Review Article
The Quality Control of Tea by Near-Infrared Reflectance (NIR) Spectroscopy and Chemometrics

Ming-Zhi Zhu,1,2,3,4 Beibei Wen,1,2,3,4 Hao Wu,5 Juan Li,1 Haiyan Lin,2 Qin Li,3 Yinhua Li,2 Jianan Huang,1,2,3 and Zhonghua Liu2,3,4

1Key Laboratory of Tea Science of Ministry of Education, Hunan Agricultural University, Changsha 410128, China
2Hunan Provincial Key Laboratory of Crop Germplasm Innovation and Utilization, Hunan Agricultural University, Changsha 410128, China
3National Research Center of Engineering Technology for Utilization of Functional Ingredients from Botanicals, Hunan Agricultural University, Changsha 410128, China
4Collaborative Innovation Centre of Utilization of Functional Ingredients from Botanicals, Hunan Agricultural University, Changsha 410128, China
5Key Laboratory of Tea Plant Biology of Henan Province, College of Life Science, Xinyang Normal University, Xinyang 464000, China

Correspondence should be addressed to Jianan Huang; jian7513@sina.com and Zhonghua Liu; larkin-liu@163.com

Received 29 June 2018; Revised 24 October 2018; Accepted 12 November 2018; Published 2 January 2019

Guest Editor: Sarfaraz Ahmed Mahesar

Copyright © 2019 Ming-Zhi Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Tea is known to be one of the most popular beverages enjoyed by two-thirds of the world’s population [1]. In 2017, 5.68 million tons of tea were produced all over the world, in which 2.55 million tons were produced in China. The tea quality is influenced by various factors, such as cultivars, picking standard, tea processing technology, storage condition, and time. Concern of variability in tea quality is increasing among consumers. It is of great significance to control quality for commercialized tea products [2].

Tea quality is determined by its major active components, including polyphenols, caffeine, and free amino acids. These compounds not only endow tea with unique qualities of color, aroma, and taste but also contribute various health benefits for the human body [3]. Tea polyphenols account for 18–36% of dry weight in tea leaves, and the astringent and bitter taste of tea is mainly contributed by tea polyphenols. Tea catechins (flavan-3-ols) are the major ingredients in tea polyphenols. Tea catechins include (−)-epigallocatechin gallate (EGCG), (−)-epicatechin gallate (ECG), (−)-epigallocatechin (EGC), (−)-epicatechin (EC), (−)-gallocatechin gallate (GCG), (−)-gallocatechin (GC), and (+)-catechin (C), among which EGCG is the most abundant component [4]. The consumption of EGCG has been proved to have therapeutic effects for multiple diseases, such as cancer, metabolic syndrome, obesity, and cardiovascular and neurodegenerative diseases [5, 6]. The anticancer

1. Introduction

Tea is known to be one of the most popular beverages enjoyed by two-thirds of the world’s population [1]. In 2017, 5.68 million tons of tea were produced all over the world, in which 2.55 million tons were produced in China. The tea quality is influenced by various factors, such as cultivars, picking standard, tea processing technology, storage condition, and time. Concern of variability in tea quality is increasing among consumers. It is of great significance to control quality for commercialized tea products [2].

The tea quality is determined by its major active components, including polyphenols, caffeine, and free amino acids. These compounds not only endow tea with unique qualities of color, aroma, and taste but also contribute various health benefits for the human body [3]. Tea polyphenols account for 18–36% of dry weight in tea leaves, and the astringent and bitter taste of tea is mainly contributed by tea polyphenols. Tea catechins (flavan-3-ols) are the major ingredients in tea polyphenols. Tea catechins include (−)-epigallocatechin gallate (EGCG), (−)-epicatechin gallate (ECG), (−)-epigallocatechin (EGC), (−)-epicatechin (EC), (−)-gallocatechin gallate (GCG), (−)-gallocatechin (GC), and (+)-catechin (C), among which EGCG is the most abundant component [4]. The consumption of EGCG has been proved to have therapeutic effects for multiple diseases, such as cancer, metabolic syndrome, obesity, and cardiovascular and neurodegenerative diseases [5, 6]. The anticancer
2 Journal of Spectroscopy

property of EGCG appears to involve the suppression of angiogenesis, induction of apoptosis, altered expression of cell-cycle regulatory proteins, and activation of killer caspases [7, 8]. The suppression of angiogenesis by EGCG is associated with the change in various miRNA expressions, the inhibition of the VEGF (vascular endothelial growth factor) family, etc. [9]. The beneficial health effects of EGCG are presumed to be related with its antioxidative property. Another possible mechanism is through the direct binding of EGCG to target proteins, leading to the regulation of signal transduction pathways, transcription factors, DNA methylation, mitochondrial function, and autophagy [10]. Caffeine is another major functional component in tea and provides the bitter taste for tea [11, 12]. Caffeine has the therapeutic effects for various diseases, including metabolic syndrome, type 2 diabetes, liver diseases, and cardiovascular and cerebrovascular diseases [13, 14]. Free amino acids provide umami taste for the tea infusion. Among free amino acids, theanine accounts for approximately 50% of the total free amino acids in tea leaves [15]. Theanine not only offers a brisk flavor and an attractive aroma but also alleviates the astringency and bitterness caused by polyphenols and caffeine. Several studies have proved that theanine has significant health and cognitive benefits by influencing stress levels and learning efficiency [16].

Besides the chemical components, tea quality is influenced by various factors, including the sensory attributes, classification, and geographical origins [17–19]. Multiple analytical approaches have been used for the quality control, such as colorimetric measurements, high-performance liquid chromatography (HPLC), high-performance liquid chromatography coupled with mass spectrometry (HPLC-MS), gas chromatography (GC), and gas chromatography coupled with mass spectrometry (GC-MS) [20–30]. However, these methods not only are expensive, time-consuming, and destructive but also need specialists for the operation and cannot be applied for online applications. Therefore, near-infrared reflectance (NIR) spectroscopy, a rapid, noninvasive, and nondestructive instrumental technique with simplicity in sample preparation, has been developed and applied for the quality control of tea in recent years [19]. NIR spectroscopy is a spectroscopic method using the near-infrared region of the electromagnetic spectrum ranging from 750 nm to 2500 nm (14,300–4000 cm⁻¹). An NIR spectrometer is usually composed of a light source, a monochromator, a sample presentation interface, and a detector. The NIR radiation can be absorbed, transmitted, or reflected after interaction with samples. The feedback of spectral stretching and bending of the chemical bonds (O–H, N–H, and C–H) can be captured by utilizing different measurement modes of NIR equipment. Therefore, the specific absorption of organic compounds in the NIR region can represent their chemical composition [31, 32]. Anharmonicity and Fermi resonance determine the occurrence and spectral properties, such as the frequency and intensity of NIR absorption bands. However, NIR absorption bands are typically broad and overlapping, which severely restricts the sensitivity in the classical spectroscopic sense. The original spectral data of NIR spectroscopy usually require pattern recognition methods for accurate analysis by removing the disturbance of the noise, variability, uncertainties, and unrecognized features. Nevertheless, NIR spectroscopy is characterized by high penetration depth. This property allows direct analysis of strongly absorbing or even highly scattering samples, such as turbid liquids or solids, without further pretreatments [32].

Generally, the whole procedures of NIR spectroscopy include spectral data acquisition, data preprocessing, spectral data preprocessing, calibration models building with a set of samples, and models validating using a set of independent samples [33]. A typical NIR spectrum of tea is shown in Figure 1 [34]. The preprocessing of spectral data should be used for eliminating the noise and baseline shift from the background and instrument [33]. The commonly used preprocessing methods in tea analysis include standard normal variate (SNV), multiplicative scatter correction (MSC), and Savitzky–Golay (SG) smoothing [35, 36]. Various variables selection methods, such as synergy interval partial least squares (Si-PLS) and successive projections algorithm (SPA), are used for the screen of useful variables [37]. Multiple unsupervised and supervised pattern recognition methods have been used for the qualitative analysis (the discrimination of tea categories, varieties, and geographical origins) and quantitative analysis (the determination of chemical components in tea and optimization of processing conditions). These pattern recognition methods include principal component analysis (PCA), artificial neural network (ANN), linear discriminant analysis (LDA), support vector machine (SVM), soft independent modeling of class analogy (SIMCA), partial least squares (PLS), and backpropagation artificial neural network (BP-ANN) (Table 1) [33].

In this article, we review the most recent advances and applications of NIR spectroscopy and chemometrics for the quality control of tea, including the measurement of chemical compositions, the evaluation of sensory attributes, the identification of categories and varieties, and the discrimination of geographical origins.

2. The Application of NIR Spectroscopy in Tea

2.1. Chemical Composition. The major compositions in tea include polyphenols, catechins, caffeine, free amino acids, and moisture. These compositions are closely relevant to the overall quality of tea, and they thus are the key indexes of tea quality. The monitoring of these compositions contents in tea is critical for the quality control [54]. NIR spectroscopy has been successfully used for the prediction of major compositions contents in tea in recent years. Nonetheless, only one or several components were simultaneously measured by NIR spectroscopy in previous studies [55, 56]. Lee et al. firstly determined the contents of nine individual catechins and caffeine by NIR spectroscopy. These nine catechins include EGCG, (−)-epigallocatechin-3-(3'-O-methyl) gallate (EGCG-3Me), EGC, ECG, EC, C, GCG, GC, and gallic acid. The calibration models for EGCG, EGC, ECG, EC, C, total catechins, and caffeine exhibited accurate prediction, with high $r^2$ (coefficient of determination in the
Figure 1: A typical absorbance (log(1/R)) spectrum of tea in the wavenumbers from 12,000 to 3800 cm$^{-1}$ [34].

prediction set, >0.9) and RSP (the ratio of standard deviation of reference data to SEP (C) in the external validation set, >4.1) values [38]. In addition, Zareef et al. used Fourier transform near-infrared reflectance (FT-NIR) spectroscopy for the simultaneous prediction of four compositions in black tea including amino acids, caffeine, theaflavins, and water extract. For quantitative analysis of these components, four kinds of chemometrics algorithms including PLS, Si-PLS, genetic algorithm PLS (GA-PLS), and backward interval PLS (Bi-PLS) were used for the establishment of prediction models. The results showed that GA-PLS was suitable for the quantitative analysis of amino acids and water extract and Bi-PLS was the best method for the quantification of caffeine and theaflavins (TFs) [39].

Tea usually can be divided into six categories in China, and more than 100 famous tea varieties or brands exist in China. However, a common model is lacking for simultaneously evaluating various quality parameters of various teas. Recently, Wang et al. developed a common cross-category FT-NIR model for simultaneous determination of polyphenols, caffeine, and free amino acids in various Chinese teas, including green tea, black tea, oolong tea, and dark tea. Baseline offsets, random noise, and biases were removed, and characteristic signals were enhanced by a hybrid method, which combines MSC and first-order derivative and SG. Two variable selection methods, random frog (RF) and competitive adaptive reweighted sampling (CARS), were used for selecting key variables for PLS calculation. Both enhanced RF-PLS and CARS-PLS models simplified the model complexity, enhanced the model performance, and gave satisfactory prediction precision. NIR coupled with enhanced cross-category models thus has the potential for the simultaneous prediction of the major ingredients in various Chinese teas [35].

The determination of total polyphenols in tea by using NIR spectroscopy has been studied in detail. However, most of these research studies were performed in research laboratories. Furthermore, these research studies mainly used commercial NIR instruments, which are nonspecific, expensive, and sensitive to environmental variation, and they are thus not suitable for online detection in tea industrial usage. The new trend of tea quality monitoring is to supervise the whole production line so as to ensure the high quality and consistency of tea products [40, 41]. Qi et al. developed a portable and low-cost optical visible and near-infrared reflectance (VIS-NIR) spectroscopy system, including a light source, a backscattering fiber probe, a grating system equipped with a slit, a detector, and a computer supported with data acquisition and control software. The genetic algorithm-synergy interval partial least squares (GA-Si-PLS) algorithm was used for monitoring the total polyphenols content in tea, and coefficients of variation (CVs) were <5% for most of the samples. This optical sensors system thus possessed great potential for the real-time and online monitoring of tea quality in processing enterprises [40]. In addition, summer-autumn tea leaves are the raw material of instant black tea products. The oxidation of the summer-autumn tea extract is a critical treatment for the production of instant black tea products, and the total polyphenols content is the key index for the oxidation degree. Pan et al. developed an in situ monitoring installation, including an oxidation system of the tea extract and VIS-NIR spectroscopy system, to monitor the total polyphenols content during tea oxidation. The ACO-PLS (ant colony optimization-partial least squares) algorithm was extremely suitable for the modeling of this monitoring installation, and CVs for most of the samples were less than 10%. This monitoring installation thus was a promising tool for in situ monitoring of tea oxidation [41].

NIR hyperspectral imaging has also been used for the prediction of chemical compositions in tea. Compared with NIR spectroscopy, NIR hyperspectral imaging could simultaneously obtain spectral and spatial information by the integration of spectroscopy and digital imaging. The componental and constructional characteristics of a sample could be acquired by the spectrum for each pixel and the gray scale image for each narrow band. Texture information is another significant image feature. It is more similar to human visual perception, which facilitates the direct identification of complex features in the sample [42, 57]. Deng et al. predicted the moisture content in Longjing tea leaves with NIR hyperspectral imaging. The property of continuous texture near the veins was validated according to the variable rates of water loss in the mesophylls and vein cells. Then, the three-dimensional Gabor filter (TDGF) algorithm and its corresponding filter bank were used for describing the textures of tea leaves. The overall metrics showed that the combination of spectrum and TDGF textures facilitated PLS regression modeling to predict the moisture content of Longjing tea [42].

2.2. Sensory Attributes. Sensory attributes of tea include color, taste, aroma, and appearance, which are the key factors of tea quality as well as indicators of commercial values. Traditional methods for evaluating the sensory attributes rely basically on experienced panels, also known as tea tasters. However, the results of traditional sensory evaluation are purely subjective, which are easily affected by experience, gender, mental state, physical condition, and other factors. Therefore, objective methods are the principal concern for the evaluation of tea sensory attributes. The NIR spectroscopy is an ideal solution for the rapid, accurate, and
| Commodity                  | Attributes                                                                 | Methods          | Wavelength scanned | Spectral pretreatment | Calibration models | Results No. of samples | References |
|---------------------------|---------------------------------------------------------------------------|------------------|--------------------|------------------------|--------------------|------------------------|------------|
| Tea leaves                | Caffeine, catechin (gallic acid, GC, EGC, C, EGCG, EC, GCG, ECG)          | NIR              | 400–2500 nm        | Win ISI Score          | MPLS               | r² for caffeine: 0.97; GA: 0.85; GC: 0.78; EGC: 0.95; C: 0.91; EGCG: 0.97; EC: 0.95; GCG: 0.85; ECG: 0.94 | 665 [38]  |
| Black tea                | Amino acids, caffeine, theaflavins, water extract                         | FT-NIR           | 4000–10,000 cm⁻¹   | SNV, MSC               | PLS, Si-PLS, GA-PLS, Bi-PLS | Using GA-PLS, R_p for amino acids: 0.9498; water extract: 0.8785; using Bi-PLS, R_p for caffeine: 0.9232; theaflavins: 0.924 CARS-PLS (R_p) achieved best predictive performance for total polyphenols: 0.994; caffeine: 0.986; free amino acids: 0.993 | 95 [39]  |
| Black, dark, oolong, and green tea | Total polyphenols, caffeine, free amino acids                           | FT-NIR           | 4000–10,000 cm⁻¹   | MSC combined with first-order derivative and SG smoothing | PLS1, PLS2, RF-PLS, CARS-PLS | CARS-PLS (R_p) achieved best predictive performance for total polyphenols: 0.994; caffeine: 0.986; free amino acids: 0.993 | 145 [35]  |
| Green tea                | Total polyphenols                                                        | VIS-NIR          | 300–1000 nm        | SNV                    | PLS, Si-PLS, CARS-Si-PLS, GA-Si-PLS | Prediction set (R_p) for PLS: 0.8043; Si-PLS: 0.8804; GA-Si-PLS: 0.8859; CARS-Si-PLS: 0.8753 | 50 [40]  |
| Tea extract              | Total polyphenols                                                        | VIS-NIR          | 300–1000 nm        | SNV                    | PLS, Si-PLS, GA-PLS, CARS-PLS, ACO-PLS | Prediction set (RMSEP) for PLS: 0.7699; Si-PLS: 0.8766; GA-PLS: 0.8993; CARS-PLS: 0.8897; ACO-PLS: 0.8853 | 85 [41]  |
| Longjing tea leaves      | Moisture content                                                         | Hyperspectral imaging | 874.41–1733.91 nm | Smoothing filter (3 * 3 window) MNF rotation, 2D filter LoG (Laplacian of Gaussian) | PCPLS1-9, SPA-PLS | r² for PCPLS1-9: 0.9491, 0.8826, 0.9531, 0.8905, 0.9548, 0.9105, 0.9713, 0.9071, and 0.9610; SPA-PLS: 0.9216 | 30 [42]  |
| "Biluochun" green tea    | Sensory attributes                                                       | FT-NIR           | 4000–10,000 cm⁻¹   | SNV                    | Si-PLS, PCA, BPNN, BP-AdaBoost | BP-AdaBoost model revealed its superior performance, R_p = 0.7717, R_p = 0.8935 | 70 [37]  |
| Black tea                | Color sensory quality                                                    | VIS-NIR          | 200–1100 nm        | SNV                    | GA-BPANN PLS, SI-PLS, SI-CARS-PLS, SI-CARS-ELM, SI-CARS-SVM, SI-CARS-ELM-AdaBoost | ELM-AdaBoost was used for the validation, R_p² = 0.893 | 127 [18] |
| Black tea                | Theaflavin, thearubigin                                                   | NIR              | 1000–1799 nm       | MSC, SG 1st derivative, Min/Max, SNVT | | | 78 [43] |
| Commodity (Anji-white) | Attributes | Methods | Wavelength scanned | Spectral pretreatment | Calibration models | Results | No. of samples | References |
|-----------------------|------------|---------|--------------------|----------------------|--------------------|---------|---------------|------------|
| Green tea | Lutein, Chl-b, Chl-a, Phe-b, Phe-a, β-carotene | VIS-NIR | 400–2498 nm | ANOVA | PLS, SPA, MLR | MLR gave superior prediction ($R^2$) for lutein: 0.975; Chl-b: 0.973; Chl-a: 0.993; Phe-b: 0.919; Phe-a: 0.962; β-carotene: 0.965 | 135 | [44] |
| White tea and albino tea | Tea polyphenols, free amino acids, moisture, ash contents | FT-NIR | 4000–12,400 cm$^{-1}$ | MSC, SNV, SG smoothing, KND, 1st and 2nd derivatives | DPLS, DA | DPLS: 98.48; DA: 100 | 70 | [45] |
| Green, yellow, white, black, and oolong tea | Region of interest | VIS-NIR | 589, 635, 670, 783 nm | SNV | LDA, Lib-SVM, ELM | Lib-SVM was the best model, $r^2 = 98.39\%$ | 206 | [46] |
| Iron Buddha tea | Total polyphenols | VIS-NIR | 800–2500 nm | SNV | PLS (LS-SVM and BPNN) | SVM-ECOC model provided a classification accuracy of 97.41 ± 0.16% | 6 | [19] |
| Pu-erh tea | Metabolomics analysis | NIR | 3600–12,500 cm$^{-1}$ | OPUS 7.2 software from Bruker Optics | PCA, PLS, HCA, PLS-DA | Classification accuracies: LS-SVM: 95.0%; BPNN: 97.5% | 180 | [47] |
| Green tea (Anji-white) | | NIR | 4000–12,000 cm$^{-1}$ | Smoothing, 2nd derivative, SNV | OCPLS, SIMCA | With SNV preprocessing, OCPLS provided sensitivity of 0.886 and specificity of 0.951; SIMCA provided sensitivity of 0.886 and specificity of 0.938 and achieved best classification performance | 248 | [36] |
| Green tea | Catechin, EC, EGC, ECG, EGCG, GCG | NIR | 1050–2500 nm | 1st derivative | PLS, BP-ANN, SVM | Accuracy (%): PLS: 100.000; BP-ANN: 95.455; SVM: 98.485 | 220 | [49] |
| Green tea | | NIR | 4000–9000 cm$^{-1}$ | 2nd derivative, SNV | OVR-PLSDA, OVO-PLSDA, PLSDA-softmax, ES-PLSDA | Total accuracy (%): OVR-PLSDA: 64.68; OVO-PLSDA: 84.94; PLSDA-softmax: 92.99; ES-PLSDA: 93.77 | 1540 | [50] |
Table 1: Continued.

| Commodity       | Attributes                          | Methods          | Wavelength scanned          | Spectral pretreatment          | Calibration models | Results                                                                                         | No. of samples | References |
|-----------------|-------------------------------------|------------------|-----------------------------|--------------------------------|-------------------|-------------------------------------------------------------------------------------------------|----------------|------------|
| Black tea       | Caffeine, water extract, total polyphenols, free amino acids | NIR              | 4000–12,500 cm⁻¹           | SNV, MSC, Min/Max             | PLS               | (1) \( R \) in the prediction set for caffeine: 0.955; water extracts: 0.962; total polyphenols: 0.954; free amino acids: 0.927 (2) Identification accuracy (%) 94.30 Classification accuracy (%): SIMCA: 88.00; PLSDA: 92.00; SPA-LDA: 100 The sensitivity of PLSDA model for raw data: 0.971; SNV: 1.000; 2nd derivative: 0.886; smoothing: 0.971 | 140            | [51]       |
| Green and black tea | —                                  | NIR              | 3800–14,000 cm⁻¹           | 1st derivative, SG smoothing  | SIMCA, PLSDA, SPA-LDA |                                                                                                 | 82             | [52]       |
| Oolong tea      | —                                   | NIR              | 4000–12,000 cm⁻¹           | SNV, 2nd derivative, smoothing| PLSDA             |                                                                                                 | 570            | [53]       |
| Oolong tea      | Polyphenols, alkaloids, protein, volatile and nonvolatile acids, aroma compounds | NIR and NMR     | 3300–12,500 cm⁻¹           | SNV, 2nd derivative, smoothing| PCA, PLSDA        | Discrimination accuracy (%) for NMR + NIR data: 86.20–95.80; NMR data: 68.20–78.70; NIR data: 80.00–89.30 | 90             | [17]       |

Abbreviations: ACO, ant colony optimization; ANOVA, one-way analysis of variance; Bi-PLS, backward interval PLS; BP-ANN, backpropagation artificial neural network; BPNN, backpropagation neural network; C, (+)-catechin; CARS-PLS, competitive adaptive reweighted sampling-partial least squares; Chl-a, chlorophyll a; Chl-b, chlorophyll b; EC, (−)-epicatechin; ECG, (−)-epicatechin gallate; EGC, (−)-epigallocatechin; EGG, (−)-epigallocatechin gallate; ELM, extreme learning machine; ES, ensemble strategy; NIR, near-infrared reflectance; FT-NIR, Fourier transform near-infrared reflectance; GA, genetic algorithm; GC, (−)-gallocatechin; GCG, (−)-gallocatechin gallate; ISOMAP, isometric mapping; KND, Karl Norris derivative filter; LDA, linear discriminant analysis; PLS, partial least squares; PLSDA, partial least squares discriminant analysis; Lib-SVM, library support vector machine; MDS, multidimensional scaling; Min/Max, min/max normalization; MLR, multiple linear regression; MNN, minimal noise fraction; MPLS, modified partial least squares; MSC, multiplicative scattering correction; NMR, nuclear magnetic resonance; OCV, one-class partial least squares; OVO-PLS, one-versus-one partial least squares discriminant analysis; OVR-PLS, one-versus-rest partial least squares discriminant analysis; PCA, principal component analysis; Phe-a, pheophytin a; Phe-b, pheophytin b; \( Q^2 \), cross-validated correlation coefficient; \( R^2 \), coefficient of determination in the prediction set; \( R^2_p \), correlation coefficient in the prediction set; \( R^2_c \), determine coefficient; RF-PLS, random frog-partial least squares; SG smoothing, Savitzky–Golay smoothing; SIMCA, soft independent modeling of class analogy; Si-PLS, synergy interval partial least squares; SNV, standard normal variance; SVNT, standard normal variance transformation; SPA-LDA, successive projections algorithm associated with linear discriminant analysis; SVM, support vector machine; SVM-ECOC, error-correcting output code (ECOC) model containing support vector machine (SVM); t-SNE, t-distributed stochastic neighbor embedding; VIS-NIR, visible and near-infrared reflectance; —, not mentioned.
noninvasive sensory evaluation of tea [18]. Jiang and Chen used FT-NIR spectroscopy for predicting the sensory properties of green tea infusion. The Si-PLS algorithm was applied for the selection of significant spectral regions, and the modified BP-AdaBoost algorithm was used for calibrating the models. The BP-AdaBoost algorithm showed its superiority in modeling, with the $R_p$ (the correlation coefficient in the prediction set) of 0.7717, RPD (the ratio performance deviation in the prediction set) of 1.59, $R_c$ (the correlation coefficient in the calibration set) of 0.8554, and RMSECV (the root mean square error of cross-validation) of 5.0305. Thus, the FT-NIR spectroscopy technique proved to be a rapid, accurate, and noninvasive analytical method for the evaluation of sensory quality in green tea. Nonetheless, tea sensory properties should be individually characterized by NlR spectroscopy [37].

Qín et al. investigated the feasibility for predicting the color sensory attribute in black tea by using VIS-NIR spectroscopy. The spectra information and color information were acquired for the modeling. Spectra information-based models obtained better performance than color parameters-based models. The excellent performance for predicting the color sensory quality was acquired by genetic algorithm-backpropagation artificial neural network (GA-BPANN) models, with the $R$ of 0.8935 and the root mean square error of 0.392 in the prediction set [18]. Furthermore, TFs and thearubigins (TRs) are the major pigments that determine the color and brightness of black tea infusion. During the fermentation process, the color of black tea leaves changes remarkably from green to red and then to brown. When the TRs/TFs ratio is approximately equal to 10:1, the fermentation process of black tea reaches the optimum point, and the most beautiful color was produced in tea infusion. The TRs/TFs ratio thus is a critical parameter for evaluating the fermentation degree and sensory quality characteristics of black tea. Dong et al. used NlR spectroscopy for the prediction of the TRs/TFs ratio value during the Congou black tea fermentation process. The combination of Si-PLS and CARS could effectively select the characteristic wavelength variables related to the TRs/TFs ratio, with a variable compression ratio up to 98.6%. Based on these characteristic variables, an extreme learning machine (ELM) combined with an adaptive boosting (AdaBoost) algorithm (ELM-AdaBoost) was used for constructing the prediction model. The prediction performance of the SI-CARS-ELM-AdaBoost model was higher than that of other nonlinear models including extreme learning machine (ELM), SVM, linear models, and full-spectrum PLS model. The rapid and accurate prediction of the TRs/TFs value was acquired during fermentation, with a determinate coefficient ($R^2$) of 0.893, relative standard deviation (RSD) below 10%, RPD above 3, and root mean square error of prediction (RMSEP) of 0.0044 [43]. Similarly, Li et al. found that color of green tea had close correlations with the contents of six lipid-soluble pigments, including chlorophyll a, chlorophyll b, lutein, β-carotene, pheophytin a, and pheophytin b. VIS-NIR spectroscopy was used for rapid and simultaneous determination of six lipid-soluble pigments in green tea. Based on multiple linear regression (MLR) with the characteristic wavelengths, the quantitative models of the six pigments showed excellent performance, with $R^2$ of 0.975, 0.973, 0.993, 0.919, 0.962, and 0.965, respectively [44]. The color sensory quality of tea thus could be evaluated or controlled by the rapid determination of pigments with NlR spectroscopy [43, 44].

By conducting Pearson’s correlation analysis between chemical components and taste score, Chen et al. found that eight ingredients (water extracts, total polyphenols, total catechins, caffeine, free amino acids, TFs, theaflavin-3-gallate, and theaflavin-3′-gallate) in the black tea were the main contributors to the taste quality, while gallic acid, EGCG, EC, and theaflavin-3,3′-digallate had weak correlations with taste quality. Then, the FT-NIR spectroscopy system coupled with the backpropagation-AdaBoost (BP-AdaBoost) algorithm was used for simultaneous prediction of taste quality and these eight taste-related compounds content in black tea. BP-AdaBoost models showed superior predictions for taste quality and taste-related compounds content in black tea, with the $R_p > 0.76$, and the RMSEP <1.7% for all models [45].

2.3. Classification and Authentication. Tea usually can be divided into six categories, including green tea (unfermented), white tea (partly fermented), oolong tea (semifermented), black tea (fully fermented), and dark tea (postfermented). Rapid and feasible classification of two or three tea categories has been achieved by NlR spectroscopy [58, 59]. Recently, NlR hyperspectral imaging has also been used for the tea classification. Ning et al. used VIS-NIR hyperspectral imaging for the classification of five Chinese tea categories, including green, black, oolong, yellow, and white teas. Hyperspectral data were extracted within the range of 400–1000 nm wavelength from a total of 206 tea samples. Four dominant wavelengths (589, 635, 670, and 783 nm) were selected as spectral features, and textural features were extracted by the gray-level cooccurrence matrix (GLCM) at these four dominant wavelengths. The classification models of library support vector machine (Lib-SVM), LDA, and ELM were constructed based on spectral features, full spectra, textural features, and data fusion. The model of Lib-SVM based on data fusion or full spectra was the best model, with the correct classification rate of 98.39% [46]. Nonetheless, the applications of VIS-NIR hyperspectral imaging described above only cover five types of teas. Furthermore, tea characteristics measured by VIS-NIR spectral imaging (400–1000 nm) dominated by physical characteristics and the pigments. Compared with VIS-NIR spectral imaging, the NIR spectral imaging provides more detailed chemical information, which offers a better classification system. NIR hyperspectral imaging (950–1760 nm) has also been used for the classification of six different commercial tea products, including green, black, oolong, yellow, white, and pu–erh teas. Before data modeling, the NIR imaging data should be preprocessed to reduce the disturbances of light scattering caused by the uneven and inhomogeneous leaf surface. By
using the data visualization method of t-distributed stochastic neighbor embedding (t-SNE), the six commercial tea products could be effectively divided into three categories based on the extent of processing: minimal processing, oxidation, and fermentation. A multiclass error-correcting output code (ECOC) model containing SVM binary learners was further constructed for the tea classification according to the product type. The ECOC-SVM model provided excellent classification accuracy up to 97.41% for the six commercial tea products [19].

The NIR technology has also been used for the authentication of tea storage periods. Xiong et al. used the VIS-NIR imaging system (405–970 nm) to classify the Iron Buddha tea based on the storage period (years of 2004, 2007, 2011, 2012, and 2013). The classification accuracies of 97.5% and 95.0% were acquired by using backpropagation neural network (BPNN) and least squares-support vector machine (LS-SVM) models [47]. Similarly, Wang et al. used NIR spectroscopy for the storage period classification of pu-erh raw tea, which has been stored for 1–10 years. Obvious difference between new and aged pu-erh raw teas was found, and 85% of the samples could be identified without any false-positive result. However, the remaining 15% samples could not be successfully clustered into the right year of production, and they also could not be clustered into the wrong year either [48].

2.4. Geographical Origins. The tea qualities of different geographical origins are somewhat jagged, due to the disparity of geographical and natural conditions (altitude, climate, soil, microelement, etc.), tea cultivars, cultivation traditions, and processing procedures. The same kind of tea from different geographical origins might vary dramatically in prices and quality [17]. Therefore, almost all of the famous teas are labeled with their origins, such as Anji-white tea, Anxi-Tieguanyin tea, and Yingde-black tea [36, 53]. However, some merchants fraudulently falsify the geographical origins of tea for illegal profits. It is urgent to enforce quality control against various counterfeits [53]. NIR spectroscopy has been successfully used for determining the geographical origin of various teas in recent years. Anji-white tea, one of the most famous green teas, has been documented as a protected geographical indication product in China. 167 representative Anji-white tea samples were gathered from the original producing areas, and non-Anji-white tea samples with similar appearances were collected from unprotected producing areas in China. NIR spectroscopy coupled with SIMCA or one-class partial least squares (OCPLS) was used for the geographical origin discrimination of these samples. Based on the SNV preprocessing, the sensitivity and specificity were 0.886 and 0.938 for SIMCA and 0.886 and 0.951 for OCPLS, respectively. Although it is hard to achieve the exhaustive analysis of all types of potential counterfeits, NIR spectrometry coupled with SNV-OCPLS and SNV-OCPLS models could rapidly detect most of the non-Anji-white teas in the Chinese market [36]. Besides, Zhuang et al. used the NIR spectroscopy to classify the green tea from two geographical origins. 100% identification accuracies in training and testing were acquired by the classification model of PLS [49]. Although NIR spectroscopy coupled with chemometrics algorithms has been used for the discrimination of tea geographical origins, the discrimination is usually limited to small scale [49, 60]. However, the class number of teas has increased significantly in recent years. For instance, Longjing tea, a top-quality green tea in China, has more than 20 geographical origins. A substantial difference in price exists among these geographical origins. More complex large-class-number classification would pose new challenges to the traditional pattern recognition, due to increasing data complexity and class overlapping, and degraded model generalization performance. Fu et al. proposed a novel ensemble strategy (ES) to solve the problem of large-class-number classification. ES combined the one-versus-one (OVO) and one-versus-rest (OVR) strategies to design a set of classifiers with reduced class numbers. The pattern recognition of ES, OVO, OVR, and softmax function was compared to discriminate the geographical origins of 25 Longjing tea samples by using NIR spectroscopy and partial least squares discriminant analysis (PLSDA). The highest total accuracy was acquired by ES-PLSDA with the value of 0.9377, while the total accuracies of OVO-PLSDA, OVR-PLSDA, and PLSDA-softmax were 0.8494, 0.6468, and 0.9299, respectively. ES pattern recognition thus achieved improved performance in large-class-number classification [50].

The geographical origins of black teas have been discriminated by NIR spectroscopy. Ren et al. constructed an NIR spectroscopy for rapidly determining the geographical origins of black tea. Different geographical origins including Anhui, Hubei, and Yunnan in China, India, Kenya, Sri Lanka, and Burma were remarkably recognized by a factorization method, with an accuracy rate of 94.3%. Meanwhile, the contents of major constituents in black tea including water extracts, caffeine, total polyphenols, and free amino acids were predicted well by the PLS algorithm, with the correlation coefficient (R) values of 0.962, 0.955, 0.954, and 0.927, respectively, in the calibration set [51]. Furthermore, Diniz et al. used NIR spectroscopy for simultaneous classification of tea samples according to their geographical origins (Brazil, Argentina, or Sri Lanka) and varieties (green or black). The successive projections algorithm associated with the linear discriminant analysis (SPALDA) was used for the variable selection, and its recognition accuracy was compared with that of SIMCA and partial least squares-discriminant analysis (PLSDA). Argentinean green tea, Brazilian green tea, Argentinean black tea, Brazilian black tea, and Sri Lankan black tea were successfully discriminated by the SPA-LDA model with 100% classification accuracy, while SIMCA and PLSDA models were not able to achieve 100% classification accuracy. Although simultaneous classification of teas according to their geographical origins and varieties was successfully realized by the SPA-LDA model, a larger testing of tea samples must be implemented to guarantee any generalization of the proposed methodology [52].

Tieguanyin tea is one of the most famous oolong teas. It is a protected geographical indication product in China. The
geographical origin of Tieguanyin tea is restricted to Anxi County, a small town in Fujian Province of China. 450 representative samples of Tieguanyin tea were collected from Anxi County, which is the original production area of Tieguanyin tea. Another 120 counterfeits with a similar appearance were gathered from nonprotective areas in China. NIR spectroscopy coupled with PLSDA was used for the geographical origin discrimination of these samples. The sensitivity and specificity of the PLSDA model based on SNV transformation reached 0.93 and 1.00, respectively. NIR spectrometry combined with the SNV-PLSDA model thus could discriminate the geographical origins of Tieguanyin tea rapidly [53]. Recently, the combinational analysis of NIR spectroscopy and proton nuclear magnetic resonance (\(^1\)H NMR) has been used for distinguishing 90 Tieguanyin tea samples, which were collected from three different growing places (Xiandu, Xianghua, and Xiping towns) in the Fujian Province of China. \(^1\)H NMR spectroscopy could offer the structure and content information of compounds in samples, which is complementary to the NMR data [61, 62]. The \(^1\)H NMR spectroscopy provided accurately qualitative information of 26 components (polyphenols, amino acids, and saccharides) in Tieguanyin tea. Compared with NIR (80.0–89.3% of accuracy) or NMR (68.2–78.7% of accuracy) analysis alone, a better discrimination accuracy of geographical origins of oolong tea could be achieved by combining the NIR and NMR data (86.2–95.8% of accuracy). The combination of NIR and NMR approaches could be used as an effective way to identify the geographical origin of tea. More Tieguanyin tea collected from more original producing areas or even different tea varieties could be included to validate the effectiveness of this combined method in the future works [17].

3. Conclusion and Prospects

As a rapid, nondestructive, and inexpensive technique, NIR spectroscopy has been extensively applied for analyzing multiple aspects of tea quality control in recent years, such as chemical compositions, sensory attributes, classification, authentication, and geographical origins. It is anticipated that NIR spectroscopy may progressively become a routine method for the tea quality control and expand to the food safety field of tea [63]. However, some challenges still impede the pervasive application of NIR spectroscopy for the quality control of tea. Although the performance of the NIR spectrometer has been significantly improved in recent years by increasing the sensitivity and reducing the background noise, improving accuracy and ensuring stability of the NIR spectrometer are still required. Innovative calibrations and prediction models with higher accuracy should be developed. More robust calibrations should be constructed for the simultaneous analysis of various teas and multiple quality attributes by using larger sample sets. Moreover, it is difficult for beginners and nonresearchers to select and optimize the appropriate algorithms and models. Intelligent software packs, which could select the optimal algorithms and models automatically from various algorithms and models, should be developed for the more widespread commercial application of NIR spectroscopy. In addition, NIR spectroscopy offers the exciting prospect potentially for real-time and online monitoring of the whole progress of tea production. The whole monitoring of tea production by NIR spectroscopy could objectively measure the chemical compositions and sensory attributes, detect unwanted problems immediately, and assure the quality of the final products.

Disclosure

Ming-Zhi Zhu and Beibei Wen are the co-first authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors’ Contributions

Ming-Zhi Zhu and Beibei Wen contributed equally to this work.

Acknowledgments

This work was supported by China Postdoctoral Science Foundation (2018M632962).

References

[1] E. H. Xia, H. B. Zhang, J. Sheng et al., “The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis,” Molecular Plant, vol. 10, no. 6, pp. 866–877, 2017.
[2] K. W. Ng, Z. J. Cao, H. B. Chen, Z.-Z. Zhao, L. Zhu, and T. Yi, “Oolong tea: a critical review of processing methods, chemical composition, health effects, and risk,” Critical Reviews in Food Science and Nutrition, vol. 5, pp. 1–24, 2017.
[3] M. Z. Zhu, N. Li, M. Zhao, W. Yu, and J. L. Wu, “Metabolomic profiling delineate taste qualities of tea leaf pubescence,” Food Research International, vol. 94, pp. 36–44, 2017.
[4] A. Chowdhury, J. Sarkar, T. Chakraborti, P. K. Pramanik, and S. Chakraborti, “Protective role of epigallocatechin-3-gallate in health and disease: a perspective,” Biomedicine & Pharmacotherapy, vol. 78, pp. 50–59, 2016.
[5] M. P. Kapoor, M. Sugita, Y. Fukuzawa, and T. Okubo, “Physiological effects of epigallocatechin-3-gallate (EGCG) on energy expenditure for prospective fat oxidation in humans: a systematic review and meta-analysis,” Journal of Nutritional Biochemistry, vol. 43, pp. 1–10, 2017.
[6] R. Y. Gan, H. B. Li, Z. Q. Sui, and H. Corke, “Absorption, metabolism, anti-cancer effect and molecular targets of epigallocatechin gallate (EGCG): an updated review,” Crit Rev Food Sci Nutr, vol. 58, no. 6, pp. 924–941, 2018.
[7] M. S. Butt, R. S. Ahmad, M. T. Sultan, M. M. N. Qayyum, and A. Naz, “Green tea and anticancer perspectives: updates from last decade,” Critical Reviews in Food Science and Nutrition, vol. 55, no. 6, pp. 792–805, 2015.
[8] C. S. Yang and J. G. Hong, “Prevention of chronic diseases by tea: possible mechanisms and human relevance,” Annual Review of Nutrition, vol. 33, pp. 161–181, 2013.
[9] B. Rashidi, M. Malekzadeh, M. Goodarzi, A. Masoudifar, and H. Mirzaei, “Green tea and its anti-angiogenesis effects,” Biomedicine and Pharmacotherapy, vol. 89, pp. 949–956, 2017.
FT-NIR spectroscopy coupled chemometrics algorithms," *Analytical Methods*, vol. 10, no. 25, pp. 3023–3031, 2018.

[40] S. Qi, Q. Ouyang, Q. S. Chen, and J. Zhao, "Real-time monitoring of total polyphenols content in tea using a developed optical sensors system," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 97, pp. 116–122, 2014.

[41] W. X. Pan, J. W. Zhao, Q. S. Chen, and L. Yuan, "In situ monitoring of total polyphenols content during tea extract oxidation using a portable spectroscopy system with variables selection algorithms," *RSC Advances*, vol. 5, no. 75, pp. 60876–60883, 2015.

[42] S. G. Deng, Y. F. Xu, X. L. Li, and Y. He, "Moisture content prediction in tealeaf with near infrared hyperspectral imaging," *Computers and Electronics in Agriculture*, vol. 118, pp. 38–46, 2015.

[43] C. W. Dong, J. Li, J. J. Wang et al., "Rapid determination by near infrared spectroscopy of theaflavins-to-thearubigins ratio during Congou black tea fermentation process," *Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy*, vol. 205, pp. 227–234, 2018.

[44] X. L. Li, J. J. Jin, C. J. Sun, D. Ye, and Y. Liu, "Simultaneous determination of six main types of lipid-soluble pigments in green tea by visible and near-infrared spectroscopy," *Food Chemistry*, vol. 270, pp. 236–242, 2019.

[45] Q. S. Chen, M. Chen, Y. Liu et al., "Application of FT-NIR spectroscopy for simultaneous estimation of taste quality and taste-related compounds content of black tea," *Journal of Food Science and Technology-Mysore*, vol. 55, no. 10, pp. 4363–4368, 2018.

[46] J. M. Ning, J. J. Sun, S. H. Li, M. Sheng, and Z. Zhang, "Classification of five Chinese tea categories with different fermentation degrees using visible and near-infrared hyperspectral imaging," *International Journal of Food Properties*, vol. 20, no. 2, pp. 1515–1522, 2017.

[47] C. W. Xiong, C. H. Liu, W. J. Pan et al., "Non-destructive determination of total polyphenols content and classification of storage periods of Iron Buddha tea using multispectral imaging system," *Food Chemistry*, vol. 176, pp. 130–136, 2015.

[48] T. Wang, X. L. Li, H. C. Yang et al., "Mass spectrometry-based metabolomics and chemometric analysis of Pu-erh teas of various origins," *Food Chemistry*, vol. 268, pp. 271–278, 2018.

[49] X. G. Zhuang, L. L. Wang, Q. Chen, X. Wu, and J. Fang, "Identification of green tea origins by near-infrared (NIR) spectroscopy and different regression tools," *Science China-Technological Sciences*, vol. 60, no. 1, pp. 84–90, 2017.

[50] H. Y. Fu, Q. B. Yin, L. Xu et al., "Challenges of large-class-number classification (LCNC): a novel ensemble strategy (ES) and its application to discriminating the geographical origins of 25 green teas," *Chemometrics and Intelligent Laboratory Systems*, vol. 157, pp. 43–49, 2016.

[51] G. X. Ren, S. P. Wang, J. M. Ning et al., "Quantitative analysis and geographical traceability of black tea using Fourier transform near-infrared spectroscopy (FT-NIRS)," *Food Research International*, vol. 53, no. 2, pp. 822–826, 2013.

[52] P. H. G. D. Diniz, A. A. Gomes, M. F. Pistonesi, B. S. F. Band, and M. C. U. de Araújo, "Simultaneous classification of teas according to their varieties and geographical origins by using NIR spectroscopy and SPA-LDA," *Food Analytical Methods*, vol. 7, no. 8, pp. 1712–1718, 2014.

[53] S. M. Yan, J. P. Liu, L. Xu et al., "Rapid discrimination of the geographical origins of an Oolong tea (Anxi-Tieguanyin) by near-infrared spectroscopy and partial least squares discriminant analysis," *Journal of Analytical Methods in Chemistry*, vol. 2014, Article ID 704971, 6 pages, 2014.

[54] B. N. Singh, A. K. S. Rawat, R. M. Bhagat, and B. R. Singh, "Black tea: phytochemicals, cancer chemoprevention, and clinical studies," *Critical Reviews in Food Science and Nutrition*, vol. 57, no. 7, pp. 1394–1410, 2017.

[55] Z. M. Guo, Q. S. Chen, L. P. Chen, W. Huang, C. Zhang, and C. Zhao, "Optimization of informative spectral variables for the quantification of EGCG in green tea using fourier transform near-infrared (FT-NIR) spectroscopy and multi-variate calibration," *Applied Spectroscopy*, vol. 65, no. 9, pp. 1062–1067, 2011.

[56] V. R. Sinija and H. N. Mishra, "FT-NIR spectroscopy for caffeine estimation in instant green tea powder and granules," *LWT-Food Science and Technology*, vol. 42, no. 5, pp. 998–1002, 2009.

[57] F. Zhu, D. R. Zhang, Y. He, F. Liu, and D.-W. Sun, "Application of visible and near infrared hyperspectral imaging to differentiate between fresh and frozen–thawed fish fillets," *Food and Bioprocess Technology*, vol. 6, no. 10, pp. 2931–2937, 2013.

[58] J. W. Zhao, Q. S. Chen, X. Y. Huang, and C. H. Fang, "Qualitative identification of tea categories by near infrared spectroscopy and support vector machine," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 41, no. 4, pp. 1198–1204, 2006.

[59] Q. S. Chen, J. W. Zhao, C. H. Fang, and D. Wang, "Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM)," *Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy*, vol. 66, no. 3, pp. 568–574, 2007.

[60] Q. S. Chen, J. W. Zhao, and H. Lin, "Study on discrimination of Roast green tea (Camellia sinensis L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition," *Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy*, vol. 72, no. 4, pp. 845–850, 2009.

[61] M. Z. Zhu, G. L. Chen, J. L. Wu, N. Li, Z.-H. Liu, and M.-Q. Guo, "Recent development in mass spectrometry and its hyphenated techniques for the analysis of medicinal plants," *Phytochemical Analysis*, vol. 29, no. 4, pp. 365–374, 2018.

[62] M. Z. Zhu, T. Liu, and M. Q. Guo, "Current advances in the metabolomics study on Lotus seeds," *Frontiers in Plant Science*, vol. 7, 2016.

[63] F. Y. H. Kutsanzedgie, Q. Chen, M. M. Hassan, M. Yang, H. Sun, and H. Rahman, "Near infrared system coupled chemometric algorithms for enumeration of total fungi count in cocoa beans neat solution," *Food Chemistry*, vol. 240, pp. 231–238, 2018.
