

Article

Identification of Coffee Varieties Using Laser-Induced Breakdown Spectroscopy and Chemometrics

Chu Zhang ^{1,2}, Tingting Shen ^{1,2}, Fei Liu ^{1,2,*}  and Yong He ^{1,2} 

¹ College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, China; chuzh@zju.edu.cn (C.Z.); ttingshen@zju.edu.cn (T.S.); yhe@zju.edu.cn (Y.H.)

² Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture, Hangzhou 310058, China

* Correspondence: fliu@zju.edu.cn; Tel.: +86-571-8898-2825; Fax: +86-571-8898-2143

Received: 7 November 2017; Accepted: 27 December 2017; Published: 31 December 2017

Abstract: We linked coffee quality to its different varieties. This is of interest because the identification of coffee varieties should help coffee trading and consumption. Laser-induced breakdown spectroscopy (LIBS) combined with chemometric methods was used to identify coffee varieties. Wavelet transform (WT) was used to reduce LIBS spectra noise. Partial least squares-discriminant analysis (PLS-DA), radial basis function neural network (RBFNN), and support vector machine (SVM) were used to build classification models. Loadings of principal component analysis (PCA) were used to select the spectral variables contributing most to the identification of coffee varieties. Twenty wavelength variables corresponding to C I, Mg I, Mg II, Al II, CN, H, Ca II, Fe I, K I, Na I, N I, and O I were selected. PLS-DA, RBFNN, and SVM models on selected wavelength variables showed acceptable results. SVM and RBFNN models performed better with a classification accuracy of over 80% in the prediction set, for both full spectra and the selected variables. The overall results indicated that it was feasible to use LIBS and chemometric methods to identify coffee varieties. For further studies, more samples are needed to produce robust classification models, research should be conducted on which methods to use to select spectral peaks that correspond to the elements contributing most to identification, and the methods for acquiring stable spectra should also be studied.

Keywords: coffee varieties; laser-induced breakdown spectroscopy; discrimination; chemometrics; wavelength selection

1. Introduction

Coffee is one of the most popular beverages in the world. According to the International Coffee Organization (ICO), the estimated global coffee consumption in 2014 was 149.2 million bags (at 60 kg per bag) [1]. Coffee variety is one of the key factors that influences coffee quality. The identification of coffee beans has been studied with traditional laboratory-based chemical methods [2–4] and spectroscopic techniques [5–10]. Spectroscopic techniques have been widely studied due to their advantages of speed and few, or no sample preparation, and being both easy to operate and contactless. Near-infrared (NIR) spectroscopy, mid-infrared (MIR) spectroscopy, nuclear magnetic resonance (NMR) spectroscopy, and Raman spectroscopy have been used in the coffee industry to study coffee authentication, variety, producing area, and quality determination. These methods focus on the samples' chemical group and showed satisfactory results in the coffee industry [5–10].

Different samples have different kinds of elements with different contents. In many cases, element detection, such as element identification, element-content determination, and other related fields, is necessary. Laser-induced breakdown spectroscopy (LIBS) is a spectroscopic technique that identifies the elemental composition of samples [11]. LIBS analyzes the spectral emission from

laser-induced plasmas by focusing a pulse laser beam on the sample's surface. LIBS then identifies the spectral peaks and the spectral intensity corresponding to the element in the sample. LIBS can detect almost all elements, while simultaneously analyzing all those samples' elements that are above the detection limit. When compared with traditional element measurement methods such as atomic absorption spectrometry (AAS), inductively coupled plasma mass spectrometry (ICP-MS), and inductively coupled plasma optical emission spectrometry (ICP-OES), LIBS offers the advantages of little or no sample preparation, along with a fast, simple, and relatively low-cost analysis.

LIBS is very popular in element detection because of its ability to detect almost all of them. LIBS has been used for qualitative and quantitative element detection for soil [12,13], plants [14,15], food [16,17], minerals [18,19], and planets like Mars [20–22]. In recent years, studies have focused on the establishment of LIBS as an effective detecting technique within the coffee industry. Gondal et al. analyzed the LIBS spectrum of a coffee sample, revealing the presence of Mg, Ca, Al, Cu, Na, Ba, Br, Co, Cr, Ce, Mn, and Mo [23]; Nufiqurakhmah et al. used the LIBS technique to identify the elemental contents of coffee beans and selected Ca, W, Sr, Mg, and H to differentiate regular and luwak beans from the Arabica variant [24]; Ferreira et al. attempted to classify roasted and ground coffee with low-cost LIBS devices and obtained results that showed accuracies above 84% for ensemble methods in some spectral ranges [25]; Anggraeni et al. used the LIBS technique to differentiate green coffee beans from Arabica and Robusta variants, and they confirmed that Ca, W, Mg, Be, Na, and Sr were element identifiers of green coffee beans [26]. The above articles show that LIBS is a promising technique for detecting coffee products. However, their qualitative analysis methods lack optimization and the accuracy rates have yet to be improved.

The LIBS technique generates a large amount of data, including background and spectral peaks that are element-related. The simplest use of LIBS serves to identify elements by observing spectral peaks. Dealing with LIBS spectra by extracting useful information for further qualitative and quantitative analysis is far more complex [27–29]. Chemometric methods are widely used to handle spectral data in different spectroscopy techniques, and researchers are now trying to improve the accuracy of qualitative and quantitative analysis using LIBS spectra. The qualitative analysis of LIBS spectra focuses on identification and classification, and the quantitative analysis focuses on element-concentration determination [27–29]. By extracting useful information and building qualitative and quantitative models, chemometric methods can be used to reduce the errors generated by random factors and to reveal sample differences caused by different factors. Preprocessing LIBS spectra, which includes baseline correction, noise reduction, normalization, outlier removal, and overlapping-peaks identification, is important for the reduction of errors caused by random factors such as instruments, the environment, and the operation [27–34]. A series of supervised (partial least squares-discriminant analysis (PLS-DA), artificial neural network (ANN), support vector machine (SVM), etc.) and unsupervised (principal component analysis (PCA), hierarchical cluster analysis (HCA), etc.) chemometric methods have been widely introduced as ways of handling LIBS data for qualitative analysis [27–29]. To quantitatively determine element concentrations, univariate and multivariate methods are being introduced in order to build regression models [27–29]. Although chemometric methods have proven to be effective, their adoption in LIBS spectra analysis still poses a great challenge.

The objective of this study was to identify coffee varieties using LIBS with chemometric methods. Our specific objectives were: (1) to identify the main elements in coffee; (2) to reduce noise by wavelet transform; (3) to select those elements that contribute most to coffee-variety identification; and (4) to build calibration models for coffee-variety identification using chemometric methods.

2. Materials and Methods

2.1. Sample Preparation

We collected four coffee-bean cultivars from China, including Typica Arabica coffee from Yunnan Province, Catimor Arabica coffee from Yunnan Province, Fushan Robusta coffee from Hainan Province, and Xinglong Robusta coffee from Hainan Province. For each variety, we ground twenty coffee beans, which we then mixed thoroughly as a sample and stored in a plastic bag. The coffee powders were pressed into a tablet by a tablet-press machine (FY-24, SCJS, Tianjin, China). In total, 24 samples from each variety were prepared. The Typic Arabica coffee from Yunnan Province, Catimor Arabica coffee from Yunnan Province, Fushan Robusta coffee from Hainan Province, and Xinglong Robusta coffee from Hainan Province were assigned, respectively, the category values 1, 2, 3, and 4.

2.2. LIBS System

The experimental system used in this study consists of a Q-switch Nd:YAG nanosecond pulsed laser (Vlite-200, Beamtech, Beijing, China). A second harmonic laser (532 nm, pulse duration of 8 ns, beam diameter of 7 mm) was used to ignite the sample with the help of a plano-convex lens ($f = 50$ mm). A detection system, consisting of an Echelle spectrograph (ME5000, Andor, Belfast, UK) and an intensified charge-coupled device (ICCD) camera (DH334, Andor), was used to collect plasma emission spectra in the range of 230 to 880 nm. The samples were placed on a X-Y-Z translation stage.

2.3. LIBS Data Acquisition

Before measurement, 0.5 g of coffee powder from each sample was pressed into a 15-mm diameter pellet for 60 s at a pressure of 12 tons. In order to improve the signal-to-noise rate, the delay and integration times were optimized to 1.8 μ s and 5 μ s, respectively. The laser was fired with a pulse energy of 60 mJ at 1 Hz. The spot size of the focused laser beam was 600 μ m. For each sample, ten successive spectra were accumulated at each location, and 16 different locations were measured in ambient air with the help of the translation stage.

2.4. Chemometric Methods

We frequently used wavelet transform (WT) as a spectral-smoothing method. WT decomposes the original spectral data into its high-frequency and low-frequency parts. The noise was generally presented as the high-frequency part; we used WT for spectral-noise reduction by dealing with this high-frequency part via a threshold method. The preprocessed high-frequency and low-frequency parts were then reconstructed as the preprocessed spectra [35–37].

Principal component analysis is a widely used qualitative method in spectral data analysis. PCA transforms the original variables into new variables, which are linear combinations of the original variables. These new variables (called principal components) are orthogonal and ranked according to the explained variances. The first principal component (PC) explained most of the variances, and was followed by the second PC, the third PC, and so on. Generally, the first few PCs explained most of the variances. The score biplots, based on two different PCs, are commonly used to present the sample distributions in a score space [38].

Partial least squares-discriminant analysis is a supervised pattern recognition method based on partial least squares regression (PLSR). The PLSR explores the linear relationship between the spectral data and the corresponding features (physical, chemical, and category). PLSR transforms the original data into the new orthogonal variables (called latent variables (LVs)). The first few LVs that carried the most information were used for calibration. The PLS-DA uses the integer that represents the category instead of the chemical or physical features. The PLS-DA conducts the regression procedure; the prediction value of PLS-DA was a real number with decimals. A threshold value should be set to determine the category which the sample belongs to. In this study, the threshold value was set

to 0.5 [39], and leave-one-out cross validation was conducted in order to obtain an optimal number of LVs.

Radial basis function neural network (RBFNN) is a widely used 3-layer feed-forward neural network. It can approximate any nonlinear function. RBFNN has a good generalization ability and a fast convergence speed. RBFNN creates a direct connection between the input layer and the hidden layer while connecting the hidden layer and the output layer via output weights [40]. The optimal spread value should be determined for RBFNN. In this study, we built RBFNN models using spread values from 0.1 to 1 with a step of 0.1, and from 1 to 100 with a step of 1. The optimal spread value was determined by the RBFNN model with the highest classification accuracy.

A support vector machine is a supervised discriminant method. It maps original data into a higher dimension space, and constructs hyperplanes that have a maximum distance to the nearest sample of a given category. The samples were then classified via the hyperplanes. Kernel functions are important in establishing SVM models, and radial basis function (RBF) is a widely used kernel function. The penalty coefficient (C) and kernel function parameter (γ) should be determined [41]. In this study, a grid-search procedure was used to determine the optimal combination of C and γ . The search range for C and γ ranged from 2^{-8} to 2^8 . The optimal combination of C and γ was determined by the SVM model with the highest classification accuracy.

2.5. Model Evaluation and Software

The performances of the classification models were evaluated via the classification accuracy of the calibration and prediction sets, defined as the percentage of correctly classified samples taken from all the samples. The PLS-DA, RBFNN, and SVM models and the PCA analysis were conducted on the Matlab R2010b (The Math Works, Natick, MA, USA).

3. Results

3.1. LIBS Spectra Preprocessing

Once the LIBS spectra were acquired and ready to be analyzed, the spectral data needed to be preprocessed to minimize both the influence of noise and the variations caused by the matrix effects, the experimental conditions, the sample status, and the LIBS system.

Given that the spectral intensity of some wavelengths had negative values due to spectral noise, a simple baseline correction was applied to the spectral data in order to eliminate the negative values. After this baseline correction, the LIBS spectra were preprocessed by WT to reduce the noise. The LIBS spectra were generally disturbed by the noise caused by the environment, the sample, and the instruments. The reduction of spectra noise was meant to enhance the signal-to-noise ratio, which is important in spectral data analysis. WT was an efficient method for reducing noise and keeping the spectral features. Its use consisted in the decomposition of the original data and the reconstruction of the preprocessed data. In this study, we used WT with a wavelet basis function Daubechies 5 (db5) and a decomposition level 10, after trials. Given that the spectra in the range of 229.98–246.51 nm had no obvious spectral peaks, we only analyzed the 246.53–880.26 nm spectra (20,937 variables).

Figure 1 shows the raw and the preprocessed 312.02–322.02 nm spectra of a randomly selected spectrum, where a small peak occurred. We observed that the WT-preprocessed spectrum was much smoother than the raw spectrum. However, the preprocessed spectrum's peak intensities were slightly lower than the raw spectrum's. The results show that WT reduced the noise efficiently, as the peaks were kept.

An efficient method for us to deal with the variations caused by matrix effects and experimental conditions was to use normalization [22]. In this study, the LIBS spectra were normalized [22] after being preprocessed by WT. The total areas under the 246.53–880.26 nm spectrum were calculated, and each spectrum was normalized to the total area.

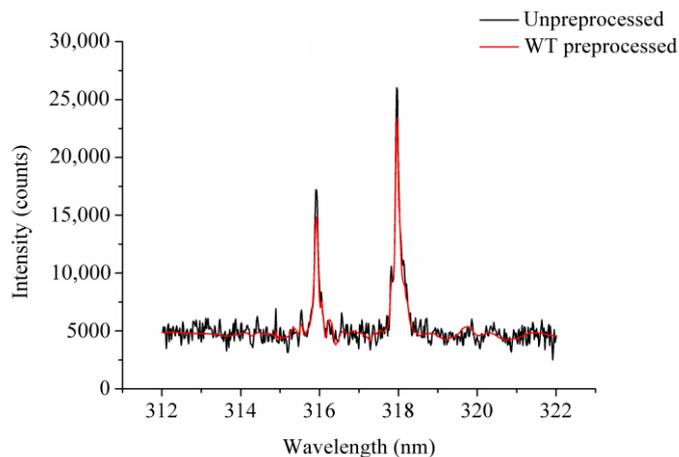


Figure 1. The un-preprocessed spectrum and the spectrum preprocessed by wavelet transform (WT).

Each sample's LIBS spectra were collected from 16 of the sample's sites. Due to the site-to-site variations, each site's spectrum showed differences, and some of the spectra with larger variations were treated as outliers [22]. To remove the spectra with larger variations, the relative standard deviation (RSD) of the C I 247.86 nm intensity was calculated for each sample. In our study, the rule was to keep 75% of the original spectra of each variety after removing outliers [22]. To identify outliers, 75% of the original spectra were randomly selected to compute the RSD, a procedure that was repeated 100,000 times. The spectra with minimum RSDs were kept for further analysis, and the remaining 25% were identified as outliers. Thus, a total of 1152 spectra (288 from each cultivar) were selected for further analysis after outlier removal.

3.2. Features of LIBS Spectra

Figure 2a shows the preprocessed LIBS spectra of one randomly selected sample from each variety. We were able to assign the main elements we observed in the LIBS spectra to C I, Mg I, Mg II, Al II, CN, H, Ca II, Fe I, K I, Na I, N I, and O I. We noted that there were big differences in the spectral intensities of different peaks, of up to 10^6 , and that some of the small peaks were not observable in Figure 2a. We found that the LIBS spectra of the four coffee varieties were similar, with differences in their spectral intensities. Figure 2b–e show some typical peaks of Mg II, Ca II, K I, and Fe I. In observing these peaks, we identified intensity differences for the different coffee varieties.

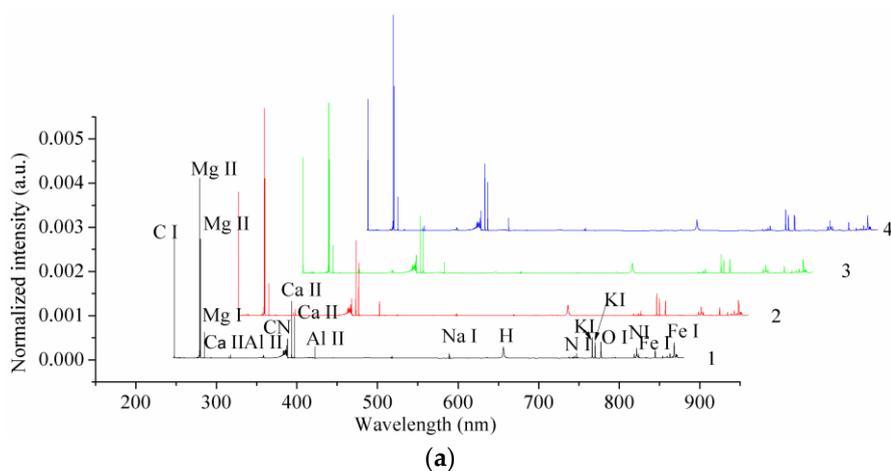


Figure 2. Cont.

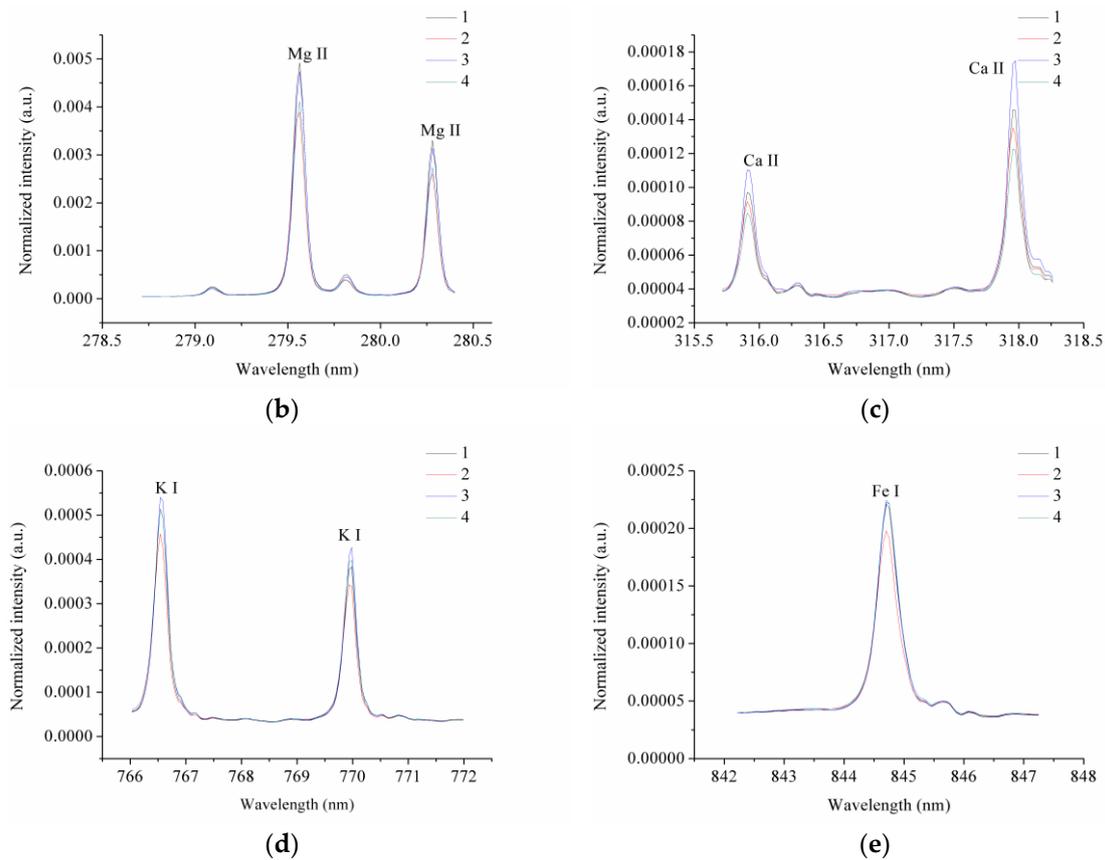


Figure 2. (a) The average laser-induced breakdown spectroscopy (LIBS) spectra of four different coffee varieties, and some typical peaks of (b) Mg II, (c) Ca II, (d) K I, and (e) Fe I. (a.u.: arbitrary unit).

3.3. Principal Component Analysis

A PCA analysis was performed on the LIBS spectra data (20,937 variables). The first 3 PCs explained 81.17%, 9.90%, and 2.76% of the total variance, respectively. The scores scatter plot of PC1 vs. PC2, PC1 vs. PC3, and PC2 vs. PC3 are shown in Figure 3. Here, we noticed that the samples from different varieties could group together, and that they overlapped and were difficult to distinguish from each other. The identification of coffee varieties required further analysis.

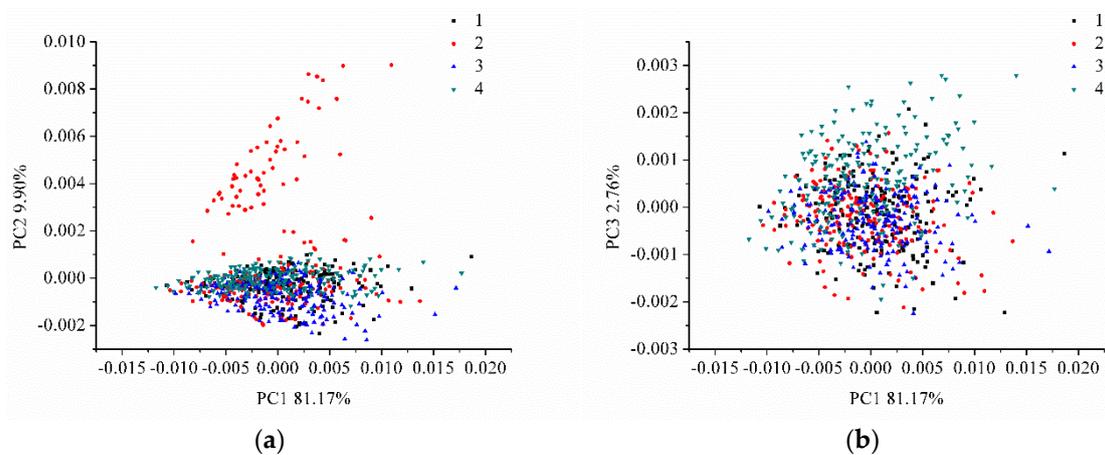


Figure 3. Cont.

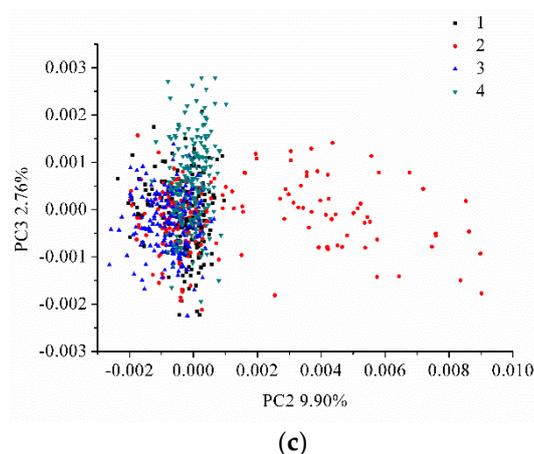


Figure 3. The scores scatter plot of (a) PC1 vs. PC2, (b) PC1 vs. PC3, and (c) PC2 vs. PC3, for the four coffee varieties. (PC: principal component).

3.4. Classification Models on the Full LIBS Data

The samples for each variety were randomly divided into calibration and prediction sets at a 2:1 ratio. We used PLS-DA, RBFNN, and SVM to build classification models on the full LIBS data (20,937 variables). PLS-DA, RBFNN, and SVM were able to deal efficiently with a large amount of data. The results of the PLS-DA, RBFNN, and SVM models are shown in Table 1.

Table 1. The results of the partial least squares-discriminant analysis (PLS-DA), radial basis function neural network (RBFNN) and support vector machine (SVM) models on the full LIBS data.

| Models | ^a LVs/Spread/(C, γ) | Calibration Accuracy (%) | | | | | Prediction Accuracy (%) | | | | |
|--------|--|--------------------------|-------|-------|-------|-------|-------------------------|-------|-------|-------|-------|
| | | 1 | 2 | 3 | 4 | Total | 1 | 2 | 3 | 4 | Total |
| PLS-DA | 20/*/* | 95.83 | 98.96 | 96.35 | 91.67 | 95.70 | 42.71 | 82.29 | 75.00 | 60.42 | 65.10 |
| RBFNN | */100/* | 100 | 100 | 100 | 100 | 100 | 82.29 | 66.67 | 80.21 | 98.96 | 82.03 |
| SVM | */*/(9.1896, 0.0039) | 100 | 100 | 100 | 100 | 100 | 68.75 | 80.21 | 90.63 | 94.79 | 83.59 |

^a LVs are the number of latent variables in the PLS-DA model; spread is the spread value in RBFNN; nodes are the number of nodes in the hidden layer of SVM; * means that no parameters exist for the given model.

The PLS-DA classification result was acceptable, with a classification accuracy of 95.70% and 65.10% for the calibration and prediction sets, respectively. The RBFNN and SVM models obtained better results than the PLS-DA model, yielding in both cases a classification accuracy of 100% for the calibration set and of over 80% for the prediction set. The results of the PLS-DA, RBFNN, and SVM models on the full LIBS spectra indicated that it was feasible to use LIBS spectra with chemometric methods as a way of identifying coffee varieties. Each cultivar's classification performance was different depending on the discriminant model used.

3.5. Analysis of Principal Component Analysis Loadings

The full LIBS spectra contained 20,937 variables; some of these (peaks) corresponded to the elements, while most were background information (useless information). Thus, it was important to select the useful variables for coffee-variety identification and to remove the useless background information. PCA loadings is an effective method for identifying useful information [42]. The first few PCs contained the most useful information, and loadings of each PC indicated the importance of the wavelengths. LIBS is a technique that identifies the elemental composition of the samples, as the qualitative and quantitative analyses of the LIBS spectra focus on the spectral emission of elements.

As presented in Figure 3, the first three PCs accounted for more than 93% of total variances; we used loadings of these first three PCs to identify important wavelengths.

Figure 4 shows the loading plots of the first three PCs using LIBS data from the calibration set. Wavelengths with higher loadings were selected and marked in Figure 4. We noted that, on the peaks, the loading plot correlated well with the LIBS spectra. We were able to assign the selected wavelength variables in Figure 4a to C I, Mg I, Mg II, Al II, CN, H, Ca II, Fe I, K I, N I, and O I, while the selected wavelength variables in Figure 4c were matched to Al II and Na I. Table 2 shows the selected spectral lines and their corresponding elements. The unmarked peaks in Figure 4b,c were on both sides of the corresponding peaks in Figure 4a due to the fact that there were data points (which were not peaks) on both sides of the major peaks with high-emission intensity.

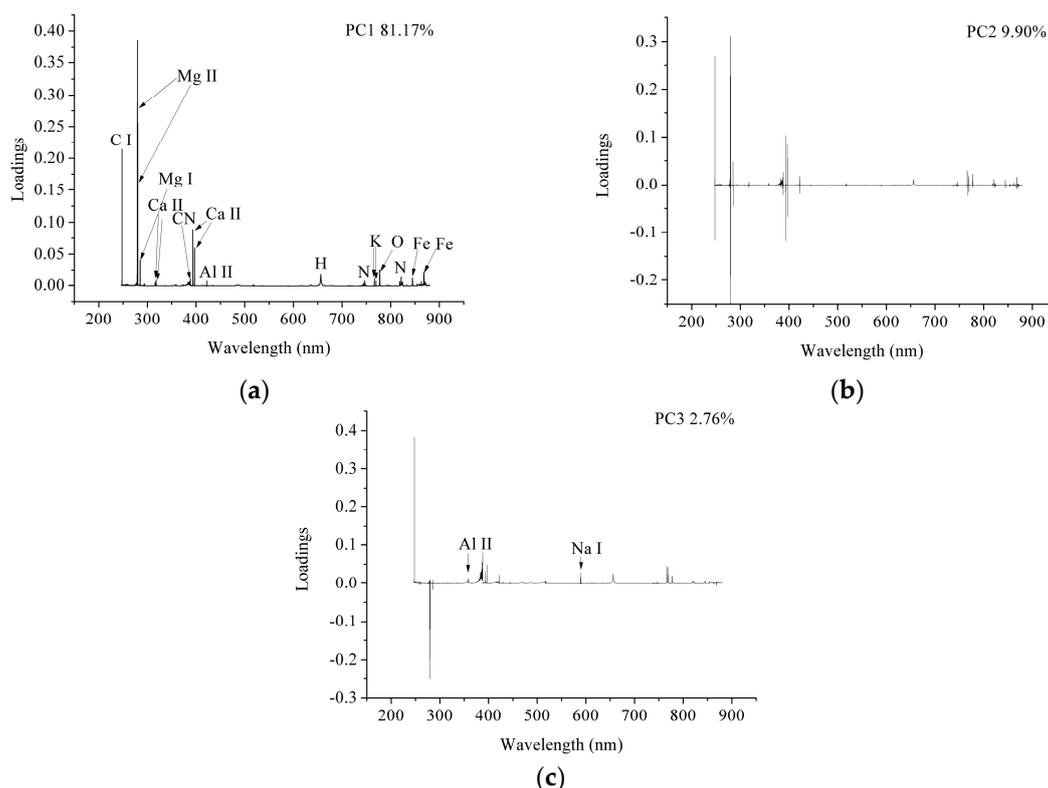


Figure 4. The loading plots of the first three PCs, (a) PC1, (b) PC2, and (c) PC3 for the 4 coffee varieties.

Table 2. The spectral lines selected by PCA loadings and the corresponding element assignments.

| Wavelength (nm) | Element |
|-----------------|---------|-----------------|---------|-----------------|---------|-----------------|---------|
| 247.87 | C I | 317.97 | Ca II | 422.68 | Al II | 769.98 | K I |
| 279.56 | Mg II | 358.57 | Al II | 588.99 | Na I | 777.33 | O I |
| 280.28 | Mg II | 388.32 | CN | 656.38 | H | 821.69 | N I |
| 285.22 | Mg I | 393.39 | Ca II | 746.92 | N I | 844.75 | Fe I |
| 315.93 | Ca II | 396.85 | Ca II | 766.54 | K I | 868.10 | Fe I |

3.6. Classification Models on Selected Wavelengths

We used PLS-DA, RBFNN, and SVM to build calibration models for the 20 selected wavelength variables. The results are shown in Table 3. The classification accuracies of the PLS-DA model were 56.77% and 57.81% for the calibration and prediction sets, with 9 LVs. These results were unsatisfactory. The RBFNN and SVM models performed better, both having a classification accuracy of over 80% for their prediction sets. The results indicated the feasibility of identifying coffee varieties by using

spectral peaks that correspond to the elements. Each variety's classification differed depending on the results of the given discriminant model.

Table 3. The results of the PLS-DA, RBFNN and SVM models on selected wavelength variables.

| Models | LVs/Spread/(C, γ) | Calibration Accuracy (%) | | | | | Prediction Accuracy (%) | | | | |
|--------|---------------------------|--------------------------|-------|-------|-------|-------|-------------------------|-------|-------|-------|-------|
| | | 1 | 2 | 3 | 4 | Total | 1 | 2 | 3 | 4 | Total |
| PLS-DA | 9/*/* | 29.30 | 89.06 | 56.77 | 51.56 | 56.77 | 14.58 | 97.92 | 66.67 | 52.08 | 57.81 |
| RBFNN | */0.2/* | 80.73 | 77.60 | 89.06 | 100 | 86.85 | 56.25 | 77.08 | 87.5 | 100 | 80.21 |
| SVM | */*/(84.4485, 0.0359) | 74.48 | 67.71 | 90.10 | 98.44 | 82.68 | 46.88 | 84.38 | 90.63 | 100 | 80.47 |

4. Discussion

The main goal of this study was to identify coffee varieties by using LIBS with chemometric methods. To achieve this goal, four different coffee varieties were collected and measured with a LIBS system developed in our laboratory. In this study, we conducted a simple procedure that included spectral preprocessing, outlier discarding, feature selection, and model establishment.

We used WT to reduce spectral noise. As the analysis of LIBS spectra involved the identification of spectral peaks representing the elements, it was quite important to suppress LIBS spectra noise while keeping the useful information. However, the LIBS spectra could simply be divided into two parts: the background information and the spectral peaks corresponding to the elements. For some of the peaks, the spectral intensity was much higher than the background information, while for others the spectral intensity was only a little higher. One should take care when reducing noise to avoid the elimination of unobvious spectral peaks; WT has proved to be efficient in studies that focused on LIBS [35,36], and it performed well in this study. However, we observed in Figure 1 that the intensity of the peaks dropped as well. The results showed that although WT can be used to reduce noise, studies should also focus on improvements that would aim to reduce noise while keeping peak intensities. Generally, moving average (MA) smoothing and Savitzky-Golay (SG) smoothing have been used as denoising methods in near-infrared spectroscopy and mid-infrared spectroscopy [43,44]. However, to our knowledge, very few studies have looked into their use as a means of preprocessing LIBS spectra. In a future study, we will examine the use of MA and SG to reduce LIBS-spectra noise. The acquisition of LIBS spectra was affected by many factors, and the spectra of different sampling points showed great variations. Preprocessing LIBS spectra could help reduce the influences caused by instruments, samples, and the atmosphere during spectra acquisition. Equally, methods to reduce influences during spectra acquisition should also be studied as a way of improving the identification of coffee varieties.

The full LIBS data contained 20,937 variables, which is quite substantial. According to previous studies [45,46], the coffee varieties could be determined via the concentrations of different elements. The full spectrum contained all of the sample's elemental information, which was used for coffee-variety identification. As mentioned above, the LIBS data contained the background information and the spectral peaks that corresponded to the elements. The background information may have interfered with the performance of the chemometric methods, increasing the computational task and model complexity. The PCA loading plots in Figure 4 also indicate that the spectral peaks had a larger influence than the background information. The performances of the classification models using full spectra pointed to the effectiveness of spectral line selection. The models that used full spectra performed slightly better than those that used the selected spectral lines. The results indicated that the background information influenced the model performances. The reason for this might be that the background information, such as matrix effects, was generated via samples during the spectra acquisition procedure; these samples might be different due to the different variety samples, and the classification models might therefore perform better with all the variables rather than with the spectral peaks. A similar phenomenon, where classification models using full spectra performed better, can be found in the literature [47]. Moreover, in other studies, when in order to build classification models

researchers manually selected spectral regions with spectral peaks instead of the full spectra and spectral peaks, they obtained satisfactory results. It should be noted that in these regions there was also background information [48–50]. On the other hand, Vors et al. have shown that optimal data preprocessing and variable-selection methods have a great influence on classification performances [51]. Some other studies have also indicated that these are also influenced by the selection of optimal variables [52–54]. It was therefore important to select variables or regions that were related to the elements that contribute the most to coffee-variety identification. This procedure helps significantly reduce the number of variables and the computation task, and simplifies the model. In our study, for example, the number of variables dropped from 20,937 to 20, resulting in a 99.90% reduction; the classification performances differed only slightly. The computation of the SVM and RBFNN models using the full-spectra calibration set ($768 \times 20,937$) takes more than 12 h on a personal computer (computer hardware: CPU: Intel Core (TM) i7-6700; RAM: 16 GB; Graphics card: NVIDIA GeForce GTX 750 Ti; 256 GB solid state disk). By using selected wavelengths, this computation task was significantly reduced from hours to minutes. Future studies should also explore the methods used to evaluate the importance of spectral peaks (elements) in LIBS data analysis.

The models using the selected wavelength variables performed worse than the models on the full LIBS data. The SVM and RBFNN models performed better than the corresponding PLS-DA models. The nonlinear neural networks (RBFNN and SVM) showed better results than the linear PLS-DA models. The reason might be attributable to the nonlinear effects in the LIBS data [55]. The overall results indicate that LIBS, combined with chemometric methods, could be used for coffee-variety identification. SVM and RBFNN showed some advantages over the linear PLS-DA models. However, studying more samples and more chemometric methods should help obtain better identification results.

5. Conclusions

We used LIBS in combination with chemometric methods, including discriminant models and variable selection, to identify coffee varieties. Preprocessing the LIBS spectra using WT effectively reduced spectral noise. Via PCA loadings we selected variables corresponding to C I, Mg I, Mg II, Al II, CN, H, Ca II, Fe I, K I, Na I, N I, and O I as important variables for identifying coffee varieties. PLS-DA, RBFNN, and SVM models, built by using full spectra and important variables, showed acceptable results, indicating the effectiveness of chemometrics in LIBS analysis. The overall results pointed to the feasibility of identifying coffee varieties by using LIBS with chemometric methods. For future studies, more samples are needed in order to produce more robust classification models. Additionally, chemical analyses of elements in coffee beans should be conducted for further proof of this method's effectiveness. Further research should also focus on the usage of different methods as a way of selecting the elements that contribute most to classification. This study provides an example of a qualitative analysis of LIBS data with chemometric methods for researchers conducting further studies. Our attempt to select the most contributed variables (elements) in LIBS spectra should also help with the development of on-line qualitative analysis systems.

Acknowledgments: This study was supported by National Natural Science Foundation of China (31671579), and China Postdoctoral Science Foundation (2017M610370).

Author Contributions: C.Z., F.L., and Y.H. conceived and designed the experiments; C.Z., T.S., and F.L. performed the experiments; C.Z., F.L., and Y.H. analyzed the data; C.Z., F.L., and Y.H. contributed reagents/materials/analysis tools; C.Z., T.S., F.L., and Y.H. wrote the manuscript; all authors have read and confirmed the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. International Coffee Organization. Trade Statistics Tables. Available online: <http://www.ico.org/> (accessed on 11 October 2017).

2. Fernandes, E.A.D.N.; Tagliaferro, F.S.; Azevedo-Filho, A.; Bode, P. Organic coffee discrimination with INAA and data mining/KDD techniques: New perspectives for coffee trade. *Accredit. Qual. Assur.* **2002**, *7*, 378–387. [[CrossRef](#)]
3. Alonsosalces, R.M.; Serra, F.; Reniero, F.; Héberger, K. Botanical and geographical characterization of green coffee (*Coffea arabica* and *Coffea canephora*): Chemometric evaluation of phenolic and methylxanthine contents. *J. Agric. Food Chem.* **2009**, *57*, 4224–4235. [[CrossRef](#)] [[PubMed](#)]
4. Wada, K.; Ohgama, S.; Sasaki, H.; Shimoda, M.; Osajima, Y. Classification of Various Trade Varieties of Coffee by Coupling of Sensory Data and Multivariate Analyses. *Agric. Biol. Chem.* **2014**, *51*, 1745–1752. [[CrossRef](#)]
5. Esteban-Díez, I.; González-Sáiz, J.M.; Sáenz-González, C.; Pizarro, C. Coffee varietal differentiation based on near infrared spectroscopy. *Talanta* **2007**, *71*, 221–229. [[CrossRef](#)] [[PubMed](#)]
6. Downey, G.; Briandet, R.; Wilson, R.H.; Kemsley, E.K. Near- and mid-infrared spectroscopies in food authentication: Coffee varietal identification. *J. Agric. Food Chem.* **1997**, *45*, 4357–4361. [[CrossRef](#)]
7. Huck, C.W.; Guggenbichler, W.; Bonn, G.K. Analysis of caffeine, theobromine and theophylline in coffee by near infrared spectroscopy (NIRS) compared to high-performance liquid chromatography (HPLC) coupled to mass spectrometry. *Anal. Chim. Acta* **2005**, *538*, 195–203. [[CrossRef](#)]
8. Kemsley, E.K.; Ruault, S.; Wilson, R.H. Discrimination between *Coffea arabica* and *Coffea canephora* variant *robusta* beans using infrared spectroscopy. *Food Chem.* **1995**, *54*, 321–326. [[CrossRef](#)]
9. Wei, F.; Furihata, K.; Hu, F.; Miyakawa, T.; Tanokura, M. Complex mixture analysis of organic compounds in green coffee bean extract by two-dimensional NMR spectroscopy. *Magn. Reson. Chem.* **2010**, *48*, 857–865. [[CrossRef](#)] [[PubMed](#)]
10. El-Abassy, R.M.; Donfack, P.; Materny, A. Discrimination between Arabica and Robusta green coffee using visible micro Raman spectroscopy and chemometric analysis. *Food Chem.* **2011**, *126*, 1443–1448. [[CrossRef](#)]
11. Pathak, A.K.; Kumar, R.; Singh, V.K.; Agrawal, R.; Rai, S.; Rai, A.K. Assessment of LIBS for Spectrochemical Analysis: A Review. *Appl. Spectrosc. Rev.* **2012**, *47*, 14–40. [[CrossRef](#)]
12. Kim, K.R.; Kim, G.; Kim, J.Y.; Park, K.; Kim, K.W. Kriging interpolation method for laser induced breakdown spectroscopy (LIBS) analysis of Zn in various soils. *J. Anal. At. Spectrom.* **2014**, *29*, 76–84. [[CrossRef](#)]
13. Unnikrishnan, V.K.; Nayak, R.; Aithal, K.; Kartha, V.B.; Santhosh, C.; Gupta, G.P.; Suri, B.M. Analysis of trace elements in complex matrices (soil) by Laser Induced Breakdown Spectroscopy (LIBS). *Anal. Methods* **2013**, *5*, 1294–1300. [[CrossRef](#)]
14. Rai, D.; Agrawal, R.; Kumar, R.; Rai, A.K.; Rai, G.K. Effect of Processing on Magnesium Content of Green Leafy Vegetables. *J. Appl. Spectrosc.* **2014**, *80*, 878–883. [[CrossRef](#)]
15. Chauhan, D.K.; Tripathi, D.K.; Rai, N.K.; Rai, A.K. Detection of Biogenic Silica in Leaf Blade, Leaf Sheath, and Stem of Bermuda Grass (*Cynodon dactylon*) Using LIBS and Phytolith Analysis. *Food Biophys.* **2011**, *6*, 416–423. [[CrossRef](#)]
16. Yao, M.; Huang, L.; Zheng, J.; Fan, S.; Liu, M. Assessment of feasibility in determining of Cr in Gannan Navel Orange treated in controlled conditions by laser induced breakdown spectroscopy. *Opt. Laser Technol.* **2013**, *52*, 70–74. [[CrossRef](#)]
17. Martelli, M.R.; Brygo, F.; Sadoudi, A.; Delaporte, P.; Barron, C. Laser-Induced Breakdown Spectroscopy and Chemometrics: A Novel Potential Method to Analyze Wheat Grains. *J. Agric. Food Chem.* **2010**, *58*, 7126–7134. [[CrossRef](#)] [[PubMed](#)]
18. Yaroshchuk, P.; Death, D.L.; Spencer, S.J. Comparison of principal components regression, partial least squares regression, multi-block partial least squares regression, and serial partial least squares regression algorithms for the analysis of Fe in iron ore using LIBS. *J. Anal. At. Spectrom.* **2012**, *27*, 92–98. [[CrossRef](#)]
19. Yuan, T.; Wang, Z.; Lui, S.L.; Fu, Y.; Li, Z.; Liu, J.; Ni, W. Coal property analysis using laser-induced breakdown spectroscopy. *J. Anal. At. Spectrom.* **2013**, *28*, 1045–1053. [[CrossRef](#)]
20. Wiens, R.C.; Maurice, S.; Lasue, J.; Forni, O.; Anderson, R.B.; Clegg, S.; Bender, S.; Blaney, D.; Barraclough, B.L.; Cousin, L.; et al. Pre-flight calibration and initial data processing for the ChemCam laser-induced breakdown spectroscopy instrument on the Mars Science Laboratory rover. *Spectrochim. Acta B* **2013**, *82*, 1–27. [[CrossRef](#)]
21. Stipe, C.B.; Guevara, E.; Brown, J.; Rossman, G.R. Quantitative laser-induced breakdown spectroscopy of potassium for in-situ geochronology on Mars. *Spectrochim. Acta B* **2012**, *70*, 45–50. [[CrossRef](#)]

22. Sobron, P.; Wang, A.; Sobron, F. Extraction of compositional and hydration information of sulfates from laser-induced plasma spectra recorded under Mars atmospheric conditions—Implications for ChemCam investigations on Curiosity rover. *Spectrochim. Acta B* **2012**, *68*, 1–16. [[CrossRef](#)]
23. Gondal, M.A.; Baig, U.; Dastageer, M.A.; Sarwar, M. Determination of elemental composition of coffee using UV-pulsed laser induced breakdown spectroscopy. *AIP Conf. Proc.* **2016**, *1742*.
24. Nufiqurakhmah, N.; Nasution, A.; Suyanto, H. Laser-Induced Breakdown Spectroscopy (LIBS) for spectral characterization of regular coffee beans and luwak coffee bean. In Proceedings of the Second International Seminar on Photonics, Optics, and Its Applications (ISPhOA 2016), Bali, Indonesia, 11 November 2016; pp. 101500M-1–101500M-7.
25. Ferreira, E.J.; Ferreira, E.C.; Delbem, A.C.B.; Milori, D.M.B.P. Ensemble of predictors and laser induced breakdown spectroscopy for certifying coffee. *Electron. Lett.* **2011**, *47*, 967–969. [[CrossRef](#)]
26. Anggraeni, K.; Nasution, A.; Suyanto, H. Recognition of spectral identifier from green coffee beans of arabica and robusta varieties using laser-induced breakdown spectroscopy. In Proceedings of the Second International Seminar on Photonics, Optics, and Its Applications (ISPhOA 2016), Bali, Indonesia, 11 November 2016; pp. 1015019-1–1015019-6.
27. Zhang, T.L.; Shan, W.U.; Tang, H.S.; Wang, K.; Duan, Y.X.; Hua, L.I. Progress of Chemometrics in Laser-induced Breakdown Spectroscopy Analysis. *Chin. J. Anal. Chem.* **2015**, *43*, 939–948. [[CrossRef](#)]
28. Haddad, J.E.; Canioni, L.; Bousquet, B. Good practices in LIBS analysis: Review and advices. *Spectrochim. Acta B* **2014**, *101*, 171–182. [[CrossRef](#)]
29. Galbács, G. A critical review of recent progress in analytical laser-induced breakdown spectroscopy. *Anal. Bioanal. Chem.* **2015**, *407*, 7537–7562. [[CrossRef](#)] [[PubMed](#)]
30. Yaroshchuk, P.; Eberhardt, J.E. Automatic correction of continuum background in Laser-induced Breakdown Spectroscopy using a model-free algorithm. *Spectrochim. Acta B* **2014**, *99*, 138–149. [[CrossRef](#)]
31. Pořízka, P.; Klus, J.; Prochazka, D.; Képeš, E.; Hrdlička, A.; Novotný, J.; Novotný, K.; Kaiser, J. Laser-Induced Breakdown Spectroscopy coupled with chemometrics for the analysis of steel: The issue of spectral outliers filtering. *Spectrochim. Acta B* **2016**, *123*, 114–120. [[CrossRef](#)]
32. Pořízka, P.; Klus, J.; Hrdlička, A.; Vrabel, J.; Škarková, P.; Prochazka, D.; Novotný, J.; Novotný, K.; Kaiser, J. Impact of Laser-Induced Breakdown Spectroscopy data normalization on multivariate classification accuracy. *J. Anal. At. Spectrom.* **2017**, *32*, 277–288. [[CrossRef](#)]
33. Zorov, N.B.; Gorbatenko, A.A.; Labutin, T.A.; Popov, A.M. A review of normalization techniques in analytical atomic spectrometry with laser sampling: From single to multivariate correction. *Spectrochim. Acta B* **2010**, *65*, 642–657. [[CrossRef](#)]
34. Castro, J.P.; Pereirafilho, E.R. Twelve different types of data normalization for the proposition of classification, univariate and multivariate regression models for the direct analyses of alloys by laser-induced breakdown spectroscopy (LIBS). *J. Anal. At. Spectrom.* **2016**, *31*, 2005–2014. [[CrossRef](#)]
35. Schlenke, J.; Hildebrand, L.; Moros, J.; Laserna, J.J. Adaptive approach for variable noise suppression on laser-induced breakdown spectroscopy responses using stationary wavelet transform. *Anal. Chim. Acta* **2012**, *754*, 8–19. [[CrossRef](#)] [[PubMed](#)]
36. Zhang, B.; Sun, L.; Yu, H.; Xin, Y.; Cong, Z. Wavelet denoising method for laser-induced breakdown spectroscopy. *J. Anal. At. Spectrom.* **2013**, *28*, 1884–1893. [[CrossRef](#)]
37. Fu, X.; Yan, G.; Chen, B.; Li, H. Application of wavelet transforms to improve prediction precision of near infrared spectra. *J. Food Eng.* **2005**, *69*, 461–466. [[CrossRef](#)]
38. Dong, W.; Ni, Y.; Kokot, S. A Near-Infrared Reflectance Spectroscopy Method for Direct Analysis of Several Chemical Components and Properties of Fruit, for Example, Chinese Hawthorn. *J. Agric. Food Chem.* **2013**, *61*, 540–546. [[CrossRef](#)] [[PubMed](#)]
39. Liu, F.; Yusuf, B.L.; Zhong, J.; Feng, L.; He, Y.; Wang, L. Variety Identification of Rice Vinegars Using Visible and Near Infrared Spectroscopy and Multivariate Calibrations. *Int. J. Food Prop.* **2011**, *14*, 1264–1276. [[CrossRef](#)]
40. Fidêncio, P.H.; Poppi, R.J.; Andrade, J.C.D. Determination of organic matter in soils using radial basis function networks and near infrared spectroscopy. *Anal. Chim. Acta* **2002**, *453*, 125–134. [[CrossRef](#)]
41. Burges, C.J.C. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167. [[CrossRef](#)]

42. Zhang, C.; Jiang, H.; Liu, F.; He, Y. Application of Near-Infrared Hyperspectral Imaging with Variable Selection Methods to Determine and Visualize Caffeine Content of Coffee Beans. *Food Bioprocess Technol.* **2017**, *10*, 213–221. [[CrossRef](#)]
43. Gómez, A.H.; He, Y.; Pereira, A.G. Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques. *J. Food Eng.* **2006**, *77*, 313–319. [[CrossRef](#)]
44. Meza-Márquez, O.G.; Gallardo-Velázquez, T.; Osorio-Revilla, G. Application of mid-infrared spectroscopy with multivariate analysis and soft independent modeling of class analogies (SIMCA) for the detection of adulterants in minced beef. *Meat Sci.* **2010**, *86*, 511–519. [[CrossRef](#)] [[PubMed](#)]
45. Bertrand, B.; Villarreal, D.; Laffargue, A.; Posada, H.; Lashermes, P.; Dussert, S. Comparison of the Effectiveness of Fatty Acids, Chlorogenic Acids, and Elements for the Chemometric Discrimination of Coffee (*Coffea arabica* L.) Varieties and Growing Origins. *J. Agric. Food Chem.* **2008**, *56*, 2273–2280. [[CrossRef](#)] [[PubMed](#)]
46. Martín, M.J.; Pablos, F.; González, A.G. Characterization of green coffee varieties according to their metal content. *Anal. Chim. Acta* **1998**, *358*, 177–183. [[CrossRef](#)]
47. De Lucia, F.C., Jr.; Gottfried, J.L. Influence of variable selection on partial least squares discriminant analysis models for explosive residue classification. *Spectrochim. Acta B* **2011**, *66*, 122–128. [[CrossRef](#)]
48. Moncayo, S.; Manzoor, S.; Navarro-Villoslada, F.; Caceres, J.O. Evaluation of supervised chemometric methods for sample classification by Laser Induced Breakdown Spectroscopy. *Chemom. Intell. Lab. Syst.* **2015**, *146*, 354–364. [[CrossRef](#)]
49. Kong, H.; Sun, L.; Hu, J.; Xin, Y.; Cong, Z. Selection of Spectral Data for Classification of Steels Using Laser-Induced Breakdown Spectroscopy. *Plasma Sci. Technol.* **2015**, *17*, 964–970. [[CrossRef](#)]
50. Moncayo, S.; Rosales, J.D.; Izquierdohornillos, R.; Anzano, J.; Caceres, J.O. Classification of red wine based on its protected designation of origin (PDO) using Laser-induced Breakdown Spectroscopy (LIBS). *Talanta* **2016**, *158*, 185–191. [[CrossRef](#)] [[PubMed](#)]
51. Vors, E.; Tchepidjian, K.; Sirven, J.B. Evaluation and optimization of the robustness of a multivariate analysis methodology for identification of alloys by laser induced breakdown spectroscopy. *Spectrochim. Acta B* **2016**, *117*, 16–22. [[CrossRef](#)]
52. Tian, Y.; Yan, C.; Zhang, T.; Tang, H.; Li, H.; Yu, J.; Bernard, J.; Cheney, L.; Martin, S.; Delepine-Gilon, N. Classification of wines according to their production regions with the contained trace elements using laser-induced breakdown spectroscopy. *Spectrochim. Acta B* **2017**, *135*, 91–101. [[CrossRef](#)]
53. Xu, L.; Liang, L.; Zhang, T.; Tang, H.; Wang, K.; Li, H. A method of improving classification precision based on model population analysis of steel material for laser-induced breakdown spectroscopy. *Anal. Methods* **2014**, *6*, 8374–8379. [[CrossRef](#)]
54. Putnam, R.A.; Mohaidat, Q.I.; Daabous, A.; Rehse, S.J. A comparison of multivariate analysis techniques and variable selection strategies in a laser-induced breakdown spectroscopy bacterial classification. *Spectrochim. Acta B* **2013**, *87*, 161–167. [[CrossRef](#)]
55. Stipe, C.B.; Hensley, B.D.; Boersema, J.L.; Buckley, S.G. Laser-induced breakdown spectroscopy of steel: A comparison of univariate and multivariate calibration methods. *Appl. Spectrosc.* **2010**, *64*, 154–160. [[CrossRef](#)] [[PubMed](#)]

