology, and spectroscopy technology is widely used to detect the composition of tea on the basis of both internal and external characteristic of tea [5]–[12]. However, due to lack of spatial information, the above mention method could not comprehensively judge

The associate editor coordinating the review of this manuscript and approving it for publication was Gang Mei.

the inside and outside quality of the tea, which restricted the in-depth study of tea polyphenols. Especially, it is difficult to accurately determine tea polyphenols, that how quickly it quantify the composition of tea which is still a difficult task. Therefore, it was necessary to establish an effective method for detecting the tea polyphenols content of yellow tea.

In recent years, with the development of imaging spectrometers, more and more researchers are paying attention to hyperspectral images. Hyperspectral imaging technology combine the advantage of traditional imaging and spectroscopy techniques, which could achieve objective and rapid prediction of material composition, due to the simultaneous acquisition of spatial information and spectral information of the object being detected [13]. As a measure of the variation of biophysical parameters and the characteristics of the reflection spectrum, spectral analysis of tea studied due to the geographical location of tea growth or the amount of sunlight absorbed [14]–[18]. Many scholars conducted research on the phenolic components of tea, which mainly concentrated in the study of green tea and black tea [19]–[23], while there has been little research on tea polyphenols in yellow tea. In addition, hyperspectral images have the characteristics of many bands, continuous, narrow spectral range, and large data volume. Therefore, how to achieve feature extraction of high-dimensional spectral feature space was still a challenge.

At present, feature extraction has been the premise and basis for hyperspectral data processing [13]. As another feature of the hyperspectral image, the texture was often used in the tea field because of its rotational invariance and resistance to noise, reflecting the visual characteristics of homomorphism in the image [24]. In particular, the gray-scale co-occurrence matrix was used to study the spatial correlation properties of gray scale. Yang et al. used HIS and gray level co-occurrence matrix (GLCM) to estimate the free amino acids of yellow tea [24]. Mishra et al. used spectral and textural information to classify green tea [25]. Local binary pattern (LBP) and GLCM were integrated to classify fresh tea leaves [26]. GLCM and Gabor filters were combined for tea classification [27]. GLCM was extracted according to the image of the selected wavelength to classify various tea leaves [28]. Although progress has been made to estimate textures of tea based on GLCM extraction, there are still some obstacles in practical applications. Like how to extract detailed texture features and multiscale spectral information from hyperspectral image. In view of the high-dimensional features of hyperspectral images, it was very challenging to mine low-redundant and high-identification effective information from high-dimensional hyperspectral images. Therefore, there was a need to develop powerful methods to improve the validity of spatial feature representation of spectral data.

In the last 20 years, wavelet transform has been successfully applied to the feature extraction of hyperspectral data [29], [30], because it could better represent the smooth region and detail information of the image, which provided a new idea for the extraction and analysis of

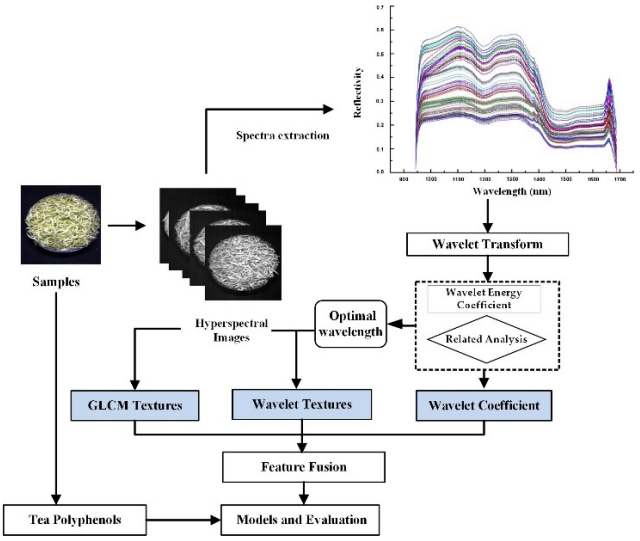

FIGURE 1. The flow chart for yellow tea polyphenols content estimation based on multi-feature fusion.

texture information. Li et al. combined with wavelet transform and GLCM to extract features from multi-spectral images and estimate tea leaf pigments, which achieved good results [31]. However, as far as we know, there is little research on the combination of wavelet transform and GLCM for hyperspectral images of yellow tea.

The wavelet transform could realize not only the dimensionality reduction in the spectral domain, but also the energy concentration and the sparsity of the coefficients in the wavelet transform domain in the spatial domain. Therefore, the tea polyphenols of yellow tea were estimated using wavelet transform in this study. The research work mainly included the following main contributions: (1) to extract the spectrum from the hyperspectral image of yellow tea, which was transformed into wavelet coefficient of different scales by continuous wavelet transform (CWT); (2) to obtain wavelet coefficient features and characteristic wavelengths from wavelet coefficient energy; (3) to extract the GLCM textures and wavelet textures from hyperspectral images at characteristic wavelength; (4) to construct the estimation model based on multi-feature fusion to estimate tea polyphenols content. The flow chart for yellow tea polyphenols content estimation based on multi-feature fusion was shown in Figure 1.

II. DATA AND METHODS

A. EXPERIMENTAL MATERIALS

Five kinds of yellow tea were purchased from the local store, including Pingyang huangtang (PY, produced by Pingyang County, Zhejiang Province, China), Mogan huangya (MG, produced by Deqing, Zhejiang Province, China), Huoshan huangya (HS, produced by Huoshan, Anhui Province, China), Mengding huangya (MD, produced by Mengdingshan, Sichuan Province, China), Junshan yinzhen (JS, produced by Yueyang, Hunan Province, China), which are famous yellow tea in China, and it was difficult for them

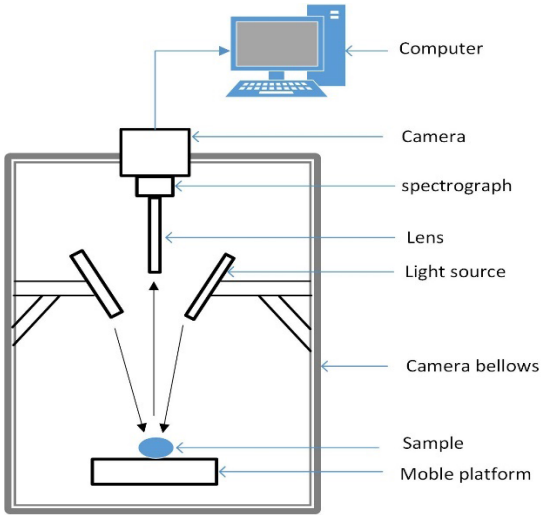

FIGURE 2. Hyperspectral imaging system.

to distinguished directly with the naked eye. Twenty samples were selected for each variety, and a total of 100 samples were used as research objects. The samples were randomly divided into a calibration set and a validation set according to a ratio of 3:1. The calibration set was used to train model, and the validation set was used to verify the performance of the model.

B. HYPERSPECTRAL IMAGING SYSTEM

The hyperspectral image acquisition system used in this experiment that included a spectral imager (Imspector V17E, Spectral Imaging Ltd., Oulu, Finland) and a camera as a CCD camera (IPX-2M30, Imperx Inc., Boca Raton, FL, USA), two 150W halogen lamps (3900, Illumination Technologies Inc., New York, USA), one data acquisition black box, reflective linear tube and electronically controlled displacement platform (MTS120, Beijing Optical Instrument Factory, China), Image acquisition and analysis software (Spectral Image Software, Isuzu Optics Corp., Taiwan, China). The four tungsten halogen lamps of reflected light source were evenly distributed on the ring bracket in the dark box, and the light source is irradiated in a direction of 45° with respect to the vertical direction [24]. HIS is shown in Figure 2.

C. HYPERSPECTRAL IMAGE ACQUISITION

Each yellow tea sample was weighed 20g ± 0.5g using hyperspectral system to obtain 100 sets of hyperspectral data of 636 × 814 × 508. The exposure time and the height of the objective lens were 2ms and 28cm respectively, and the speed of transporting the mobile platform was 8.0mm/s. The spectral imager has a spectral resolution of 5 nm. In order to eliminate the influence of dark current, the originally obtained hyperspectral image was corrected according to formula (1).

$$I_c = \frac{I_r - I_b}{I_w - I_b} \tag{1}$$

TABLE 1. Statistical description of tea polyphenols content (%).

Data Set	Samples	Min	Max	Mean	SD
Whole	100	10.65	21.70	16.81	2.75
Calibration	75	10.65	21.04	16.81	2.82
Validation	25	11.84	21.70	16.83	2.56

where I_c and I_r were the corrected and original hyperspectral image, I_w and I_b were the reference image with white and black image, respectively.

D. TEA POLYPHENOLS CONTENT MEASUREMENT

Weigh 1g from the tea ground sample, which was put it in a 250mL Erlenmeyer flask, by adding 80mL of boiling distilled water and leaching in a boiling water bath for 30 minutes, then pumping and washing, pouring the filtrate into a 100mL volumetric flask, and cooling to room temperature. The tea polyphenols content was calculated according to the formula (2) by using a calibrated potassium permanganate solution.

$$X = \frac{(A - B) \times \omega \times 0.00582/0.318}{m \times \mathrm{V}_1/\mathrm{V}_2} \tag{2}$$

where, X represents the content (%) of tea polyphenols, A and B represent the consumption of potassium permanganate by the yellow tea samples and blank (mL) respectively, and ω represents the concentration of potassium permanganate (%).Where m represents the mass of the sample (g), V1 and V2 indicate the test solution for measurement and the volume of the test solution (mL). To make sure that both data sets were representative of the range of tea polyphenols content. As shown in Table 1, the content of tea polyphenols ranged from 10.65% to 21.70%, with an average of 16.81%, and the data distribution trend of the calibration set was also consistent with the validation set, which indicated that the data distribution of the two data sets were reasonable.

E. MULTI-FEATURE FUSION

Wavelet transform (WT) was a method of signal time-frequency analysis and processing, which was widely used because of overcoming the disadvantage that the window size did not change with frequency. CWT is used to analyze spectrum in detail. It has the ability to decompose continuous, multi-scale original signals. Moreover, the wavelet transform extracts the weak information hidden in the spectral signal by decomposing the spectral signal into sub-signals of different frequencies, so as to realize the overall structural characteristics of the spectral information [32]. Five kinds of yellow tea samples were collected and the hyperspectral images were captured from HSI. The reflectance spectrum of yellow tea from the region of interest was extracted, which was transformed into wavelet coefficients at different scales by CWT.

The wavelet transform is the inner product of a square integrable function f(t) and a wavelet function ψ(t) which has

good local properties in the time-frequency domain:

$$W_f(a,b) = < f, \psi_{a,b} > = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t)\psi^*(\frac{t-b}{a})dt \quad (3)$$

where $<^*,^*>$ denotes the inner product; a > 0, which is the scale factor; b is the displacement factor, * denotes the complex conjugate, $\psi_{a,b}(t)$ is called the wavelet basis function.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}) \quad (4)$$

$\psi(t)$ is called the mother wavelet, and $\psi(t)$ must meet the permissive condition:

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0, \quad \text{or} \quad \int_{-\infty}^{+\infty} \frac{|\psi(\omega)|^2}{|\omega|} d\omega = C_\phi < \infty \quad (5)$$

To reduce the dimension of the feature vector, the statistics of the wavelet coefficients were used as features. In this study, the maximum value of the coefficients of each scale after decomposition was used as a feature. At the same time, the optimal wavelength was screened by the sum of the energy of the wavelet coefficients.

X_i denote the wavelet coefficient after CWT decomposition, i represent different scale factor($i = 2^1, 2^2, 2^3 \ldots$), then the wavelet coefficient energy is the square of the wavelet coefficients at each scale, E_i is the wavelet coefficient energy, and S is the sum of the wavelet coefficient energy.

$$E_i = X_i^2 \quad (6)$$

$$S = \sum_{i=1}^{n} E_i \quad (7)$$

Wavelet transform was also used to decompose images at different scales to obtain different levels of contour information and detail information. A small-scale decomposition of the contour of a scale may represent the contour and detail information at a smaller scale. Usually, an image undergone a wavelet transform to generate three high-frequency sub-band images, including HL, LH, and HH, that represent the high-frequency components of horizontal, vertical and diagonal directions respectively, which reflected the edges, outlines, and textures of the image signal of different directions as mention above.

When the size of image I is $M \times N$, the pixel of the image is (x, y), the wavelet coefficient of the image decomposition is I_{ij}^D, and the image filtered by the formulas (8)-(10).

$$I_{xy}^{LH} = \frac{1}{N_h} \sum_{j=0}^{N_h-1} h(j) I^L_{(x/2-1)[(y+i-1)mod\ N]} \quad (8)$$

$$I_{xy}^{HL} = \frac{1}{N_l} \sum_{i=0}^{N_l-1} l(i) I^H_{(x/2-1)[(y+i-1)mod\ N]} \quad (9)$$

$$I_{xy}^{HH} = \frac{1}{N_h} \sum_{j=0}^{N_h-1} h(j) I^H_{(x/2-1)[(y+i-1)mod\ N]} \quad (10)$$

TABLE 2. GLCM texture features and its formulas.

Texture feature	Formula
Mean	$Mean = \sum_{i,j}^{L-1} ip_{i,j}$
Variance	$Var = \sum_{i,j=0}^{L-1} ip_{i,j}(i - \sum_{i,j}^{L-1} ip_{i,j})^2$
Homogeneity	$Hom = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} \frac{\{p(i,j)\}^2}{1+(i-j)^2}$
Contrast	$Con = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} (i-j)^2 p(i,j)^2$
Dissimilarity	$Dis = \sum_{i,j=0}^{L-1} ip_{i,j} \lvert i-j \rvert$
Entropy	$Ent = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} p(i,j)^2$
Second Moment	$Sm = \sum_{i,j=0}^{L-1} ip_{i,j}^2$
Correlation	$Corr = \sum_{i,j=0}^{L-1} ip_{i,j} \left[\frac{(i - \sum_{i,j}^{L-1} ip_{i,j})(j - \sum_{i,j}^{L-1} ip_{i,j})}{\sqrt{Var_i * Var_j}} \right]$

Where, L represented the size of the window, $p_{i,j}$ represented the gray value of the pixel in the ith row and jth column of the matrix.

where, $l(i)(i = 0, 1, 2, \ldots, N_l - 1)$, $h(j)(j = 0, 1, 2, \ldots, N_h - 1)$ is the impulse response of the low pass and high pass filters. $x = 0, 2, 4 \ldots, M$, $y = 0, 1, 2, \ldots, N$, N_l and N_h is the lengths of the low pass and high pass filters. LH, HL and HH indicate the details at horizontal, vertical and diagonal directions, respectively.

In this study, wavelet texture features expressed by $f = \{E, EN\}$, including energy (E) and entropy (EN) which were calculated according to the equation (11)-(12) based on wavelet coefficients.

$$E = \frac{1}{MN} \sum_{i=1}^{M}\sum_{j=1}^{N} \left| I_{ij}^{D} \right|^2 \quad (11)$$

$$EN = -\frac{1}{MN} \sum_{i=1}^{M}\sum_{j-1}^{N} I_{ij}^{D}(\log 2(I_{ij}^{\Gamma})) \quad (12)$$

where, D = |LH,HL,HH|.

GLCM is a method of describing textures in terms of gray-scale spatial correlation. GLCM involved a matrix function of pixel distance and angle, which reflected the comprehensive information of an image including direction, interval, amplitude of change and speed, to calculate the correlation between a certain distance in the image and a two-point gray scale in a certain direction. In this study, eight features were used to describe the texture features of GLCM [33], as shown in the Table 2.

F. ESTIMATION MODEL AND EVALUATION

Support vector machine (SVM) has been widely used in the classification and regression of spectral data due to its excellent generalization ability and small sample learning ability. SVR is an important branch of SVM. In this paper, by looking for an optimal function, the mathematical expectation of the loss function reached the minimum between the predicted value and the true value of tea polyphenols based on SVR [34].

Partial least squares regression (PLSR) is widely used in many fields because of its simple and robust, small computational complexity and high prediction accuracy [35]. PLSR could extract the strongest comprehensive variable for the dependent variable and eliminate the multiple information by decomposing and screening the original data, which overcome the adverse effects of the independent multi collinearity in modeling.

Correlation analysis is a statistical analysis method for studying the correlation between two or more variables. The correlation coefficient (r) is the relationship between variables. In this study, the correlation coefficient was determined by correlation analysis between wavelet coefficients and tea polyphenols.

Performance of model for the calibration set and validation set was evaluated based on the determination coefficient (R^2) and the root mean square error (RMSE) [24]. In general, R^2 was close to 1, and the model performed better.

III. RESULTS

A. A SPECTRAL FEATURE

The hyperspectral characteristics of all yellow tea samples with tea polyphenols content from 10.65% to 21.7% are shown in Figure 3. It could be seen from Figure 3 that the spectral curve of different yellow tea varieties grew and decremented in the spectral range. The reflectance increased greatly at wavelengths of 1102 nm and 1139 nm. The reflectivity of various varieties of yellow tea tended to be stable at 1450-1650nm. Due to external conditions, such as light intensity, oxygen concentration in the air, instrument system error, etc., a relatively cluttered curve appeared at begin and end of the spectrum curve, which were removed, including 908-943 nm and 1689-1735 nm. However, hyperspectral data has many bands and high correlation between bands, which will lead to data redundancy. Therefore, it was necessary to select representative characteristic wavelengths to achieve model optimization.

B. WAVELET COEFFICIENT FEATURES OF DIFFERENT SCALES

To reveal the correlation between the collected hyperspectral with the polyphenols content of yellow tea. In this study, Daubechies10 was used as a wavelet basis function, a total of 10 layers (1- 10 with step size of 1) of wavelet decomposition experiments were performed. And the hyperspectral reflectance data of the yellow tea sample was

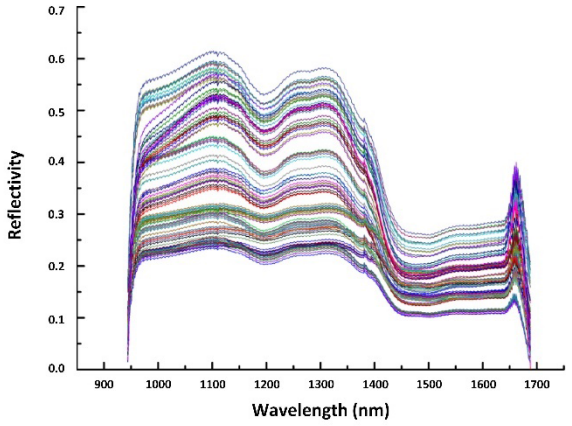

FIGURE 3. Original spectrum of yellow tea samples.

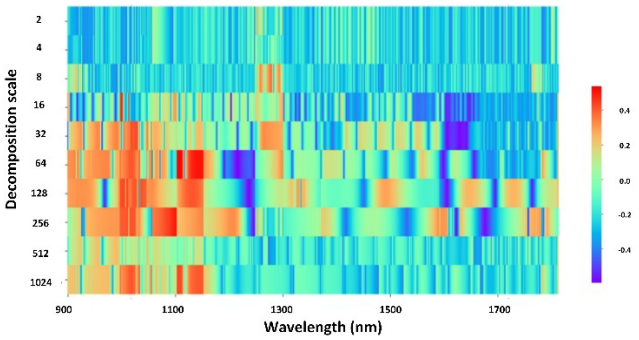

FIGURE 4. Correlation coefficient between wavelet coefficients and tea polyphenols.

decomposed by CWT, which was according to the scales of 2^1, 2^2, 2^3, 2^4, 2^5, 2^6, 2^7, 2^8, 2^9, 2^{10}, the correlation between the transformed wavelet coefficients and the polyphenols content of yellow tea was analyzed. The results are shown in Figure 4, in which the red and blue regions indicated strong correlation bands. After CWT decomposition, the correlation between the wavelet coefficients of various scales and the polyphenols content of yellow tea were improved to some extent.

It can be seen from Figure 4 that as the decomposition scale increase, the correlation of the wavelet coefficients of the original spectral data is larger. Among them, the overall correlation reached the maximum at the sixth decomposition scale. The possible reason was that the information contain in different decomposition scales was related to different factors affecting the reflectivity of tea polyphenols in yellow tea, and the low-scale coefficients reflected the noise of the original spectrum and the slight absorption characteristics. The correlation of low-scale coefficients is small due to the different details of the same type of sample, and the difference between the five kinds of yellow tea was also large.

The high-scale wavelet coefficients contained low-frequency information of the original spectral signal,

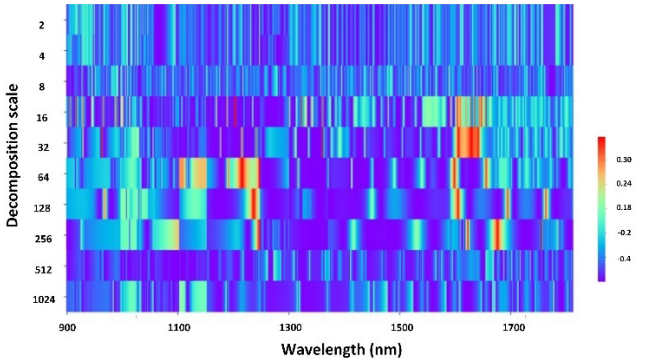

FIGURE 5. Square of the correlation coefficient between wavelet coefficients and tea polyphenols.

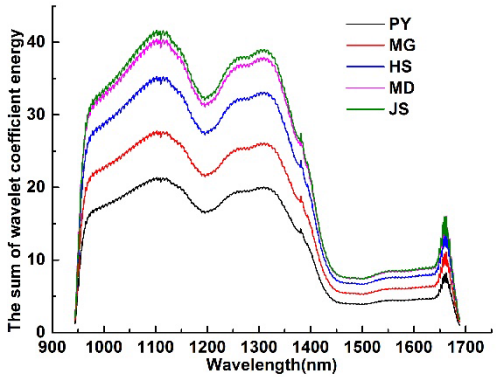

FIGURE 6. The sum of wavelet coefficient energy for yellow tea samples.

reflecting the apparent absorption characteristics of the original spectrum and the structural changes of the tea polyphenols, which determined the shape of the entire original spectrum. As the high-scale coefficients gradually filtered out individual differences in the samples, the correlation between the same type of samples was increasing. The correlations of the wavelet coefficients of the five varieties of yellow tea were also different. However, the all five teas belonged to the yellow tea family. Their different scales of wavelet coefficients were basically consistent with the change trend of tea polyphenols.

Multi-scale wavelet decomposition contributed to the spectral response of tea polyphenols. The low-scale wavelet coefficient features could capture the local absorption characteristics of tea polyphenols, and the high-scale wavelet characteristics could reflect the effect of tea polyphenols on the reflectance amplitude. Therefore, to better optimize the wavelet coefficients, the coefficient graph was further drawn based on the square of the correlation coefficient, as shown in Figure 5. It can be seen that the excellent wavelet coefficients are mainly at 959 nm and 1561 nm of the fourth scale, 1202 nm and 1228 nm of the sixth scale, 1321 nm, 1520 nm and 1540 nm of the fifth scale, and 1535 nm and 1643 nm of the seventh scale, a total of nine wavelet coefficient features, which were used to build the model.

C. OPTIMAL WAVELENGTH

In fact, the wavelet transform decomposes the signal into reconstructed signals of different frequency bands. The energy of the wavelet transform has an equivalence relation with the energy of the original signal. Figure 6 shows the sum of the wavelet coefficient energy, which is calculated by adding the wavelet coefficients of different scales for all samples, and represent the wavelet energy characteristics of the yellow tea of the variety. It was found that sum of the wavelet coefficient energy corresponding to different varieties of yellow tea was consistent in the range of 947-1696 nm. The sum of wavelet coefficient energy was expressed as PY > MG > HS > MD > JS. Among them, the tea polyphenols at the wavelength of 1102 nm and 1309 nm have obvious spectral absorption characteristics, which were selected as optimal wavelengths. Moreover, the spectral reflectance changed with the content of the substance, and the spectral absorption characteristics of tea polyphenols were mainly caused by the frequency doubling and frequency combination of basic chemical bonds such as O-H and C-H in the molecule. The absorption capacity in the spectral range of 1006-1102 nm was gradually enhanced, for the O-H bond in the tea polyphenols was strongly vibrated in the vicinity of the second-order frequency doubling region. In addition, the reflectance of tea polyphenols around 1309 nm began to increase due to the influence of CH2 groups. Therefore, 1102 nm and 1309 nm are characteristic wavelengths of the tea polyphenols content, and every sample could acquire two characteristic wavelengths images.

D. TEXTURE FEATURE EXTRACTION

Some tea samples exhibit similar texture for spatial features. Therefore, to reduce the redundancy feature, it was necessary to extract the GLCM texture features from characteristic wavelengths images. At the same time, in order to extract significant GLCM texture features, including central regions and boundaries, the size of the window during GLCM texture feature extraction should not be too large. Therefore, the GLCM (gray-level co-occurrence matrix) was extracted from the 1102nm and 1309nm grayscale images by ENVI version 4.8 (ITT Visual Information Solutions, Boulder, CO, USA) with default window (3×3), including mean, variance, homogeneity, contrast, dissimilarity, entropy, angular second moment (ASM), and correlation. A total of 16 GLCM texture features were present due to each set of channels from two wavelengths for each tea sample.

Wavelet decomposition with two scales was performed on the characteristic hyperspectral images at 1102 nm and 1309 nm, respectively. The wavelet texture features were extracted from the high-frequency sub-graphs since the high frequency subgraph has good spatial selectivity. First, the wavelet decomposition subgraphs of the high frequency subgraph were extracted, and the results are shown in Figure 7 (Take 1102 nm as an example) in which HL1, LH1 and HH1 represented the first layer of high frequency subgraphs at horizontal, vertical and diagonal directions of wavelet

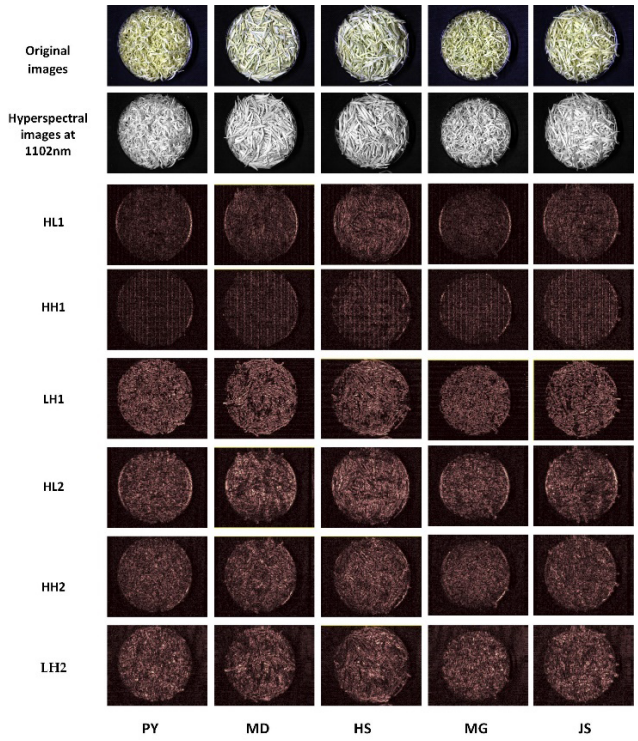

FIGURE 7. Results of wavelet decomposition of characteristic hyperspectral image at 1102 nm. (HL1, LH1 and HH1 represented the first layer of high frequency subgraphs at horizontal, vertical and diagonal directions of wavelet decomposition. HL2, LH2 and HH2 represented the second layer of high frequency subgraphs.)

decomposition. HL2, LH2 and HH2 represented the second layer of high frequency subgraphs, respectively. The wavelet component has directional selectivity, which was consistent with visual characteristics. In addition, the first layer of wavelet subgraphs of the two feature hyperspectral images represented the contour texture and the second layer represented the detail texture. The wavelet textures of the same tea were different from hyperspectral images at different characteristic wavelengths. The wavelet textures of different varieties of tea were also different, especially the difference between the high frequency subgraphs of the second layer which was more obvious.

The multi-scale information obtained by wavelet decomposition not only maintained the spatial characteristics of the original hyperspectral image, but also reflected the high-frequency information of the hyperspectral image. To more intuitively describe the wavelet texture, energy and entropy were extracted from these frequency subgraphs. Two wavelet texture features were extracted from each of the six subgraphs. Therefore, a total of 24 wavelet texture features were obtained from the hyperspectral image at two characteristic wavelengths.

E. TEA POLYPHENOLS CONTENT PREDICTION MODEL

To evaluate the impact of WT on the data in detail, wavelet coefficient features, wavelet texture, GLCM texture, wavelet texture & GLCM texture, fusion data with all features were set as independent variables to build different models, including PLSR and SVR. The results are shown in Table 3. It was found that the effects of the models established by different variables were significantly different. The R^2 of the model constructed using the GLCM texture features were 0.709 and 0.785. Whether based on PLS regression methods or SVR regression methods, models with multi-feature fusion perform best. Among them, the model combining the two texture features was 6.3% and 29.4% higher than the PLSR-based GLCM texture and wavelet texture model respectively, which was 4.3% and 27.5% higher than the SVR-based GLCM texture and wavelet texture model respectively.

We also found that most models of calibration and validation based on SVR method performed better, whether it was a single feature or a multi-feature fusion as the input variable of the model. Moreover, to avoid over fitting, it was necessary to perform Kaiser-Meyer-Olkin (KMO) analysis after the feature variables were extracted. In this study, KMO was 0.697 for wavelet coefficients (9 variables), 0.674 for GLCM textures (16 variables), and 0.706 for wavelet textures (24 variables), which indicated a weak correlation between features. Therefore, these variables were used directly for modeling.

IV. DISCUSSION

In the fusion method based on wavelet transform, the determination of wavelet basis function and wavelet decomposition layer was the key problem to be solved. First of all, in the wavelet basis function selection, the better the regularity and the smaller the vanishing moment, the wavelet base was preferred. Daubechies as a wavelet basis function, which makes the signal reconstruction process smoother. The Daubechies wavelet was characterized by the stronger localization ability in the frequency domain, and the better the division effect of the frequency band, when the vanishing moment order was larger. In addition, with the specific wavelet basis, the result of feature fusion was greatly affected by the number of wavelet decomposition layers. When the wavelet decomposition layer was low (<2), the feature fusion result could not reflect the texture and structure information of the hyperspectral image. As the number of decomposition layers increased the fused image contain more features. The worse the spectral information of the hyperspectral image remains, especially a layer larger than 6. Therefore, six layers of decomposition were selected in this study.

In fact, the performance of wavelet transform has been excellent in this study. Many studies have shown that wavelet transform was successfully applied in the spectral and spatial domains of tea. The texture features that were extracted from the subgraphs based on wavelet decomposition were used to distinguish tea [36]. Wavelet transform was used to enhance the absorption of the spectrum [37]. The wavelet packet transform was used to extract spectral features [38]. WT was used to extract wavelet features from Raman spectroscopy [39]. Discrete wavelet transform was used to extract the spectral characteristics to predict tea moisture content [40].

TABLE 3. Performance of different regression models with different features for prediction of tea polyphenols content.

Models	Features variables	Number of variables	Calibration set		Validation set	
			R^2	RMSE	R^2	RMSE
PLSR	Wavelet coefficients	9	0.774	1.421	0.683	1.149
	GLCM textures	16	0.709	1.544	0.695	1.359
	Wavelet textures	24	0.534	1.931	0.485	1.802
	GLCM & Wavelet textures	40	0.757	1.412	0.705	1.360
	Multi-feature fusion	49	0.856	1.077	0.827	1.037
SVR	Wavelet coefficients	9	0.738	1.739	0.563	1.338
	GLMC textures	16	0.785	1.625	0.757	1.420
	Wavelet textures	24	0.595	1.893	0.592	1.660
	GLCM & Wavelet textures	40	0.821	1.404	0.812	1.131
	Multi-feature fusion	49	0.871	0.830	0.838	0.896

The texture features of the enhanced image could be achieved by the fusion of wavelet textures and GLCM textures [41]. However, few researches have been done on hyperspectral images at characteristic wavelengths. In this study, the spectral data was subjected to wavelet transform processing, and the ''spectral information'' of the hyperspectral image was further explored, so that the ''spectral information'' of the yellow tea hyperspectral image was fully utilized. Moreover, by performing wavelet transform on the image and generating a high-frequency wavelet coefficient image, the image noise could be significantly reduced, and wavelet texture features could be extracted to make up for the lack of detailed information of GLCM texture. Obviously, the ability of a wavelet transform was to represent textures by changing the spatial resolution, and represent the characteristics of the spectrum by extracting the wavelet coefficients.

In addition, comparison by model of the R^2 for the calibration set model constructed using only wavelet coefficients which is 0.774 with PLSR and 0.738 with SVR, which indicate that the wavelet coefficients can reflect the characteristics of the spectrum. Therefore, the spectral-based wavelet transform method can effectively predict the content of tea polyphenols. Furthermore, the model with multi-feature fusion including wavelet coefficient features, GLCM texture features and wavelet texture features is greatly improve, and satisfactory results are obtained regardless of the calibration set or the validation set. From the comparative analysis in Table 3, we also find that the results of different models are also different, and the performance of the SVR model is better than that of the PLSR. The R^2 of the validation set model with SVR reach to 0.838 and RMSE is 0.896. As a result, the accuracy of the model based on multi-feature fusion is significantly improves.

V. CONCLUSION

In this study, five kinds of yellow tea samples were collected, and the hyperspectral images were captured from HSI. Then, the reflectance obtained from hyperspectral images was transformed into wavelet coefficients at different scales by CWT. Wavelet coefficient features and characteristic wavelengths were extracted according to wavelet coefficient energy, GLCM textures and wavelet textures were obtained from hyperspectral images at characteristic wavelength, respectively. The partial least squares regression (PLSR) and support vector regression (SVR) are constructed to estimate the polyphenols content of yellow tea using multi-feature fusion, and for the better performance of SVR model ($R^2 = 0.871$ for calibration, $R^2 = 0.838$ for validation), which accuracy is 1.7% and 1.3% higher than the PLSR model. In addition, for the Calibration set, the accuracy of the model based on Multi-feature fusion is 32.8%%, 9.6%, 29.3% and 3.1% higher than that of wavelet coefficient, GLMC texture features, wavelet texture features and GLCM & Wavelet textures, respectively. In short, a method based on HSI and wavelet transform was provided which could quickly and accurately predict tea polyphenols content of yellow tea, and provide a reference for estimating other tea components.

ACKNOWLEDGMENT

The authors would like to thank S. Ye and S. Hussain for their help with data collection and check. They are grateful to the reviewers for their suggestions and comments, which significantly improved the quality of this article.

REFERENCES

[1] R. Li, K. Jia, X. Chen, and H. Xiao, ''A novel curcuminoid-tea polyphenol formulation: Preparation, characterization, and *in vitro* anti-cancer activity,'' *J. Food Biochem.*, vol. 41, no. 2, 2017, Art. no. e12332.

[2] H. Qin, X. Deng, B. Li, W. Dai, S. Jiao, Y. Qin, and M. Zhang, ''Volatiles, polysaccharides and total polyphenols in Chinese rose tea infusions and their antioxidant activities,'' *J. Food Process. Preservation*, vol. 42, no. 1, 2018, Art. no. e13323.

[3] Y. Qian, S. Zhang, S. Yao, J. Xia, Y. Li, X. Dai, W. Wang, X. Jiang, Y. Liu, and M. Li, ''Effects of vitro sucrose on quality components of tea plants (Camellia sinensis) based on transcriptomic and metabolic analysis,'' *BMC Plant Biol.*, vol. 18, no. 1, p. 121, 2018.

[4] W. Jia, G. Liang, Z. Jiang, and J. Wang, ''Advances in electronic nose development for application to agricultural products,'' *Food Anal. Meth.*, vol. 12, no. 10, pp. 2226–2240, 2019.

[5] Y. Sun, J. Wang, S. Cheng, and Y. Wang, "Detection of pest species with different ratios in tea plant based on electronic nose," *Ann. Appl. Biol.*, vol. 174, no. 2, pp. 209–218, 2019.
[6] S. Ghosh, B. Tudu, N. Bhattacharyya, and R. Bandyopadhyay, "A recurrent Elman network in conjunction with an electronic nose for fast prediction of optimum fermentation time of black tea," *Neural Comput. Appl.*, vol. 31, no. 2, pp. 1165–1171, 2019.
[7] X. Yang, Y. Liu, L. Mu, W. Wang, Q. Zhan, M. Luo, H. Tian, C. Lv, and J. Li, "Discriminant research for identifying aromas of non-fermented Puerh tea from different storage years using an electronic nose," *J. Food Process. Preservation*, vol. 42, no. 10, 2018, Art. no. e13721.
[8] R. Zhi, L. Zhao, and D. Zhang, "A framework for the multi-level fusion of electronic nose and electronic tongue for tea quality assessment," *Sensors*, vol. 17, no. 5, p. 1007, 2017.
[9] J. Jin, S. Deng, X. Ying, X. Ye, T. Lu, and G. Hui, "Study of herbal tea beverage discrimination method using electronic nose," *J. Food Meas. Characterization*, vol. 9, no. 1, pp. 52–60, 2015.
[10] W. Peng, L. Wang, Y. Qian, T. Chen, B. Dai, B. Feng, and B. Wang, "Discrimination of unfermented pu'er tea aroma of different years based on electronic nose," *Agricult. Res.*, vol. 6, no. 4, pp. 436–442, 2017.
[11] D. Lelono, K. Triyana, S. Hartati, and J. Istiyanto, "Classification of Indonesia black teas based on quality by using electronic nose and principal component analysis," in *Proc. Adv. Sci. Technol. Soc., 1st Int. Conf. Sci. Technol.*, vol. 1755, 2016, Art. no. 020003.
[12] S. Sarkar, P. Bhondekar, M. Macaš, R. Kumar, R. Kaur, A. Sharma, A. Gulati, and A. Kumar, "Towards biological plausibility of electronic noses: A spiking neural network based approach for tea odour classification," *Neural Netw.*, vol. 71, no. 7, pp. 142–149, 2015.
[13] Q. Chen, C. Zhang, J. Zhao, and Q. Ouyang, "Recent advances in emerging imaging techniques for non-destructive detection of food quality and safety," *Trends Anal. Chem.*, vol. 52, pp. 261–274, Dec. 2013.
[14] X. Zhou, J. Sun, X. Wu, B. Lu, N. Yang, and C. Dai, "Research on moldy tea feature classification based on WKNN algorithm and NIR hyperspectral imaging," *Spectrochimica Acta A, Mol. Biomolecular Spectrosc.*, vol. 206, no. 7, pp. 378–383, 2019.
[15] L. Li, S. Xie, J. Ning, Q. Chen, and Z. Zhang, "Evaluating green tea quality based on multisensor data fusion combining hyperspectral imaging and olfactory visualization systems," *J. Sci. Food Agricult.*, vol. 99, no. 4, pp. 1787–1794, 2019.
[16] Y. Tu, M. Bian, Y. Wan, and T. Fei, "Tea cultivar classification and biochemical parameter estimation from hyperspectral imagery obtained by UAV," *PeerJ*, vol. 6, no. 5, 2018, Art. no. e4858.
[17] K. Timothy, R. Jinchang, and M. Stephen, "Effective classification of Chinese tea samples in hyperspectral imaging," *Artif. Intell. Res.*, vol. 2, no. 4, pp. 1927–6974, 2013.
[18] J. Zhao, Q. Chen, J. Cai, and Q. Ouyang, "Automated tea quality classification by hyperspectral imaging," *Appl. Opt.*, vol. 48, no. 19, pp. 3557–3564, 2009.
[19] H. Sun, Y. Chen, M. Cheng, X. Zhang, X. Zheng, and Z. Zhang, "The modulatory effect of polyphenols from green tea, oolong tea and black tea on human intestinal microbiota *in vitro*," *J. Food Sci. Technol.*, vol. 55, no. 1, pp. 399–407, 2018.
[20] N. Hocker, C. Wang, J. Prochotsky, A. Eppurath, L. Rudd, and M. Perera, "Quantification of antioxidant properties in popular leaf and bottled tea by high-performance liquid chromatography (HPLC), spectrophotometry, and voltammetry," *Anal. Lett.*, vol. 50, no. 10, pp. 1640–1656, 2017.
[21] X. J. Yu, K. S. Liu, Y. He, and D. Wu, "Color and texture classification of green tea using least squares support vector machine (LSSVM)," *Key Eng. Mater.*, vol. 460, no. 1, pp. 774–779, 2011.
[22] H. Pan, D. Zhang, B. Li, Y. Wu, and Y. Tu, "A rapid UPLC method for simultaneous analysis of caffeine and 13 index polyphenols in black tea," *J. Chromatograph. Sci.*, vol. 55, no. 5, pp. 495–496, 2017.
[23] G. Tenore, P. Campiglia, D. Giannetti, and E. Novellino, "Simulated gastrointestinal digestion, intestinal permeation and plasma protein interaction of white, green, and black tea polyphenols," *Food Chem.*, vol. 169, no. 169, pp. 320–326, 2015.
[24] B. Yang, Y. Gao, H. Li, S. Ye, H. He, and S. Xie, "Rapid prediction of yellow tea free amino acids with hyperspectral images," *PLoS ONE*, vol. 14, no. 2, pp. 1–17, 2019.
[25] D. Dutta, P. Das, U. Bhunia, U. Singh, S. Singh, J. Sharma, and V. Dadhwal, "Retrieval of tea polyphenol at leaf level using spectral transformation and multi-variate statistical approach," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 36, no. 11, pp. 22–29, 2015.
[26] Z. Tang, Y. Su, M. Er, F. Qi, L. Zhang, and J. Zhou, "A local binary pattern based texture descriptors for classification of tea leaves," *Neurocomputing*, vol. 168, no. 5, pp. 1011–1023, 2015.
[27] Ö. Akar and Ö. Güngör, "Integrating multiple texture methods and NDVI to the Random Forest classification algorithm to detect tea and hazelnut plantation areas in northeast Turkey," *Int. J. Remote Sens.*, vol. 36, no. 2, pp. 442–464, 2015.
[28] J. Ning, J. Sun, S. Li, M. Sheng, and Z. Zhang, "Classification of five Chinese tea categories with different fermentation degrees using visible and near-infrared hyperspectral imaging," *Int. J. Food Properties*, vol. 20, no. 2, pp. 1515–1522, 2017.
[29] L. M. Bruce, J. Li, and Y. Huang, "Automated detection of subpixel hyperspectral targets with adaptive multichannel discrete wavelet transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 4, pp. 977–980, Apr. 2002.
[30] T. Cheng, B. Rivard, and G. Sánchez-Azofeifa, "Spectroscopic determination of leaf water content using continuous wavelet analysis," *Remote Sens. Environ.*, vol. 115, no. 2, pp. 659–670, 2011.
[31] X.-L. Li, Y. He, and Z.-J. Qiu, "Textural feature extraction and optimization in wavelet sub-bands for discrimination of green tea brands," in *Proc. Int. Conf. Mach. Learn. Cybern.*, vol. 3, 2008, pp. 1461–1466.
[32] S. A. Ouadfeul and L. Aliouane, "Random seismic noise attenuation data using the discrete and the continuous wavelet transforms," *Arabian J. Geosci.*, vol. 7, no. 7, pp. 2531–2537, 2014.
[33] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
[34] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 3, pp. 247–259, 2011.
[35] Y. Fu, G. Yang, J. Wang, X. Song, and H. Feng, "Winter wheat biomass estimation based on spectral indices, band depth analysis and partial least squares regression using hyperspectral measurements," *Comput. Electron. Agricult.*, vol. 100, pp. 51–59, Jan. 2014.
[36] G. Gill, A. Kumar, and R. Agarwal, "Nondestructive grading of black tea based on physical parameters by texture analysis," *Biosyst. Eng.*, vol. 116, no. 2, pp. 198–204, 2013.
[37] J. Zhang and Z. Zhao, "A method for determination of thiamethoxam in tea infusion by wavelet transform of self-enhanced absorption spectrum," *Food Anal. Methods*, vol. 10, no. 3, pp. 659–665, 2017.
[38] X. Li, L. Luo, Y. He, and N. Xu, "Determination of dry matter content of tea by near and middle infrared spectroscopy coupled with wavelet-based data mining algorithms," *Comput. Electron. Agricult.*, vol. 98, no. 7, pp. 46–53, 2013.
[39] X. Li, C. Sun, L. Luo, and Y. He, "Nondestructive detection of lead chrome green in tea by Raman spectroscopy," *Sci. Rep.*, vol. 5, no. 10, 2015, Art. no. 15729.
[40] X. Li, C. Xie, Y. He, Z. Qiu, and Y. Zhang, "Characterizing the moisture content of tea with diffuse reflectance spectroscopy using wavelet transform and multivariate analysis," *Sensors*, vol. 12, no. 7, pp. 9847–9861, 2012.
[41] A. Bakhshipour, A. Sanaeifar, S. Payman, and M. Guardia, "Evaluation of data mining strategies for classification of black tea based on image-based features," *Food Anal. Methods*, vol. 11, no. 4, pp. 1041–1050, 2018.

BAOHUA YANG was born in Yilan, Heilongjiang, China, in 1974. She received the master's degree in computer application from the Hefei University of Technology, China.

She was a Visiting Scholar with Kansas State University, USA, for three months. Since 2008, she has been an Associate Professor at the School of Information and Computer Science, Anhui Agricultural University. Her current research interests include machine learning, pattern recognition, and hyperspectral image analysis.

YUE ZHU was born in Anqing, Anhui, China, in 1996. She received the B.S. degree in communication engineering from the School of Economics and Technology, Anhui Agricultural University. She is currently pursuing the M.S. degree in computer science and technology with the College of Information and Computer Science, Anhui Agricultural University.

Her current research interests include machine learning and intelligent detection.

MENGXUAN WANG was born in Suzhou, Anhui, China, in 1995. He received the B.S. degree in software engineering from the Suzhou College. He is currently pursuing the M.S. degree in agricultural engineering and information technology with the College of Information and Computer Science, Anhui Agricultural University.

His current research interests include machine learning and intelligent detection.

JINGMING NING was born in Anhui, China, in 1975. He received the Ph.D. in tea science from Anhui Agricultural University, China, in 2011. He is currently the Director of the Institute of Tea and Food Industrialization Technology, Anhui Agricultural University, the Deputy Secretary-General of the Yellow Tea Working Group of the China Tea Standardization Technical Committee, and an Expert of the Processing and Resource Utilization of the Tea Industry System in Anhui Province of China. He is also a member of the Tea Processing Committee of the China Tea Society. His current research interests include tea processing and tea quality analysis.