

# Application of Molecular spectroscopy and chemometrics for authentication and quality control of tea (*Camellia sinensis* L.): a review

Mitsalina Fildzah Arifah<sup>1</sup>, Agustina A.M.B. Hastuti<sup>1,2</sup> and Abdul Rohman<sup>1,2\*</sup>

1. Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia.

2. Center of Excellence Institute for Halal Industry and Systems (PUI-PT IHIS), Universitas Gadjah Mada, Yogyakarta 55281, Indonesia.

## Info Article

Submitted: 01-01-2021

Revised: 22-01-2022

Accepted: 13-04-2022

\*Corresponding author  
Abdul Rohman

Email:  
abdulkimfar@gmail.com

## ABSTRACT

Tea, derived from *Camellia sinensis* L., is considered as the most popular beverages in the world. The quality of teas may vary depending on harvesting location and geographical origins, thus the traceability of teas according to their origins is very essential to assure tea's quality. Due to economic reasons, high quality tea products may be added with foreign materials or adulterated with low quality ones, as a consequence, some analytical methods have been proposed and developed to quality control of tea. During the recent years, the application of molecular spectroscopic techniques (UV-Vis, Fluorescence, Near Infrared, Mid Infrared, Raman) in combination with multivariate data analysis has emerged as rapid and reliable analytical tool in the quality control of food, including food authentication. The objective of this review is to update the application of molecular spectroscopy (UV-Vis, fluorescence, infrared and Raman) for the quality control and authentication of tea products either geographical origins issue or detection of potential adulterants. The variables obtained during molecular spectral measurement involve hundreds or thousands of data, which make data analysis rather complex. Fortunately, the specific chemometrics tools can solve the problems arising from big data coming from analyte signals, spectral interferences and overlapping peaks. This review paper provides an overview of the recently developed approaches and latest research carried out in molecular spectroscopic techniques in combination with chemometrics for the quality control and for authentication of teas.

**Keywords:** tea, chemometrics, molecular spectroscopy, quality control, authentication analysis.

## ABBREVIATION

ANN = Artificial Neural Network; HCA = Hierarchical Cluster Analysis; kNN = k-Nearest neighbours; LDA = Linear discriminate analysis; MSC = multiplicative scatter correction; NPLS-DA = N-way partial least square discriminate analysis; PCA = Principal Component Analysis; PLS-DA = partial least square discriminate analysis; PCA-DA = Principal Component Analysis-Discriminant Analysis; SNV = standard normal variate; RMSEC = root mean square error of calibration; RMSEP = root mean square error of prediction.

## INTRODUCTION

Tea, originated from China, is considered as one of the most beverages throughout the world. Currently, tea has been planted and produced in several geographical regions including Indonesia,

Argentina, India, Turkey, Vietnam, Sri Lanka, Japan, Kenya, Iran, and etc.) (Shuai *et al.*, 2022). Tea is also export commodity in Indonesia and has been traded in the global market. Some communities believe that drinking tea is not only pleasurable culture but also providing some beneficial effects to human health due to some polyphenolic compounds contained having strong antioxidant activities through neutralizing free radicals which are harmful in the human body. The phytochemicals in tea are also associated with the risk reduction of cancer occurrence, tissue aging, and cardiovascular diseases. Among polyphenols, catechins and gallic acid in particular are considered as the compounds that are most beneficial to human health (Navratilova *et al.*, 2019).

The aroma, flavour, colour, phytochemical composition (types and quantities) and other tea qualities and characteristics are typically affected to some factors including manufacture processing, harvesting time, cleaning the leaves, the levels of drying processes of the leaves, storage and the fermentation degree (unfermented, partially fermented, fully fermented and post-fermented). Based on these factors, tea is typically classified into white, black, green, red tea (oolong), yellow tea and dark tea (Pu-erh). White, Yellow and Green teas are a non-fermented type. Green tea and black tea are considered as the most traditional ones, which represented around 22% and 78% of the production in the world, respectively (Pons *et al.*, 2021). Compared with black and oolong tea, these teas contain higher amounts of polyphenols contributing to high antioxidant activity in teas either *in vivo* or *in vitro* (Gonçalves Bortolini *et al.*, 2021). Furthermore, The price of tea varies depending on brand and geographical origin, therefore, it is very important to authenticate tea from the geographical origins to protect consumers and the global food trade (Cengiz *et al.*, 2017).

Several analytical methods and reviews have been published for the authentication of teas from different origins, detection of adulterants (either legal or non-legal components) in teas, and for quality control of tea mainly based chromatographic methods such as gas chromatography coupled with solid phase micro-extraction and Flame ionization detector-mass spectrometer by determining volatile compounds in tea products (Inarejos-García *et al.*, 2021), ultra-high performance liquid chromatography combined with high resolution mass spectrometric detection (UHPLC-HRMS) and chemometrics by determining fingerprinting profile of tea from different origins (Navratilova *et al.*, 2019), vibrational spectroscopic techniques (infrared, Terahertz, hyper-spectral imaging, and Raman) combined with chemometrics for tea quality, tea authentication and safety analyses (Lin & Sun, 2020), and elemental analysis and isotope ratio combined with Random forest (machine learning) for discrimination of tea origins (Deng *et al.*, 2020). In this review, molecular spectroscopy combined with chemometrics offering fast and simple analytical method was highlighted. The objective of this review is to update the recent progress of the application of molecular spectroscopy (UV-Vis, fluorescence, infrared and Raman) in combination with chemometric techniques for the quality control and for authentication analyses of tea

products (geographical origins, detection of adulterants).

During performing this review, some electronic databases including Web of Science, Scopus, PubMed and Directory Open Access Journal (DOAJ) were explored. The keywords used during the literature searching were “tea AND authentication AND chemometrics AND spectroscopy” and “tea AND origin AND spectra” and “tea AND chemometrics AND spectra. In addition, the Boolean function was also applied. The abstracts of the search results were examined for potential inclusion of the articles.

### **Molecular spectroscopy and Chemometrics**

Molecular spectroscopic (MS)-based techniques can be understood as the interaction between electromagnetic radiation (EMR) with target of analytes in studied samples in the molecular levels. This technique is based on the principle that each molecule which composed of multiple atoms after absorbing EMR undergo some phenomena including electronic transition in UV-Vis spectroscopy, excitation-emission in molecular fluorescence, vibrational transition observed in infrared (near, mid) and rotational transition found in NMR spectroscopy (Zaukuu *et al.*, 2022). Every EMR region has different features and stands for a specific kind of molecular transition which corresponds to specific spectroscopic techniques. UV-Vis spectroscopy and fluorescence spectroscopy can be understood as interaction between EMR in UV-Vis region (corresponds to wavelength 200-800 nm) with samples to give electronic transition (UV-Vis) and or emission, while mid IR spectroscopy related to the interaction between EMR at region of 4000-400  $\text{cm}^{-1}$  to cause the vibrational transitions of chemical bonds or functional groups (Rohman *et al.*, 2020).

One of the characteristics inherent to molecular spectroscopic techniques is the large number of variables, which are several hundreds or even thousands, even from single measurement. The intensity in one or several wavelengths is normally not enough to determine tea characteristics in order to authenticate tea product. Therefore, chemometrics must be used to solve the problems associated with big data analysis. Chemometrics is defined as the use of statistics and mathematics to extract the chemical data into understandable information. In the application of molecular spectroscopy combined with chemometrics for tea authentication, spectral pre-processing is often employed prior to multivariate

data analysis to simplify the data and select the best features of the data (Figure 1). The chemometrics analysis of pattern recognition and/or multivariate calibration are commonly used for further data analysis (Figure 1). Table 1 compiled the application of molecular spectra and chemometrics for authentication and quality control of tea products.

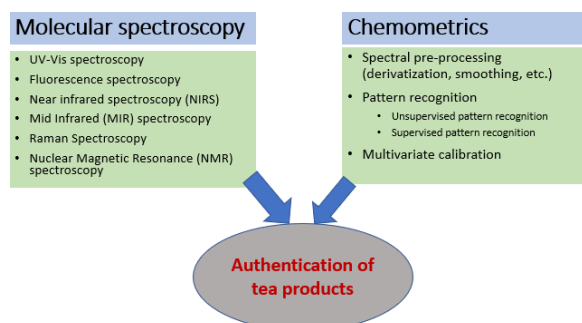


Figure 1. The combination of molecular spectroscopy and chemometrics analysis for authentication of tea products.

Molecular spectroscopy methods produce spectra, which can be understood as the intensity of radiation in many wavelengths. As molecular spectroscopy techniques scan the sample in long wavelength with a short interval, huge amount of data is produced for every single spectrum. Moreover, molecular spectra are typically very broad and may contain some interferences originating from background and overlapping peaks, leading to redundancy and co-linearity. As complex spectra are produced by molecular spectroscopy techniques, some spectral pre-processing could be employed to provide the best chemometrics model (Barbin *et al.*, 2014). Various pre-processing techniques were applied to molecular spectra for the authentication of tea, including but not limited to mean centering, median centering, baseline correction, smoothing, derivatization, standard normal variate (SNV), etc. (Aboulwafa *et al.*, 2019; Buyukgoz *et al.*, 2016; Heryanto *et al.*, 2017). SNV normalizes each spectrum by its individual mean and standard deviation to remove spectral scattering effect (Aboulwafa *et al.*, 2019). Pattern recognition techniques particularly of the unsupervised data analysis can also be used for spectral pre-processing, such as PCA (Buyukgoz *et al.*, 2016).

Analysis of tea by molecular spectroscopy in terms of authentication and quality control use a group of statistical analysis called pattern

recognition techniques to classify samples. The pattern recognition can be divided into unsupervised methods, such as HCA and PCA, and supervised methods, such as PLS-DA, ANN, etc. Unsupervised technique is a group of pattern recognition technique that do not pre-classify data. Hierarchical cluster analysis (HCA) builds hierarchy of data cluster which reflect the similarities of the data. The similarity is computed by the distance of the data, such as using Euclidean distance, Manhattan distance, Mahalanobis distance, etc (Aboulwafa *et al.*, 2019). In contrast to HCA, PCA forms new variables called “principal components” to account as much variance in the data set. This technique allows reduction of dimensionality with minimal information loss (Jolliffe & Cadima, 2016). PCA is the most common unsupervised pattern recognition technique in tea authentication study.

Supervised pattern recognition techniques use the information from predetermined groups to predict the classification of unknown samples. While various techniques of supervised pattern recognition exist, there are several common steps that should be followed during the data analysis. In general, samples with known groups are divided into a training set, a calibration set, and a test set. Those sets should be independent to each other to ensure the validity of the analysis. Then, prospective variables are chosen, which may include spectral pre-processing. With the selected variables, a model is built using chosen multivariate techniques with the training set. The model should be calibrated to optimize the parameters in the multivariate techniques. Lastly, the model is validated to evaluate the ability of the model to recognize and predict unknown samples (Berrueta *et al.*, 2007). Various supervised pattern recognition techniques are available and has been used for the authentication of tea, including kNN, SIMCA, ANN, etc. (Table I). Supervised pattern recognition is a fast-growing field due to the advancement of artificial intelligence.

Multivariate calibration employs multiple variables for the calibration of data set. In a complex data like spectra, univariate calibration will not be sufficient for quantification process in terms of tea authentication. Molecular spectroscopy is a non-selective method, i.e., poor selectivity, especially without a clean-up step a priori and/or extraction of selected compounds (Olivieri, 2018). Nevertheless, for authentication of food, much of the information in the spectra is inherent characteristic of the sample, thus using

Table I. The application of molecular spectra and chemometrics for authentication and quality control of tea products.

Issues	Methods and chemometrics	Results	References
The authentication of Java Preanger steamed green tea (JPGT), highly valuable and very expensive green teas in Indonesia	UV spectra of tea samples were scanned at 190-400 nm. Chemometrics: MSC and moving averaging smoothing; chemometrics: PCA, PLS-DA, LDA and PCA-LDA	PCA was capable of separating JPGT samples from non-JPGT samples, while PLS-DA, LDA and PCA-LDA were successful for the classification of prediction set samples with accuracy of classification levels of 100%	(Suhandy & Yulia, 2021).
The authentication of geographical origins of tea from Vietnam	UV-Vis spectral measurement at 190 to 700 nm with the informative region at (190 – 490 nm. Chemometrics: PLS-DA, PCA-DA and ANN	ANN provide better discrimination model than PCA-DA and PLS-DA model. The accuracy levels of ANN in training and prediction sets were 99% and 84%, respectively.	(Tran <i>et al.</i> , 2021).
The authentication of geographical origins from 38 green tea samples obtained from the South and the East Asian regions	UV-Vis spectra at wavelength of 200-400 nm. Chemometrics: HCA PCA, SIMCA, PLS-DA	PCA using PC1 and PC2 describing 75% and 19% indicated that the first cluster comprised green tea samples from China, while the second one contained tea samples from India and Sri Lanka. SIMCA and PLS-DA confirmed that good class separation of tea samples from the South and the East Asian can be enhanced.	(Aboulwafa <i>et al.</i> , 2019).
The authentication of Anhua brick tea from different types and different brands (23 brands).	Fluorescence spectroscopy around excitation wavelength of 280–360 nm and emission wavelength of 355–500 nm. Chemometrics: LDA, k-NN, PLS-DA, and NPLS-DA	NPLS-DA using full fluorescent data exhibited the best classification results with accuracy levels of 96.5% and 98.3%, respectively for tea types and for brands, respectively.	(Yuan <i>et al.</i> , 2022).
The authentication of green tea from different origins (China and Japan).	Fluorescence spectroscopy; Chemometrics: PCA, PARAFAC, PLS-DA	the combination of fluorescence spectra and PLS-DA provide the discrimination of green tea samples with accuracy levels of sensitivity and specificity of 98.70% and 98.68%, respectively.	(Casale <i>et al.</i> , 2018).
The discrimination of seven varieties of Chinese teas.	FT-NIR spectra; chemometrics: pre-processing spectra of SNV, MC and MSC; discrimination: ANN, LDA	ANN model using 4 PCs was better classification model than LDA models. ANN offered the identification rates of 100% either in the training set or in the prediction sets.	(Chen <i>et al.</i> , 2008).
The classification and discrimination of Chinese Dianhong black tea (CDBT) for the authentication issue.	Two NIRS methods (the portable and benchtop) combined NIR spectra using both spectrometers were recorded using the diffuse reflectance spectra in the wavelengths of 900 and 1,700 nm 11,100-5880 cm <sup>-1</sup> . Chemometrics: GA, SPA, PLS-DA, CDBT	The combination of GA-SVM models using NIR spectra obtained from the handheld spectrometer yielded the best predictive performance with the accuracy rate of classification model of 98.75% (training set) and 100% (in prediction set). the handheld system combined with a suitable chemometric and feature information selection could be used for the rapid and reliable discrimination of CDBT.	(Huang <i>et al.</i> , 2020).

Table I. The application of molecular spectra and chemometrics for authentication and quality control of tea products.

Issues	Methods and chemometrics	Results	References
Quality control of tea by determining free amino acids (FAA) in tea products (Chinese black, dark, oolong, and green teas).	NIRS at wavenumbers of 4000–10,000 cm <sup>-1</sup> using Diffuse reflectance	The combination of CARS-PLS models using NIR spectra as variables provide the best prediction model with R <sup>2</sup> and RMSEP values of 0.994 and 0.595 (for TP), 0.986 and 0.070 for CAFF, and 0.993 and 0.063 for FAA, respectively.	(Wang <i>et al.</i> , 2018)
Quality control of tea products through quantitative analysis model of caffeine (CAFF), epigallocatechin-3-gallate (EGCG), and moisture (MOI) content	NIRS at wavenumbers of 10,000–4,000 cm <sup>-1</sup> as variable during PLSR	The values of R <sub>c2</sub> and R <sub>p2</sub> for CAFF, EGCG and MOI were of (0.9957, 0.9812, 0.9885) and (0.9932, 0.9796, 0.9812), RMSECV values during cross validation of 0.16092, 5.68517 and 0.31884 with bias 0.03, 0.45, and 0.07 respectively.	(S. Zhang <i>et al.</i> , 2021)
Detecting and quantifying tartrazine in tea	FTIR spectroscopy at 4000 - 650 cm <sup>-1</sup> . The chemometrics: PLSR, backward interval PLS (BiPLS), GA-PLS, and CARS-PLS.	RCGA-PLS using variable of absorbance values at 600-1400 cm <sup>-1</sup> as robust technique for predicting tartrazine with low values of RMSECV and RMSEP of 0.8331 and 0.923, respectively. PCA at whole region 1367 - 864 cm <sup>-1</sup> could discriminate pure tea and tea added with tartrazine with concentration ranges of 0 to 30 mg/g.	(Amsaraj & Mutturi, 2021).
Discrimination of tea products according to types and origins	surface enhanced Raman spectroscopy (SERS) combined with chemometrics of PCA	SERS using absorbance values at 200–2000 cm <sup>-1</sup> combined with chemometrics of PCA and pre-processing spectra clearly separated tea products according to types (herbal tea, black and earl grey tea).	(Buyukgoz <i>et al.</i> , 2016).
Detection and quantification of lead chrome green in tea products	Raman spectra using wavenumbers of 2804–230 cm <sup>-1</sup> . Chemometrics: PLSR	PLSR provide the reliable correlation between actual and predicted values of chrome green with R <sub>p</sub> and RMSEP values of 0.936 and 0.803, respectively	Li <i>et al.</i> (2015)

multiple variables will retain the information. Multivariate calibration offers advantages of lower noise, better handling of interferences, possible exploratory aspect, and outlier control (Bro, 2003).

#### Authentication of tea using UV-Vis spectra

Simple UV-vis combined with some chemometrics was used for the authentication of geographical origins from 38 green tea samples obtained from the South and the East Asian regions. UV-Vis spectra at wavelength of 200-400 nm were subjected to multivariate analyses of HCA and PCA.

The model for their authentication validated using SIMCA and PLS-DA. UV data coupled with HCA and PCA successfully distributed the tested samples into informative clusters. PCA using PC1 and PC2 describing 75% and 19% of the data variance indicated that the first cluster comprised green tea samples from East Asian samples (China), while the second one contained tea samples from South Asian samples (India and Sri Lanka). SIMCA and PLS-DA results confirmed that good class separation of tea samples from the South and the East Asian was enhanced. Hence, PLS-DA and

SIMCA using variable of UV-Vis spectra was effective tools for classification of tea samples from different regions (Aboulwafa *et al.*, 2019).

The combination of UV-Vis spectra and chemometrics has been successfully applied for geographical origins of tea from Vietnam. For this purpose, 125 tea samples were subjected to UV-Vis spectral measurement at 190 to 700 nm with the informative region at (190 – 490 nm). PLS-DA, PCA-DA and ANN were compared to construct the classification and discrimination models. The results informed that the ANN provide better discrimination model than PCA-DA and PLS-DA model. The accuracy levels of ANN in training and prediction sets were 99% and 84%, respectively. The proposed method offered simple and rapid technique for geographical origin authentication of teas (Tran *et al.*, 2021).

UV spectroscopy coupled with PLS-DA, LDA and PCA-LDA has been successfully applied for the authentication of Java Preanger steamed green tea (JPGT), highly valuable and very expensive green teas in Indonesia. Three types of green teas were prepared, namely 200 samples of JPGT with geographic indications, 200 samples of Chun Mee green tea grade I and 200 samples of Sow Mee grade II. All UV spectra of tea samples were scanned at 190-400 nm, subjected to pre-processing of MSC and moving averaging smoothing (MAS) at 9 points, and the absorbance values of pre-processed UV spectra were used as variable during chemometrics analysis. PCA was capable of separating JPGT samples from non-JPGT samples, while PLS-DA, LDA and PCA-LDA were successful for the classification of prediction set samples with accuracy of classification levels of 100% (Suhandy & Yulia, 2021). From this result, UV spectra and chemometrics provided simple and cheap method with acceptable results and it is potential to be used in daily quality control of JPGT in tea industry.

The data fusion of two molecular spectroscopic techniques, namely UV-Vis and FTIR spectra has been applied for the classification and discrimination of tea types and teas from different origins. In this study, the samples of black tea (BT) and green tea (GT) from several provinces in Sumatra and Java Islands (Indonesia) were used. Some chemometrics techniques of PCA, kNN, and logistic regression were developed and assessed for the classification purposes. PCA models using variables of UV spectra at 190 - 500 nm and the combined region of FTIR spectra at 1576-1675, 1760-1880 and 3460-3550  $\text{cm}^{-1}$  for FTIR spectra are capable of separating the tea groups according

to types of BT and GT and teas from different origins. Furthermore, kNN and logistic regression models using the same UV-visible spectra could classify GT and BT successfully with classification accuracy of > 80%; while using the same FTIR spectra as variable, kNN provide good classification of studied tea samples from different origins with accuracy level of 90.3% for classifying tea based on its origin. The data sets of data fusion of UV-Vis and FTIR spectra exhibited good separation of tea samples with different types (BT and GT) and origins with a high area under the receiver operating characteristic (ROC) curve of >0.8 and moderate accuracy (0.548). The authors concluded that the data fusion of UV-Vis and FTIR spectra offered the complementary information for the classification of tea samples having different types (BT and GT) and different origins intended for authentication studies (Arifah *et al.*, 2022).

#### **Tea authentication using Molecular fluorescence**

Molecular fluorescence spectroscopy in combination with chemometric methods of pattern recognition has been developed as fat and reliable technique for the authentication of Anhua brick tea from different types and different brands (23 brands). There are four fluorophores intrinsically identified in the brick tea, including theabrownin, tryptophan-like, humic and fulvic-like, and heme-like substances in which the excitation and emission matrix (EEM) of fluorescent intensity revealed that the strongest intensity was observed around excitation wavelength of 280–360 nm and emission wavelength of 355–500 nm. The brick tea with different types and brands were processed based on the PARAFAC factor scores, and EEM of fluorescent data were used as variable during tea's classification using four classification chemometrics of LDA, k-NN, PLS-DA, and NPLS-DA. Among these chemometrics, NPLS-DA using full fluorescent data exhibited the best classification results with accuracy levels of 96.5% and 98.3%, respectively for tea types and for brands, respectively. The authors concluded that the developed method (fluorescence spectroscopy and chemometrics) could be a potential analytical tool for the recognition of Anhua brick tea with different types and brands (Yuan *et al.*, 2022).

The excitation and emission fluorescence spectroscopy combined with chemometrics of PCA and Parallel Factor Analysis (PARAFAC) has been used for authentication of green tea from different origins (China and Japan). Because fluorescence

spectra consist of a number of overlapping signals, the proper spectral pre-processing using SELECT was used. The ability of PLS-DA for the discrimination between two types of green tea from China and Japan was evaluated by the sensitivity and specificity. Using the optimum condition (pre-processing and selection of appropriate variable), the combination of fluorescence spectra and PLS-DA provide the discrimination of green tea samples with accuracy levels of sensitivity and specificity of 98.70% and 98.68%, respectively. This developed method demonstrates the capability of the combination of fluorescence spectroscopy and PARAFAC model for characterizing, discriminating and classification of green tea samples according to its origin (Casale *et al.*, 2018).

### Near Infrared spectroscopy (NIRS)

Due to as fingerprinting technique, NIRS coupled with chemometrics of pattern recognition has evolved as rapid and reliable method for the authentication of teas from different varieties and from different regions (origin). The combination of FT-NIR spectra and chemometrics of LDA and ANN was used for discrimination of seven varieties of Chinese teas. In this study, the number of principle components (PCs) as result from was optimised in the constructing model. FT-NIR spectra were previously subjected to three pre-processing spectra namely SNV, MC and MSC providing that SNV provided the best model. During discrimination, the performance of ANN model using 4 PCs was better than LDA models in which ANN offered the identification rates of 100% either in the training set or in the prediction sets. The overall results exhibited that FT-NIR spectra in combination with pattern recognition of ANN method can be successfully employed for rapid and reliable method for the discrimination of tea varieties (Chen *et al.*, 2008).

Two NIRS methods (the portable and benchtop) combined with chemometrics of pattern recognitions were compared for classification and discrimination of Chinese Dianhong black tea (CDBT). NIR spectra using both spectrometers were recorded using the diffuse reflectance spectra in the wavelengths of 900 and 1,700 nm 11,100-5880  $\text{cm}^{-1}$ . All spectra were pre-processed using genetic algorithm (GA) and successive projections algorithm (SPA) for acquiring the feature variables used in such classification. PLS-DA and SVM were employed for the establishment of discrimination models of CDBT. The analytical results exhibited

that the combination of GA-SVM models using NIR spectra obtained from the handheld spectrometer yielded the best predictive performance with the accuracy rate of classification model of 98.75% (training set) and 100% (in prediction set). This study demonstrated that the handheld system combined with a suitable chemometric and feature information selection could be used for the rapid and reliable discrimination of CDBT (Huang *et al.*, 2020).

NIRS in combination with PLS-DA and SIMCA was applied for the authentication of Darjeeling black tea (DBT). DBT is among tea varieties which is currently part of the register of protected designations of origin (PDO) and protected geographical indications (PGI), therefore, NIRS was used to the authentication analysis of geographical origin of DBT from the other teas and to detect the adulteration practice of DBT with others. NIR spectra subjected to spectral pre-processing of SNV and first derivative provide the least classification error (2.25%). For classification, PLS-DA is successful for the discrimination of DBT from Ceylon (CBT) and Breakfast black teas (BBT) with correct classification of 100% and 92.30% on test samples using NIR spectra previously subjected by SNV+1st derivative and mean centering, respectively (Firmani *et al.*, 2019). Further analysis based on Variable Importance in Projection (VIP) indicated that the most regions contributing to such discrimination are 4250  $\text{cm}^{-1}$  (due to vibrations of O-H, N-H and C-O bands), region around 5350  $\text{cm}^{-1}$  corresponding to 2<sup>nd</sup> overtone of C=O, region around 5775 and 5920  $\text{cm}^{-1}$  due to overtones of CH<sub>2</sub> and CH<sub>3</sub> and region centred in 7150  $\text{cm}^{-1}$  corresponding to stretching of C-H and O-H. Using SIMCA, the classification performance during modelling as expressed by sensitivity and specificity of 84.61% and 97.44%, respectively. The combination of NIRS and chemometrics (SNV and random forest performs) also successfully applied for the classification of tea samples. Random forest provided better classification models than SNM with accuracy levels of classification of 96.8% and 68.8%, respectively (Yang *et al.*, 2021).

The quality control of tea by determining some parameters related to tea products using NIRS in combination with chemometrics was also reported. NIRS in combination with chemometrics of machine learning is also successfully applied for determination and prediction of some parameters namely moisture, caffeine, tea polyphenols, and tea polysaccharides which related to quality control in

instant teas. Some chemometrics namely support vector regression (SVR), binary particle swarm optimization-SVR (BPSO-SVR), partial least squares (PLS), and BPSO-PLS were compared for these determinations. The model performances from these four chemometrics were evaluated by some statistical parameters which are coefficient correlations in calibration ( $R_c$ ) and in prediction ( $R_p$ ), RMSEC and RMSEP. The optimal chemometrics models were selected based on the lowest RMSEC and RMSEP values and the highest values of  $R_c$  and  $R_p$ . From the model evaluations, NIRS subjected to BPSO-SVR provide the optimum model with  $R_c$  for moisture, caffeine, tea polyphenols, and tea polysaccharides were 0.9983, 0.9981, 0.9960 and 0.9954 and  $R_p$  values of 0.9678, 0.9757, 0.7569, and 0.8185, respectively. The values of SEP were less than 0.9302, and absolute values of Bias were less than 0.3667. These results concluded that the combination of NIRS and chemometrics of machine learning offered the optimum chemometrics model for the detection of parameters related to quality control in instant tea components (Bai *et al.*, 2022).

Wang *et al.* (Wang *et al.*, 2018) have successfully employed NIRS in combination with Random frog and competitive adaptive reweighted sampling (CARS) for selecting the variables capable of providing the best calibration model for PLSR for prediction of Total polyphenols (TP), caffeine (CAFF), and free amino acids (FAA) in tea products (Chinese black, dark, oolong, and green teas). In this study, NIR spectra were scanned in Diffuse reflectance mode at wavenumbers of 4000–10,000  $\text{cm}^{-1}$ . As a result, the combination of CARS-PLS models using NIR spectra as variables provide the best prediction model with  $R^2$  and RMSEP values of 0.994 and 0.595 (for TP), 0.986 and 0.070 for CAFF, and 0.993 and 0.063 for FAA, respectively. The results confirmed the feasibility of NIRS-suitable chemometrics for quality control of tea from different categories.

The use of absorbance values of NIRS at wavenumbers of 10,000–4,000  $\text{cm}^{-1}$  as variable during PLSR modelling was also successfully employed for quality control of tea products through quantitative analysis model of caffeine (CAFF), epigallocatechin-3-gallate (EGCG), and moisture (MOI) content. After the optimization process, the values of  $R_c^2$  and  $R_p^2$  for CAFF, EGCG and MOI were of (0.9957, 0.9812, 0.9885) and (0.9932, 0.9796, 0.9812), RMSECV values during cross validation of 0.16092, 5.68517 and 0.31884 with bias 0.03, 0.45, and 0.07 respectively. These

results obtained from FT-NIRS and PLSR were comparable with those using reference method of HPLC indicating that FT-NIRS can be an alternative method for analysis of chemical components and tea parameters intended to quality control of tea products (S. Zhang *et al.*, 2021).

#### **Fourier transform Infrared (FTIR) spectroscopy**

The IR region of basically covers EMR spans from 14,000 to 50  $\text{cm}^{-1}$ . However, the terms of IR spectroscopy were typically exclusively to mid IR corresponding to wavenumbers region of 4000–400  $\text{cm}^{-1}$ . In this region, the interaction of IR beam and chemical bonds (functional groups) in molecules resulted the vibrational transitions which are characteristics to analytes (Rohman *et al.*, 2014). IR spectra are called as fingerprinting spectra as there are no same IR spectra for two different samples. This can be seen that IR spectra of one molecule can be differentiated by other by investigating the differences of number of peaks, peak intensities, and frequency exact of each peak. In IR spectra, each peak corresponded to functional groups absorbing IR beam along with modes of vibrations and intensities. The fingerprinting profiles of FTIR can be exploited for differentiation of tea products due to some factors (origins, cultivation, production process, etc). In addition, with the development of chemometrics software, IR spectroscopy is currently applied for quantitative analysis of components present in the samples (in tea products), according to Lambert-Beer law (Rohman *et al.*, 2017).

Due to the nature as fingerprinting analytical tools, FTIR spectra are widely employed for authentication of tea. Recently geographical origins of Iranian teas have been assessed using FTIR spectroscopy and classification chemometrics. PCA, PCA-LDA and PLS-DA were investigated and compared to get the best discrimination profiles of Iranian tea samples. A-150 tea samples were randomly divided into calibration set samples (100) and the prediction set samples (50) and added with KBr and subjected to transmission spectral measurement. FTIR spectra were subjected to pre-processing such mean centering (MC), derivatization (first and second derivatives) auto-scaling, MSC, SNV and their combinations to improve the quality of FTIR spectra capable of providing the best classification models. The results showed that using absorbance values of FTIR spectra subjected to pre-processing of MSC+1<sup>st</sup> Derivative+ (MC) and SNV + 1<sup>st</sup>



Derivative+ (MC), the Iranian tea samples from five geographical regions (Shaft, Lahijan, Leilakoo, Oshiyani, and Ramsar), can be clearly discriminated with classification accuracy of 98% for both (Esteki *et al.*, 2022). The results confirmed that FTIR spectra combined with appropriate chemometrics techniques are rapid and reliable methods for the discrimination and classification of Iranian teas according to their geographical origins.

Tea may be added with foreign components such as tartrazine to give the appearance of tea or tea powders, thus, a method capable of detecting and quantifying tartrazine in tea is needed. One of the potential methods to be used for determination of tartrazine is FTIR spectroscopy. Some chemometrics of PLSR and its variants including backward interval PLS (BiPLS), genetic algorithm PLS (GA-PLS), and competitive adaptive reweighted sampling PLS (CARS-PLS) for variable selection were compared for establishing the calibration models for the determination of tartrazine. In addition, efficient real-coded GA (RCGA) was also employed as a variant of GA-PLS regression. Among these, RCGA-PLS using variable of absorbance values at 600–1400  $\text{cm}^{-1}$  was selected as a robust technique for predicting tartrazine with low values of RMSECV and RMSEP of 0.8331 and 0.923, respectively. PCA at the whole region 1367–864  $\text{cm}^{-1}$  could discriminate pure tea and tea added with tartrazine with concentration ranges of 0 to 30 mg/g. From this result, RCGA-PLS provides the superior method compared to other variable selection algorithms by achieving the least RMSECV and RMSEP (Amsaraj & Mutturi, 2021).

### Raman spectroscopy

Due to its property as a fingerprinting analytical technique, Raman spectroscopy offers rapid and reliable tools for detection of foreign components in tea products and for quality control of tea such as determination of tea freshness (Y. Zhang *et al.*, 2020). Li *et al.* (2015) have developed Raman spectra and chemometrics for detection and quantification of hazardous substances of lead chromate green in tea. Lead chromate, typically used to disguise as high-quality, is forbidden to be added into tea. In this study, 160 tea infusion samples were added with lead chromate with different concentrations and were prepared, and their Raman spectra were scanned in the range of 2804–230  $\text{cm}^{-1}$ . The intensities of Raman spectra were used as variables during prediction modelling of lead chromate. Raman spectra were subjected to wavelet transformation (WT) to extract information in

different time and frequency domains, for the establishment of lead chromate prediction model. PLSR was used for the correlation between actual and predicted values of Raman spectra offering good performance in prediction as indicated by  $R_p$  and RMSEP values of 0.936 and 0.803, respectively. These results indicated the feasibility of Raman spectra and appropriate chemometrics for non-destructive quantitative analysis of lead chromate green in tea products.

The fingerprinting profiling based on surface-enhanced Raman spectroscopy (SERS) combined with chemometrics of PCA is successfully applied for the differentiation of tea products intended for quality control based on cultivation and processing. In this study, seven different varieties of tea samples (herbal tea with types of rose hip, chamomile, linden, green and sage tea), black tea and earl grey tea) were measured by SERS in the range of 200–2000  $\text{cm}^{-1}$  with an analysis time of 20 s. PCA using this variable was used for the separation of each tea variety. In addition, some pre-processing Raman spectra were also applied to give the best model. The result showed that herbal tea samples were clearly separated using Raman spectra previously processed using mean centering, smoothing and median centering pre-processing, while samples of black and earl grey tea were clearly separated using PCA after the Raman spectra were pre-processed using baseline correction and derivatization. The combination of SERS and PCA is proven for differentiation of tea samples according to their types and cultivation (Buyukgoz *et al.*, 2016).

### CONCLUSION

Considered as the most consumed drinks, tea must be assured in terms of quality and authenticity. The quality of teas depends on several factors such as tea types and tea's origins. For specific reasons and driven by economical profits, tea products may be falsely labelled, therefore, analytical chemists have developed analytical methods capable of detecting tea falsification and tea authentication based on origins. Molecular spectroscopy (UV-Vis, Fluorescence, Near Infrared, Mid Infrared, Raman) in combination with multivariate data analysis or chemometrics have been proven to evolve as rapid and reliable methods for tea authentication. Next, molecular-spectroscopic methods applied for tea authentication must be standardized through collaborative studies among qualified laboratories.

## ACKNOWLEDGEMENT

The authors acknowledge Directorate of Research Universitas Gadjah Mada for the financial support of this study through Program Rekognisi Tugas Akhir (RTA) year 2021 with contract number: 1185/UN1.P.III/SK/HUKOR/2021 awarded to Prof. Dr. Abdul Rohman.

## REFERENCES

- Aboulwafa, M. M., Youssef, F. S., Gad, H. A., Sarker, S. D., Nahar, L., Al-Azizi, M. M., & Ashour, M. L. (2019). Authentication and discrimination of green tea samples using UV-vis, FTIR and HPLC techniques coupled with chemometrics analysis. *Journal of Pharmaceutical and Biomedical Analysis*, *164*, 653–658. <https://doi.org/10.1016/j.jpba.2018.11.036>
- Amsaraj, R., & Mutturi, S. (2021). Real-coded GA coupled to PLS for rapid detection and quantification of tartrazine in tea using FT-IR spectroscopy. *Lwt*, *139*(November 2020), 110583. <https://doi.org/10.1016/j.lwt.2020.110583>
- Arifah, M. F., Hastuti, A. A. M. B., & Rohman, A. (2022). Utilization of UV-visible and FTIR spectroscopy coupled with chemometrics for differentiation of Indonesian tea: an exploratory study. *Indonesian Journal of Pharmacy*, *33*(2), 200–207. <https://doi.org/10.22146/ijp.3795>
- Bai, X., Zhang, L., Kang, C., Quan, B., Zheng, Y., Zhang, X., Song, J., Xia, T., & Wang, M. (2022). Near-infrared spectroscopy and machine learning-based technique to predict quality-related parameters in instant tea. *Scientific Reports*, *12*(1), 1–8. <https://doi.org/10.1038/s41598-022-07652-z>
- Barbin, D. F., Felicio, A. L. de S. M., Sun, D. W., Nixdorf, S. L., & Hirooka, E. Y. (2014). Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview. *Food Research International*, *61*, 23–32. <https://doi.org/10.1016/j.foodres.2014.01.005>
- Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. In *Journal of Chromatography A* (Vol. 1158, Issues 1–2, pp. 196–214). <https://doi.org/10.1016/j.chroma.2007.05.024>
- Bro, R. (2003). Multivariate calibration: What is in chemometrics for the analytical chemist? *Analytica Chimica Acta*, *500*(1–2), 185–194. [https://doi.org/10.1016/S0003-2670\(03\)00681-0](https://doi.org/10.1016/S0003-2670(03)00681-0)
- Buyukgoz, G. G., Soforoglu, M., Basaran Akgul, N., & Boyaci, I. H. (2016). Spectroscopic fingerprint of tea varieties by surface enhanced Raman spectroscopy. *Journal of Food Science and Technology*, *53*(3), 1709–1716. <https://doi.org/10.1007/s13197-015-2088-5>
- Casale, M., Pasquini, B., Hooshyari, M., Orlandini, S., Mustorgi, E., Malegori, C., Turrini, F., Ortiz, M. C., Sarabia, L. A., & Furlanetto, S. (2018). Combining excitation-emission matrix fluorescence spectroscopy, parallel factor analysis, cyclodextrin-modified micellar electrokinetic chromatography and partial least squares class-modelling for green tea characterization. *Journal of Pharmaceutical and Biomedical Analysis*, *159*, 311–317. <https://doi.org/10.1016/j.jpba.2018.07.001>
- Cengiz, M. F., Turan, O., Ozdemir, D., Albayrak, Y., Perincek, F., & Kocabas, H. (2017). Geographical origin of imported and domestic teas (*Camellia sinensis*) from Turkey as determined by stable isotope signatures. *International Journal of Food Properties*, *20*(12), 3234–3243. <https://doi.org/10.1080/10942912.2017.1283327>
- Chen, Q., Zhao, J., Liu, M., & Cai, J. (2008). Nondestructive Identification of Tea (*Camellia sinensis* L.) varieties using FT-NIR spectroscopy and pattern recognition. *Czech Journal of Food Sciences*, *26*(5), 360–367. <https://doi.org/10.17221/1125-cjfs>
- Deng, X., Liu, Z., Zhan, Y., Ni, K., Zhang, Y., Ma, W., Shao, S., Lv, X., Yuan, Y., & Rogers, K. M. (2020). Predictive geographical authentication of green tea with protected designation of origin using a random forest model. *Food Control*, *107*(July 2019), 1–9. <https://doi.org/10.1016/j.foodcont.2019.106807>
- Esteki, M., Memarbashi, N., & Simal-Gandara, J. (2022). Classification and authentication of tea according to their geographical origin based on FT-IR fingerprinting using pattern recognition methods. *Journal of Food Composition and Analysis*, *106*(November 2021), 104321. <https://doi.org/10.1016/j.jfca.2021.104321>
- Firmani, P., De Luca, S., Bucci, R., Marini, F., & Biancolillo, A. (2019). Near infrared (NIR) spectroscopy-based classification for the

- authentication of Darjeeling black tea. *Food Control*, 100(February), 292–299. <https://doi.org/10.1016/j.foodcont.2019.02.006>
- Gonçalves Bortolini, D., Windson Isidoro Haminiuk, C., Cristina Pedro, A., de Andrade Arruda Fernandes, I., & Maria Maciel, G. (2021). Processing, chemical signature and food industry applications of *Camellia sinensis* teas: An overview. *Food Chemistry: X*, 12(October). <https://doi.org/10.1016/j.fochx.2021.100160>
- Heryanto, R., Pradono, D. I., Marlina, E., & Darusman, L. K. (2017). Classification of Java tea (*Orthosiphon aristatus*) quality using FTIR spectroscopy and chemometrics. *Journal of Physics: Conference Series*, 835(1). <https://doi.org/10.1088/1742-6596/835/1/012012>
- Huang, J., Ren, G., Sun, Y., Jin, S., Li, L., Wang, Y., Ning, J., & Zhang, Z. (2020). Qualitative discrimination of Chinese dianhong black tea grades based on a handheld spectroscopy system coupled with chemometrics. *Food Science and Nutrition*, 8(4), 2015–2024. <https://doi.org/10.1002/fsn3.1489>
- Inarejos-García, A. M., Helbig, I., Klette, P., Weber, S., Maeder, J., & Morlock, G. E. (2021). Authentication of Commercial Powdered Tea Extracts (*Camellia sinensis* L.) by Gas Chromatography. *ACS Food Science & Technology*, 1(4), 596–604. <https://doi.org/10.1021/acsfoodscitech.1c00003>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065). <https://doi.org/10.1098/rsta.2015.0202>
- Li, X. L., Sun, C. J., Luo, L. Bin, & He, Y. (2015). Nondestructive detection of lead chrome green in tea by Raman spectroscopy. *Scientific Reports*, 5, 1–9. <https://doi.org/10.1038/srep15729>
- Lin, X., & Sun, D. W. (2020). Recent developments in vibrational spectroscopic techniques for tea quality and safety analyses. *Trends in Food Science and Technology*, 104(February), 163–176. <https://doi.org/10.1016/j.tifs.2020.06.009>
- Navratilova, K., Hrbek, V., Kratky, F., Hurkova, K., Tomaniova, M., Pulkrabova, J., & Hajslova, J. (2019). Green tea: Authentication of geographic origin based on UHPLC-HRMS fingerprints. *Journal of Food Composition and Analysis*, 78(February), 121–128. <https://doi.org/10.1016/j.jfca.2019.02.004>
- Olivieri, A. C. (2018). Chemometrics and Multivariate Calibration. *Introduction to Multivariate Calibration*, 1–17. [https://doi.org/10.1007/978-3-319-97097-4\\_1](https://doi.org/10.1007/978-3-319-97097-4_1)
- Pons, J., Bedmar, À., Núñez, N., Saurina, J., & Núñez, O. (2021). Tea and chicory extract characterization, classification and authentication by non-targeted HPLC-UV-FLD fingerprinting and chemometrics. *Foods*, 10(12). <https://doi.org/10.3390/foods10122935>
- Rohman, A., bin Che Man, Y., Ismail, A., & Hashim, P. (2017). *International Journal of Food Properties FTIR spectroscopy coupled with chemometrics of multivariate calibration and discriminant analysis for authentication of extra virgin olive oil FTIR spectroscopy coupled with chemometrics of multivariate calibration*. <https://doi.org/10.1080/10942912.2017.1336718>
- Rohman, A., Ghazali, M. A. B., Windarsih, A., Irnawati, Riyanto, S., Yusof, F. M., & Mustafa, S. (2020). Comprehensive Review on Application of FTIR Spectroscopy Coupled with Chemometrics for. *Molecules*, 25(5485), 1–28.
- Rohman, A., Nugroho, A., Lukitaningsih, E., & Sudjadi. (2014). Application of vibrational spectroscopy in combination with chemometrics techniques for authentication of herbal medicine. *Applied Spectroscopy Reviews*, 49(8), 603–613. <https://doi.org/10.1080/05704928.2014.882347>
- Shuai, M., Peng, C., Niu, H., Shao, D., Hou, R., & Cai, H. (2022). Recent techniques for the authentication of the geographical origin of tea leaves from *camellia sinensis*: A review. *Food Chemistry*, 374(April 2021), 131713. <https://doi.org/10.1016/j.foodchem.2021.131713>
- Suhandy, D., & Yulia, M. (2021). Authentication of Java Preanger steamed green tea by using ultraviolet spectroscopy and discriminant analysis method. *Journal of Engineering Science and Technology*, 16(6), 4650–4665.
- Tran, T. H., Tran, Q. T., Ta, T. T., & Le, S. H. (2021). Geographical origin identification of teas

- using UV-VIS spectroscopy. *E3S Web of Conferences*, 265. <https://doi.org/10.1051/e3sconf/202126505013>
- Wang, J., Wang, Y., Cheng, J., Wang, J., Sun, X., Sun, S., & Zhang, Z. (2018). Enhanced cross-category models for predicting the total polyphenols, caffeine and free amino acids contents in Chinese tea using NIR spectroscopy. *Lwt*, 96(September 2017), 90–97. <https://doi.org/10.1016/j.lwt.2018.05.012>
- Yang, J., Wang, J., Lu, G., Fei, S., Yan, T., Zhang, C., Lu, X., Yu, Z., Li, W., & Tang, X. (2021). TeaNet: Deep learning on Near-Infrared Spectroscopy (NIR) data for the assurance of tea quality. *Computers and Electronics in Agriculture*, 190(August), 106431. <https://doi.org/10.1016/j.compag.2021.106431>
- Yuan, Y., Jiang, J., Yang, Z., Li, H., & Qiu, B. (2022). Excitation-emission matrix fluorescence spectroscopy coupled with chemometric methods for characterization and authentication of Anhua brick tea. *Journal of Food Composition and Analysis*, 109(February), 104501. <https://doi.org/10.1016/j.jfca.2022.104501>
- Zaukuu, J. L. Z., Benes, E., Bázár, G., Kovács, Z., & Fodor, M. (2022). Agricultural Potentials of Molecular Spectroscopy and Advances for Food Authentication: An Overview. *Processes*, 10(2), 1–42. <https://doi.org/10.3390/pr10020214>
- Zhang, S., Zuo, Y., Wu, Q., Wang, J., Ban, L., Yang, H., & Bai, Z. (2021). Development and Validation of Near-Infrared Methods for the Quantitation of Caffeine, Epigallocatechin-3-gallate, and Moisture in Green Tea Production. *Journal of Analytical Methods in Chemistry*, 2021. <https://doi.org/10.1155/2021/9563162>
- Zhang, Y., Gao, W., Cui, C., Zhang, Z., He, L., Zheng, J., & Hou, R. (2020). Development of a method to evaluate the tenderness of fresh tea leaves based on rapid, in-situ Raman spectroscopy scanning for carotenoids. *Food Chemistry*, 308(June 2019), 125648. <https://doi.org/10.1016/j.foodchem.2019.125648>