Food Analytical Methods (2021) 14:300–306
https://doi.org/10.1007/s12161-020-01885-2

Spectrometric Classification of Bamboo Shoot Species by Comparison of Different Machine Learning Methods

Long Tong[1] · Bin Li[1] · Yanhui Geng[1] · Lijie Chen[1] · Yanjie Li[2] · Ruishu Cao[3]

Received: 9 May 2020 / Accepted: 12 October 2020 / Published online: 16 October 2020

Abstract

The nutrition and quality of bamboo shoots from different species have a large variation, and traditional methods used for the classification of different bamboo shoot species are expensive and time-consuming. Here, the capability of near-infrared reflectance (NIR) spectroscopy to identify bamboo shoot species in a time- and cost-effective manner was examined. The NIR spectra of four bamboo shoot species were collected. Three classification models, a support vector machine (SVM), partial least squares-discriminant analysis (PLSDA), and random forest (RF), were calibrated. Several spectra pre-processing methods and their combination were trained before model calibration for the best classification model collection; each model was run 200 times for the calculation of the prediction error and model stability checking. The SVM model combined with the Det+2nd derivative had the best performance with an overall accuracy of 0.95. The use of less than 16% of the full-length NIR spectra produced a similar high accuracy of 0.91. Eight important regions, 1015, 1135, 1175, 1338, 1380, 1620, 1690, and 1750 nm, were found to be highly related to the classification of bamboo shoots. This study found that NIR spectra combined with SVM methods produced a rapid and non-destructive approach for the classification of bamboo shoot species.

Keywords NIR spectroscopy · Model · Bamboo shoot · Pre-processing method · Variable selection

Introduction

Bamboo, a major non-wood forest product that belongs to the Poaceae family, is well known for its industrial uses (Fu 2001). Adult bamboo wood is one of the most important replacements of wood resources in the wood industry (Janssen 2000). Additionally, another important property of bamboo is the edible juvenile bamboo shoot. The utilization of bamboo shoots as food is a traditional food culture in China for more than 2500 years with their rich nutrient contents and delicious taste (Satya et al. 2010). There are more than 300 different species of bamboo in Asia, and most of them produce edible shoots, but less than 100 species are utilized for food (Grosser and Liese 1971). Bamboo shoots can be easily catalogized as two types, winter bamboo shoots and spring bamboo shoots, of which spring bamboo shoots are more popular (Choudhury et al. 2012). However, the taste of these bamboo shoots varies greatly, and with a large number of suboptimal qualities of bamboo shoots in the market, the healthy development of the bamboo shoot market will be seriously influenced (Kumar et al. 2017). The quality of bamboo shoots is required due to the growing demand of the market. The traditional methods for the identification of bamboo shoots are mainly the naked eye or laboratory chemical analysis methods (i.e., wet chemical analysis), which are expensive and time-consuming. Additionally, the destruction of samples resulting from wet chemical analysis is also a concern. A fast and reliable alternative method to classify different bamboo shoot species is needed. Near-infrared reflectance (NIR) spectroscopy has the advantages of rapid analysis, being repeatable, and being non-destructive and has been widely applied to agriculture and the food factory (Nicolai et al. 2007; Porep et al. 2015).

For example, NIR spectroscopy was used to investigate the quality of fruits such as strawberry (Amodio et al. 2017), apple (Beghi et al. 2014), and banana (Zude 2003), as well as identify plants such as tea (Li and He 2008), *Eucalyptus* species (Castillo et al. 2008), and grapevine (Gutierrez et al. 2015). Additionally, NIR spectroscopy has been used to determine

✉ Yanjie Li
aj7105@gmail.com

[1] Chongqing Academy of Forestry, Chongqing 400036, China

[2] Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou 311400, Zhejiang, People's Republic of China

[3] Institute of Horticulture science, Massey University, Private Bag 11222, Palmerston North 4442, New Zealand

bamboo properties. For instance, NIR showed promise for the discrimination of three bamboo species by the scanning of leaves (Wang et al. 2016). The lignification that is associated with the crude fiber content and firmness of bamboo shoots has also been successfully predicted by NIR spectroscopy (Xu et al. 2014). Furthermore, spectroscopic determination of the chemical composition and classification of bamboo fractions was also achieved (Ramirez et al. 2015). However, fewer studies have demonstrated the ability to discriminate different bamboo shoot species based on NIR spectroscopy.

NIR spectroscopy produces profiles containing a large amount of information, which contains not only important information according to the target but also a lot of irrelevant noise (Brenchley et al. 1997). It is important to find a useful method to profitably exploit the useful spectral information for better calibration. Recently, multivariate mathematics and chemometric statistics methods, including support vector machines (SVM) (Devos et al. 2009; Vishwanathan and Murty 2002), partial least squares-discriminant analysis (PLSDA) (Ballabio and Consonni 2013), and random forest (RF) (Liaw and Wiener 2002), combined with NIR spectroscopy, have been successfully and widely used for classification. For instance, RF has been successfully applied to distinguish two pigweeds from three soybean varieties and yielded high classification accuracies ranging from 93.8 to 100% (Fletcher and Reddy 2016). SVM produced efficient and promising results in the discrimination of three types of tea (Chen et al. 2007). PLSDA combined with NIR spectroscopy could be used as a rapid and non-destructive method to discriminate castor seeds for breeding programs (Santos et al. 2014). Apart from the classification methods, the important feature selection could also highly influence the model performance. There are massive overtones and combinations of vibration information from C-H, O-H, and N-H groups that interact with NIR spectra in plant samples (Yang et al. 2018), and most of them overlap strongly, which will influence the robustness and reliability of model calibration (Inagaki et al. 2018). Pre-processing methods, such as detrending and derivatives, as well as variable selection algorithms (Caliari et al. 2017; Mancini et al. 2018) combined with chemometric statistics could efficiently reduce these bands and eliminate the irrelevant variables that influence model calibration (Rinnan et al. 2009). However, the comparison of different pre-processing methods combined with different classification methods, including SVM, PLSDA, and RF, for discrimination of multiple bamboo shoot species has not been well researched.

Therefore, the present paper aims (1) to evaluate the capacity of NIR reflectance spectroscopy to discriminate four types of bamboo shoot species using SVM, PLSDA, and RF methods; (2) to compare the performance of different NIR spectra pre-processing methods in classification models for the best discrimination of bamboo shoots; and, most importantly, (3) to identify the most important variables related to bamboo shoot identification and to test the possibility of using optimal spectral informative variables for discrimination.

Methods and Materials

Sample Collection

In spring of 2019, a total number of 747 bamboo shoots from four of the most common edible bamboo shoot species at three bamboo forest sites in Chongqing were selected for classification model building: Baijia (BJ) (*Phyllostachys bissetii*) bamboo, Gaojie (G) (*Phyllostachys prominens*) bamboo, Shui (S) (*Phyllostachys heteroclada*) bamboo, and Ping (P) (*Qiongzhuea communis* Hsueh) bamboo. Details are shown in Table 1.

NIR Spectra Collection

All fresh bamboo shoots were extracted from the soil without roots and labeled. The outside shell of the bamboo shoots was peeled off, and simultaneously, NIR spectra were collected using a field spectrometer (LF-2500, Spectral Evolution, USA) with a 5 mm diameter fiber optics probe and spectral band range from 1100 to 2500 nm at 5.5 nm intervals averaging 32 scans. For each bamboo shoot sample, three lines 120° across each other were first established, which mean each sample including three lines. In each line, NIR spectra were collected from the bottom to the top with a fixed distance of 10 mm, and then all spectra from the three lines of each sample were averaged for final use. In total, 747 averaged spectra (from 747 samples) were collected. Spectra ranging from 1100 to 2500 nm was used; wavelengths within this range are related to plant nutrition components (Gillon et al. 1999; Min et al. 2006).

Model Calibration and Validation

Three classification methods, PLSDA, SMV, and RF, and three spectra pre-processing methods and their combination, i.e., detrending (Det), first (1st), and second (2nd) derivatives using Savitzky-Golay smoothing with a window size of 15 data points and a polynomial order of 2 (Press and Teukolsky 1990), were used to obtain the best classification model. In total, six spectra pre-processing methods were used for each machine learning model: raw spectra, Det, 1st derivative, 2nd derivative, Det+1st derivative, and Det+2nd derivative. The best pro-processing and classification methods were chosen for future use. For each classification model, 80% of the data set was randomly selected for internal calibration, and the remaining 20% was used for validation. This

Table 1 Main site characteristics of four bamboo shoot species

Name	Number	Sampling time	Elevation (m)	Location name	Location coordinates (N/E)
G	115	29-Apr-2019	688	Zhongxian bamboo forest	107° 56′, 30° 19′
BJ	270	30-Apr-2019	555	Liangping bamboo forest	107° 32′, 30° 43′
S	120	29-Apr-2019	986	Fuling bamboo forest	107° 38′, 29° 36′
P	242	30-Apr-2019	909	Fuling bamboo forest	107° 40′, 29° 37′

randomized permutation has been conducted 200 times to determine the performance evaluation. This method was first mentioned by Couture et al. (2016) to estimate the model stability, which could provide the classification error based on 200 calibration models; this method was highly recommended for model calibration and validation. The overall accuracy (OVA.ACC), specificity (Spe), sensitivity (Sen), negative prediction (Neg), and positive prediction (Pos) from both calibration (Cal) and validation (Val) were used to track the model performance. The threshold for interpreting probabilities to class labels is 0.5. All analyses were conducted in R software (version 3.1.2) (R Core Team 2017). The e1071 package (Meyer et al. 2018) in R was used for SVM model performing, the caret package (Wing et al. 2017) was used for PLSDA and RF model building, the prospectr package (Stevens and Ramirez-Lopez 2014) was used for spectra pre-processing, and the ggplot2 package (Wickham 2016) was used for data visualization.

Results

Spectra Information

The average of the four bamboo shoot species original (no pre-processing) and Det+2nd derivative spectra is displayed in Fig. 1. The four types of bamboo shoots had similar curves in the original spectra, and it was difficult to observe differences with the naked eye. However, three regions were found that differed among these four bamboo shoots: 1680, 1950, and 2040 nm based on Det+2nd derivative spectra processing. The results indicated that NIR has the potential for use in bamboo shoot identification.

Model Comparison

The performance of six types of pre-processing methods combined with three classification models is shown in Table 2. The Spe, Sen, Neg, Pos, and OVA.ACC were recorded for the comparison of model performance. Despite the three different NIR classification models, six spectra pre-processing methods presented different performances compared to other pre-processing methods: The Det+2nd derivative had the highest value among Spe Sen, Neg, Pos, and OVA.ACC, followed by the Det+1st derivative, 2nd derivative, 1st derivative, and Det; raw spectra (non-processing) had the lowest value among all models. The SVM model had the best model performance among all of the six pre-processing methods. Therefore, the best pre-processing method and classification models were the Det+2nd derivative and SVM, which produced a mean and range of Spe, Sen, Neg, Pos, and OVA.ACC values all equal to 1 in the calibration and a mean Spe value of 0.98 (range 0.92–1), mean Sen value of 0.96 (range 0.83–1), mean Neg value of 0.98 (range 0.91–1), mean Pos value of 0.96 (range 0.84–1), and mean OVA.ACC value of 0.95 (range 0.91–0.99) in the validation. Very poor classification model performance with respect to the Spe, Sen, Neg, Pos, and OVA.ACC value range error was obtained from the 200 simulated times for the SVM model using the Det+2nd derivative NIR spectra.

Variable Importance of NIR Spectra Applied to Bamboo Shoot Discrimination

The importance of spectral variables used by the SVM model combined with the Det+2nd derivative is plotted in Fig. 2. The black color represents the important variables for the four bamboo shoots selected by SVM. It can be observed that the SVM model selected most of the important spectral regions that were similar among the four bamboo shoot species. Eight wavelengths bands were found that highly influenced the model accuracy, i.e., 1015, 1135, 1175, 1338, 1380, 1620, 1690, and 1750 nm; among them, the bands around 1015, 1135, and 1338 nm are considered the most important regions. The most important spectral bands were mostly located in the 1000–1800 nm region.

Model Evaluation

In total, 40 out of 256 variables considered to be important to the model classification were selected to build the SVM classification model. Then, this model was applied to the validation set as a comparison with the full-length spectra SVM model. The confusion matrix of misclassification for validation data predicted by selected variables and full-length variables in the SVM model is displayed in Fig. 3. A high classification accuracy of 0.95 was obtained from the full-length

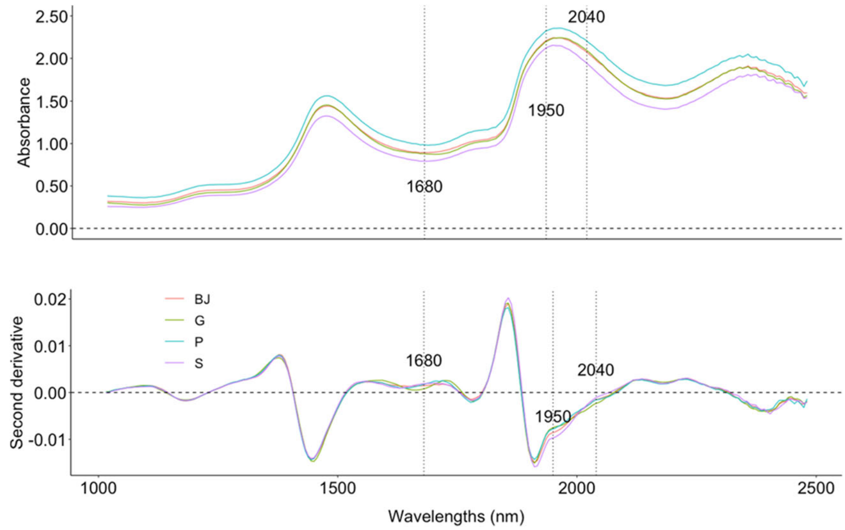

Fig. 1 The average original (Raw)-NIR (upper) and Det+2nd derivative-NIR (lower) spectra of four bamboo species. Dotted line: the position same as in (**b**). Each line represents one bamboo shoot specie; the spectra was average by species level

spectra (Fig. 3a). However, a reasonable and high classification accuracy of 0.91was also produced by the important variables, which is only 16% less than that of the full variables (Fig. 3b).

Discussion

Different classification methods may have different accuracies depending on the target materials. In this study, a high range of OVA.ACC from 0.91 to 0.99 was found using the Det+2nd spectra and SVM model. The SVM model performed better than PLSDA and RF, similar to the finding of Tankeu et al. (2018), who reported that by using of hyperspectral imaging data, the SVM model yielded better predictions for the identification of true black cohosh, and Sun et al. (2017), who showed that the SVM model combined with a hyperspectral reflectance imaging technique had high accuracy (92.96–97.28%) in the detection of chilling peaches. SVM combined with NIR spectroscopy also yielded a promising result for the prediction of the chilling storage stage of eggplants (Tsouvaltzis et al. 2020) and different rice flour types (Sampaio et al. 2020). In addition, SVM methods can be applied to image classification; for example, the SVM model has been used to classify different soil types using soil images. In contrast to these results, PLSDA had better accuracy than SVM in the identification of transgenic maize kernels (Feng et al. 2017). Spectra pre-processing methods could improve the model performance. It has been found that compared to the raw spectra without pre-processing, different pre-processing

Table 2 Distribution (95% confidence intervals) of Spe, Sen and Neg, Pos and OVA.ACC values in the calibration and validation statistics from 200 simulations of bamboo shoot model classification from SVM, PLSDA, and RF methods and six different pre-processing methods using full-length NIR spectra. Each model permutation included 80% of the data for internal calibration and the remaining 20% for validation

Model	Pre-pro	Cal					Val				
		Neg	Pos	Sen	Spe	OVA.ACC	Neg	Pos	Sen	Spe	OVA.ACC
PLSDA	1st	0.96 (0.91–1)	0.89 (0.67–1)	0.90 (0.72–1)	0.95 (0.91–1)	0.89 (0.86–0.91)	0.95 (0.85–1)	0.85 (0.61–1)	0.87 (0.63–1)	0.95 (0.80–1)	0.85 (0.79–0.9)
	2nd	0.96 (0.93–1)	0.90 (0.78–1)	0.91 (0.73–1)	0.97 (0.9–1)	0.91 (0.89–0.92)	0.95 (0.87–1)	0.87 (0.64–1)	0.87 (0.63–1)	0.96 (0.83–1)	0.87 (0.79–0.93)
	Det	0.97 (0.89–1)	0.89 (0.68–1)	0.89 (0.67–1)	0.95 (0.91–1)	0.88 (0.85–0.90)	0.93 (0.80–1)	0.81 (0.6–1)	0.81 (0.63–1)	0.93 (0.72–1)	0.79 (0.70–0.87)
	Det + 1st	0.96 (0.91–1)	0.89 (0.73–1)	0.90 (0.69–1)	0.95 (0.92–1)	0.89 (0.87–0.91)	0.94 (0.85–1)	0.85 (0.62–1)	0.87 (0.63–1)	0.95 (0.83–1)	0.85 (0.79–0.90)
	Det + 2nd	0.96 (0.94–1)	0.90 (0.78–1)	0.91 (0.69–1)	0.97 (0.91–1)	0.90 (0.89–0.92)	0.94 (0.84–1)	0.85 (0.62–1)	0.85 (0.63–1)	0.95 (0.77–1)	0.85 (0.78–0.93)
	Raw	0.96 (0.90–1)	0.89 (0.71–1)	0.9 (0.70–1)	0.96 (0.91–1)	0.89 (0.87–0.91)	0.94 (0.82–1)	0.83 (0.62–1)	0.83 (0.63–1)	0.94 (0.81–1)	0.83 (0.76–0.91)
RF	1st	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	0.95 (0.86–1)	0.86 (0.64–1)	0.85 (0.60–1)	0.95 (0.82–1)	0.85 (0.79–0.92)
	2nd	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	0.97 (0.89–1)	0.91 (0.72–1)	0.91 (0.67–1)	0.97 (0.85–1)	0.90 (0.85–0.97)
	Det	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	0.94 (0.85–1)	0.83 (0.67–1)	0.83 (0.61–1)	0.95 (0.8–1)	0.83 (0.72–0.90)
	Det+1st	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	0.96 (0.87–1)	0.89 (0.68–1)	0.88 (0.63–1)	0.96 (0.81–1)	0.88 (0.81–0.94)
	Det+2nd	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	0.97 (0.91–1)	0.91 (0.71–1)	0.91 (0.75–1)	0.97 (0.87–1)	0.91 (0.86–0.97)
	Raw	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	0.88 (0.71–0.96)	0.67 (0.6–0.88)	0.67 (0.60–0.83)	0.89 (0.62–0.98)	0.63 (0.60–0.71)
SVM	1st	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	0.98 (0.90–1)	0.93 (0.81–1)	0.92 (0.71–1)	0.98 (0.88–1)	0.92 (0.86–0.96)
	2nd	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	0.98 (0.91–1)	0.94 (0.8–1)	0.94 (0.79–1)	0.98 (0.87–1)	0.94 (0.87–0.98)
	Det	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	0.97 (0.90–1)	0.91 (0.75–1)	0.91 (0.71–1)	0.97 (0.88–1)	0.91 (0.86–0.96)
	Det+1st	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	0.98 (0.92–1)	0.94 (0.83–1)	0.94 (0.79–1)	0.98 (0.91–1)	0.93 (0.88–0.98)
	Det+2nd	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	0.98 (0.92–1)	0.96 (0.84–1)	0.96 (0.83–1)	0.98 (0.92–1)	0.95 (0.91–0.99)
	Raw	0.98 (0.95–1)	0.95 (0.85–1)	0.95 (0.89–1)	0.98 (0.94–1)	0.94 (0.93–0.96)	0.94 (0.84–1)	0.83 (0.61–1)	0.83 (0.61–1)	0.95 (0.77–1)	0.83 (0.73–0.91)

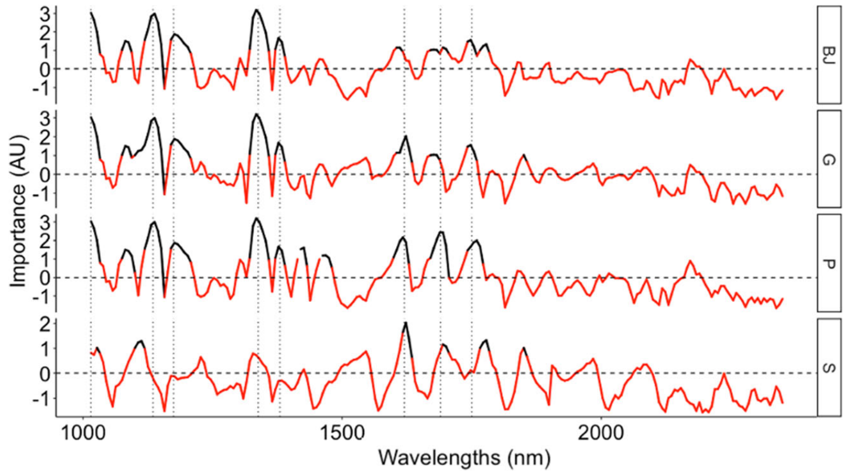

Fig. 2 Influence of bamboo shoot classification types on NIR spectra in the SVM model using Det+2nd derivative spectra. Black line: important spectral regions selected by the SVM model, and red line: less important than the black region

methods of NIR spectra could greatly improve the output performance of the PLSDA, SVM, and RF models, similar to the result reported by Qiu et al. (2018), who found that first and second derivatives with the Savitzky-Golay filter yielded a better classification accuracy (approximately 98%) using the PLSDA model for the detection of artificial aging of corn seeds compared with embryo and endosperm FT-NIR spectra. The second derivative could maximally obtain useful spectral variables for the prediction of extractive contents in *Eucalyptus bosistoana* in model calibration (Li and Altaner 2019). This study found that Det combined with the 2nd derivative yielded the best accuracy for the PLSDA, SVM, and RF models.

Eight wavelengths, i.e., 1015, 1135, 1175, 1338, 1380, 1620, 1690, and 1750 nm, in NIR spectra were considered the most important bands in this study on the classification of bamboo shoots. The bands around 1015, 1135, and 1175 nm were strongly related to the second overtone of C-H stretching vibration, which mainly represents terpenes (Ma et al. 2019). The first overtones of the C-H and O-H group stretching were mainly located at 1338 and 1380 nm, respectively (Schwanninger et al. 2011). The bands around 1620, 1690, and 1750 nm were associated with the first overtone of the C-H group, which is related to carbohydrate compounds, including fatty acids, starch, and lignin (Hourant et al. 2000; Qiu et al. 2018; Schwanninger et al. 2011). Carbohydrates and proteins are the most important components that strongly influence the quality of bamboo shoots. These components were found to strongly interact with NIR spectra in this study, which potentially proves that bamboo shoot species can be identified using NIR technology.

The results showed that the use of important variables from the NIR spectra, which were less than 16% of the full-length spectra, yielded a promising and reliable classification with a mean accuracy of 0.91, i.e., slightly less than that of the full-length spectra (0.95). NIR spectra contain much irrelevant information except for the useful chemical interacting band information, which will delay the processing time of model calibration and prediction for future use. Therefore, use of the most important extracted wavelength to build a similar and reliable model as full-length spectra could largely reduce the number of used spectral bands and improve the model performance (Workman Jr and Weyer 2012).

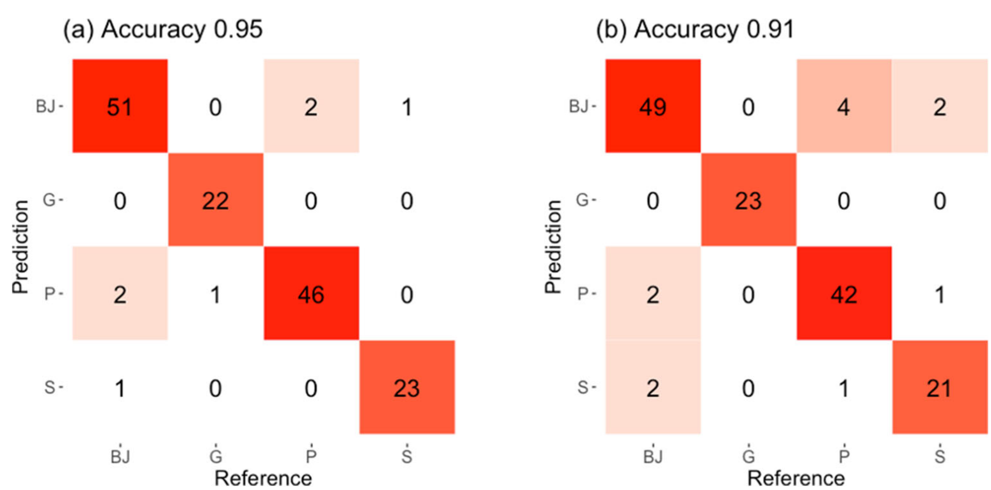

Fig. 3 The misclassification of four bamboo shoot species based on the validation of the SVM model using full-length (**a**) and selected spectral variables (**b**) based on Det + second-derivative NIR spectra

The cost and time of traditional methods for the large-scale quality assessment of bamboo shoots are limiting and will restrict the process of the bamboo shoot market. This method provides an advanced approach for bamboo shoot identification and allow for rapid measurement of a large number of samples.

Conclusions

In conclusion, use of NIR spectroscopy for the classification of different bamboo shoot species is feasible. The Det+2nd derivative pre-processing combined with the SVM method produced the highest classification accuracy and significantly avoided overfitting compared to the PLSDA and RF models. The most relevant spectra variables that were used for the SVM model could efficiently reduce the number of variables and yield a reliable accuracy as well as the full length of variables. The results indicate that the NIR technology could be a reliable and robust method for on-line bamboo shoot classification, and can support the farmers and industries in the monitoring of the high quality of bamboo shoots.

Funding This work was funded by the Chongqing Science and Technology Forestry Project: Demonstration of precise cultivation and efficient utilization of mixed bamboo shoots in a bamboo forest (No. 2019–2).

Compliance with Ethical Standards

Conflict of Interest Long Tong declares that he has no conflict of interest. Bin Li declares that he has no conflict of interest. Yanhui Geng declares that he has no conflict of interest. Lijie Chen declares that he has no conflict of interest. Yanjie Li declares that he has no conflict of interest. Ruishu Cao declares that he has no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of authors.

Informed Consent Informed consent not applicable.

References

Amodio ML, Ceglie F, Chaudhry MMA, Piazzolla F, Colelli G (2017) Potential of NIR spectroscopy for predicting internal quality and discriminating among strawberry fruits from different production systems. Postharvest Biol Technol 125:112–121

Ballabio D, Consonni V (2013) Classification tools in chemistry. Part 1: linear models. PLS-DA Anal Methods 5(16):3790–3798

Beghi R, Giovanelli G, Malegori C, Giovenzana V, Guidetti R (2014) Testing of a VIS-NIR system for the monitoring of long-term apple storage. Food Bioprocess Technol 7(7):2134–2143

Brenchley JM, Horchner U, Kalivas JH (1997) Wavelength selection characterization for NIR spectra. Appl Spectrosc 51(5):689–699

Caliari ÍP, Barbosa MH, Ferreira SO, Teófilo RF (2017) Estimation of cellulose crystallinity of sugarcane biomass using near infrared spectroscopy and multivariate analysis methods. Carbohydr Polym 158:20–28

Castillo R, Contreras D, Freer J, Ruiz J, Valenzuela S (2008) Supervised pattern recognition techniques for classification of Eucalyptus species from leaves NIR spectra. J Chil Chem Soc 53(4):1709–1713

Chen Q, Zhao J, Fang CH, Wang D (2007) Feasibility study on identification of green, black and oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM). Spectrochim Acta A Mol Biomol Spectrosc 66(3):568–574

Choudhury D, Sahu JK, Sharma G (2012) Value addition to bamboo shoots: a review. J Food Sci Technol 49(4):407–414

Couture JJ, Singh A, Rubert-Nason KF, Serbin SP, Lindroth RL, Townsend PA (2016) Spectroscopic determination of ecologically relevant plant secondary metabolites. Methods Ecol Evol 7(11): 1402–1412

Devos O, Ruckebusch C, Durand A, Duponchel L, Huvenne J-P (2009) Support vector machines (SVM) in near infrared (NIR) spectroscopy: focus on parameters optimization and model interpretation. Chemometrics Intell Lab Syst 96(1):27–33

Feng X, Zhao Y, Zhang C, Cheng P, He Y (2017) Discrimination of transgenic maize kernel using NIR hyperspectral imaging and multivariate data analysis. Sensors 17(8):1894

Fletcher RS, Reddy KN (2016) Random forest and leaf multispectral reflectance data to differentiate three soybean varieties from two pigweeds. Comput Electron Agric 128:199–206

Fu J (2001) Chinese moso bamboo: its importance. Bamboo 22(5):5–7

Gillon D, Houssard C, Joffre R (1999) Using near-infrared reflectance spectroscopy to predict carbon, nitrogen and phosphorus content in heterogeneous plant material. Oecologia 118(2):173–182

Grosser D, Liese W (1971) On the anatomy of Asian bamboos, with special reference to their vascular bundles. Wood Sci Technol 5(4):290–312

Gutierrez S, Tardaguila J, Fernandez-Novales J, Diago MP (2015) Support vector machine and artificial neural network models for the classification of grapevine varieties using a portable NIR spectrophotometer. PLoS One 10(11):e0143197

Hourant P, Baeten V, Morales MT, Meurens M, Aparicio R (2000) Oil and fat classification by selected bands of near-infrared spectroscopy. Appl Spectrosc 54(8):1168–1174

Inagaki T, Yonenobu H, Asanuma Y, Tsuchikawa S (2018) Determination of physical and chemical properties and degradation of archeological Japanese cypress wood from the Tohyamago area using near-infrared spectroscopy. J Wood Sci 64(4):347–355

Janssen JJ (2000) Designing and building with bamboo: International Network for Bamboo and Rattan inbar. Technical report, No. 20, Beijing, China, pp 2000

Kumar PS, Kumari U, Devi MP, Choudhary V, Sangeetha A (2017) Bamboo shoot as a source of nutraceuticals and bioactive compounds: a review. Indian J Nat Prod Resour 8(1):32–46

Li Y, Altaner CM (2019) Effects of variable selection and processing of NIR and ATR-IR spectra on the prediction of extractive content in *Eucalyptus bosistoana* heartwood. Spectrochim Acta A Mol Biomol Spectrosc 213:111–117

Li X, He Y (2008) Discriminating varieties of tea plant based on Vis/NIR spectral characteristics and using artificial neural networks. Biosyst Eng 99(3):313–321

Liaw A, Wiener M (2002) Classification and regression by randomForest. R news 2(3):18–22

Ma L, Peng Y, Pei Y, Zeng J, Shen H, Cao J, Qiao Y, Wu Z (2019) Systematic discovery about NIR spectral assignment from chemical structural property to natural chemical compounds. Sci Rep 9(1):1–17

Mancini M, Rinnan Å, Pizzi A, Mengarelli C, Rossini G, Duca D, Toscano G (2018) Near infrared spectroscopy for the discrimination between different residues of the wood processing industry in the pellet sector. Fuel 217:650–655

Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F (2018) e1071: Misc functions of the department of statistics, probability theory group (Formerly: E1071), TU Wien, https://CRAN.R-project.org/package=e1071. Accessed 6 May 2019

Min M, Lee WS, Kim YH, Bucklin RA (2006) Nondestructive detection of nitrogen in Chinese cabbage leaves using VIS–NIR spectroscopy. HortScience 41(1):162–166

Nicolai BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI, Lammertyn J (2007) Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. Postharvest Biol Technol 46(2):99–118

Porep JU, Kammerer DR, Carle R (2015) On-line application of near infrared (NIR) spectroscopy in food production. Trends Food Sci Technol 46(2):211–230

Press WH, Teukolsky SA (1990) Savitzky-Golay smoothing filters. Comput Phys 4(6):669–672

Qiu G, Lü E, Lu H, Xu S, Zeng F, Shui Q (2018) Single-kernel FT-NIR spectroscopy for detecting supersweet corn (*Zea mays* L. Saccharata Sturt) seed viability with multivariate data analysis. Sensors 18(4): 1010

R Core Team (2017) R: a language and environment for statistical computing. In R Foundation for Statistical Computing. Vienna, Austria

Ramirez JA, Posada JM, Handa IT, Hoch G, Vohland M, Messier C, Reu B (2015) Near-infrared spectroscopy (NIRS) predicts non-structural carbohydrate concentrations in different tissue types of a broad range of tree species. Methods Ecol Evol 6(9):1018–1025

Rinnan Å, Van Den Berg F, Engelsen SB (2009) Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends Anal Chem 28(10):1201–1222

Sampaio PS, Castanho A, Almeida AS, Oliveira J, Brites C (2020) Identification of rice flour types with near-infrared spectroscopy associated with PLS-DA and SVM methods. Eur Food Res Technol 246(3):527–537

Santos MB, Gomes AA, Vilar WT, Almeida P, Milani M, Nóbrega M, Medeiros EP, Galvão RK, Araújo MC (2014) Non-destructive NIR spectrometric cultivar discrimination of castor seeds resulting from breeding programs. J Braz Chem Soc 25(5):969–974

Satya S, Bal LM, Singhal P, Naik S (2010) Bamboo shoot processing: food quality and safety aspect (a review). Trends Food Sci Technol 21(4):181–189

Schwanninger M, Rodrigues JC, Fackler K (2011) A review of band assignments in near infrared spectra of wood and wood components. J Near Infrared Spectrosc 19(5):287–308

Stevens A, Ramirez-Lopez L (2014) An introduction to the prospectr package. https://cran.r-project.org/web/packages/prospectr/vignettes/prospectrintro.pdf. Accessed 20 June 2019

Sun Y, Gu X, Sun K, Hu H, Xu M, Wang Z, Tu K, Pan L (2017) Hyperspectral reflectance imaging combined with chemometrics and successive projections algorithm for chilling injury classification in peaches. LWT 75:557–564

Tankeu S, Vermaak I, Chen W, Sandasi M, Kamatou G, Viljoen A (2018) Hyperspectral imaging and support vector machine: a powerful combination to differentiate black cohosh (*Actaea racemosa*) from other cohosh species. Planta Med 84(06/07):407–419

Tsouvaltzis P, Babellahi F, Amodio ML, Colelli G (2020) Early detection of eggplant fruit stored at chilling temperature using different non-destructive optical techniques and supervised classification algorithms. Postharvest Biol Technol 159:111001

Vishwanathan S, Murty MN (2002) SSVM: a simple SVM algorithm. In proceedings of the 2002 international joint conference on neural networks. IJCNN'02 (cat. No. 02CH37290).IEEE 3: 2393-2398

Wang Y, Dong W, Kouba A (2016) Fast discrimination of bamboo species using VIS/NIR spectroscopy. J Appl Spectrosc 83(5):826–831

Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer, New York. http://had.co.nz/ggplot2/book. Accessed 15 July 2019

Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, Kenkel B, Benesty M, Lescarbeau R (2017) Caret: classification and regression training. R Package Version 6.0–81. http://CRAN.R-project.org/package=caret. Accessed 6 July 2019

Workman J Jr, Weyer L (2012) Practical guide and spectral atlas for interpretive near-infrared spectroscopy. CRC Press, Boca Raton

Xu F, Huang X, Dai H, Chen W, Ding R, Teye E (2014) Nondestructive determination of bamboo shoots lignification using FT-NIR with efficient variables selection algorithms. Anal Methods 6(4):1090–1095

Yang S, Han Y, Chang Y, Park J, Park Y, Chung H, Yeo H (2018) Classification of the hot air heat treatment degree of larch wood using a multivariate analysis of near-infrared spectroscopy. J Wood Sci 64(3):220–225

Zude M (2003) Non-destructive prediction of banana fruit quality using VIS/NIR spectroscopy. Fruits 58(3):135–142