foods

MDPI

Article

Application of Near-Infrared Hyperspectral Imaging with Machine Learning Methods to Identify Geographical Origins of Dry Narrow-Leaved Oleaster (*Elaeagnus angustifolia*) Fruits

Pan Gao [1,2,†], Wei Xu [3,4,†], Tianying Yan [1,2], Chu Zhang [5,6], Xin Lv [2,3] and Yong He [5,6,*]

1 College of Information Science and Technology, Shihezi University, Shihezi 832000, China; gp_inf@shzu.edu.cn (P.G.); yantianying@163.com (T.Y.)
2 Key Laboratory of Oasis Ecology Agriculture, Shihezi University, Shihezi 832003, China; lxshz@126.com
3 College of Agriculture, Shihezi University, Shihezi 832003, China; xu_wei082@163.com
4 Xinjiang Production and Construction Corps Key Laboratory of Special Fruits and Vegetables Cultivation Physiology and Germplasm Resources Utilization, Shihezi 832003, China
5 College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China; chuzh@zju.edu.cn
6 Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture and Rural Affairs, Hangzhou 310058, China
* Correspondence: yhe@zju.edu.cn; Tel.: +86-571-88982143
† These two authors contributed equally to this manuscript.

Received: 21 October 2019; Accepted: 23 November 2019; Published: 27 November 2019

Abstract: Narrow-leaved oleaster (*Elaeagnus angustifolia*) fruit is a kind of natural product used as food and traditional medicine. Narrow-leaved oleaster fruits from different geographical origins vary in chemical and physical properties and differ in their nutritional and commercial values. In this study, near-infrared hyperspectral imaging covering the spectral range of 874–1734 nm was used to identify the geographical origins of dry narrow-leaved oleaster fruits with machine learning methods. Average spectra of each single narrow-leaved oleaster fruit were extracted. Second derivative spectra were used to identify effective wavelengths. Partial least squares discriminant analysis (PLS-DA) and support vector machine (SVM) were used to build discriminant models for geographical origin identification using full spectra and effective wavelengths. In addition, deep convolutional neural network (CNN) models were built using full spectra and effective wavelengths. Good classification performances were obtained by these three models using full spectra and effective wavelengths, with classification accuracy of the calibration, validation, and prediction set all over 90%. Models using effective wavelengths obtained close results to models using full spectra. The performances of the PLS-DA, SVM, and CNN models were close. The overall results illustrated that near-infrared hyperspectral imaging coupled with machine learning could be used to trace geographical origins of dry narrow-leaved oleaster fruits.

Keywords: narrow-leaved oleaster fruits; near-infrared hyperspectral imaging; geographical origin; convolutional neural network; effective wavelengths

1. Introduction

Narrow-leaved oleaster (*Elaeagnus angustifolia*) is a shrub-like plant of Elaeagnus, which is widely distributed from the Mediterranean region to the northern hemisphere, including in northern Russia and northwestern China. Narrow-leaved oleaster fruits contain a variety of functional health components; in particular, they contain polysaccharides, phenolic acids, and flavonoids. Therefore, narrow-leaved

oleaster fruits, as a traditional medicine, are used to treat many diseases in nations and countries from Central Asia to West Asia. As a medicine and food, the fruit of narrow-leaved oleaster fruits is not only a raw material for food industry processing but also a raw material for functional food and new drugs [1–11]. It has good prospects for development and utilization in arid and semi-arid regions of Northwest China. Its unique habitat environment and long history of planting have produced unique qualities of narrow-leaved oleaster fruits in different producing areas. The qualities of narrow-leaved oleaster fruits are different depending on their place of origin, so it is urgent to establish effective methods for identification of the place of origin of narrow-leaved oleaster fruits.

At present, different scholars have isolated the bioactive components of narrow-leaved oleaster fruits [12], studied the physical and chemical properties and antioxidant properties of narrow-leaved oleaster fruits [13], used Gas Chromatography-Mass Spectrometer (GC-MS) to analyze the components of narrow-leaved oleaster fruit oil [14], and studied the diseases of narrow-leaved oleaster fruits [15]. However, there have been few studies on differentiation of the origins of narrow-leaved oleaster fruits. It is feasible to differentiate narrow-leaved oleaster fruits from different producing areas by synthesizing external morphological and microscopic characteristics and physicochemical identification of fruit powder. Manual sorting has many drawbacks, such as involving monotonous work and strong subjectivity, and being time-consuming and difficult to quantify. Physical and chemical index testing is destructive, and requires complicated sample pretreatment, a long detection cycle, and so on. It also has higher professional requirements for testers. These methods are time-consuming and laborious and cannot achieve the goal of fast and non-destructive classification. In view of the drawbacks of traditional detection methods, many applications use hyperspectral imaging for non-destructive detection due to its advantages of non-destructive, rapid, and accurate measurement, which has broad prospects.

Near-infrared hyperspectral imaging is a chemical analysis tool that can detect different absorption frequencies of specific molecules in substances. Near-infrared hyperspectral imaging can acquire spectral and image information of samples simultaneously. It can obtain comprehensive spectral information of samples. It has the characteristics of fastness and high accuracy. Near-infrared hyperspectral imaging has been widely used in geographical origins and variety identification of food [16]. C. Ru et al. used the hyperspectral imaging method of spectral image fusion in the range of visible and near-infrared (VNIR) and shortwave infrared (SWIR) to classify the geographical origin of Rhizoma Atractylodis Macrocephalae [17]. A. Noviyanto et al. used hyperspectral imaging and machine learning to distinguish honey botanical origins [18]. S. Minaei et al. used visible-near-infrared (VIS-NIR) hyperspectral imaging combined with a machine learning algorithm to predict honey floral origins [19]. M. Puneet et al. used near-infrared hyperspectral imaging to identify six different tea products [20]. Our research team has used near-infrared hyperspectral imaging for varietal and geographical origin identification of agricultural and food materials. C. Zhang et al. used near-infrared hyperspectral imaging to identify coffee bean varieties from different locations [21]. W. Yin et al. used near-infrared hyperspectral imaging to identify geographical origins of Chinese wolfberries [22]. S. Zhu et al. used near-infrared hyperspectral imaging to identify cotton seed varieties [23]. These researchers obtained good performances and illustrated the feasibility of using near-infrared hyperspectral imaging to identify the varietal and geographical origin of agricultural and food materials.

In this study, a near-infrared hyperspectral imaging system covering the spectral range of 874–1734 nm was used. This spectral range is related to various chemical compounds. Researchers have used hyperspectral imaging at this spectral range to obtain good performances for determining contents of protein [24], oil [25], water [26], total iron-reactive phenolics, anthocyanins and tannins [27], and flavanol [28], etc. Previous studies have shown that near-infrared hyperspectral imaging can achieve target classification, but there is no relevant research on the place of origin classification of dry narrow-leaved oleaster fruits. The main purpose of this study was to detect the geographical origin of dry narrow-leaved oleaster fruits based on near-infrared hyperspectral imaging technology,

combined with characteristic wavelength selection and machine learning algorithms, including deep learning, providing theoretical methods and a basis for distinguishing the different producing areas of narrow-leaved oleaster fruits.

2. Materials and Methods

2.1. Sample Preparation

Dry narrow-leaved oleaster fruits from three different geographical origins, including Miqin County, Gansu province (Gansu), China (103°4′48″ E, 38°37′12″ N); Zhongwei City, Ningxia Hui Autonomous Region (Ningxia), China (105°10′48″ E, 37°30′36″ N); and Aksu City, Xinjiang Uygur Autonomous Region (Xinjiang), China (80°17′24″ E, 41°9′00″ N), were collected. For each geographical origin, fully matured fruits were harvested in October 2018 and air-dried for consumption and trade. For each geographical origin, intact, clean, and dry narrow-leaved oleaster fruits were collected for hyperspectral image acquisition. In total, 1105, 1205, and 962 intact fruits were obtained from Gansu, Ningxia, and Xinjiang, respectively. The convolutional neural network (CNN) was trained with an independent validation set. To build discriminant models, the samples were randomly split into calibration, validation, and prediction sets. There were 539, 602, and 481 samples from Gansu, Ningxia, and Xinjiang in the calibration set, 291, 303, and 241 samples from Gansu, Ningxia, and Xinjiang in the validation set, and 275, 300, and 240 samples from Gansu, Ningxia, and Xinjiang in the prediction set, respectively. Samples of each geographical origin for hyperspectral imaging acquisition are placed and presented in Figure 1.

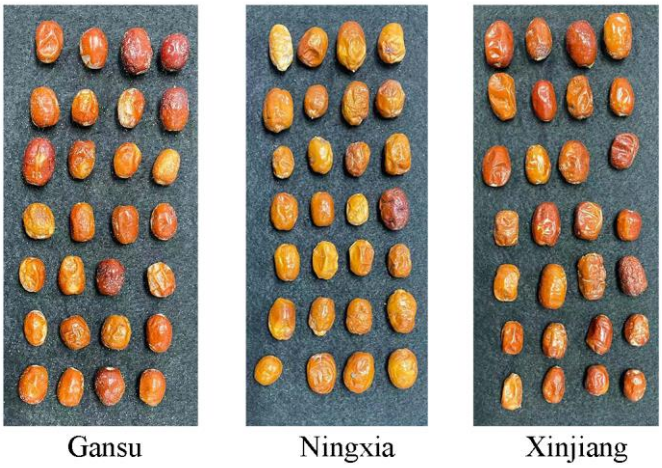

Figure 1. Samples of each geographical origin for hyperspectral imaging acquisition.

2.2. Hyperspectral Image Acquisition and Correction

A near-infrared hyperspectral imaging system was used to acquire hyperspectral images of single narrow-leaved oleaster fruits. This hyperspectral imaging system consisted of four major modules, including an imaging module, an illumination module, a sample motion module, and a software module. The imaging module consisted of an imaging spectrograph (ImSpector N17E, Spectral Imaging Ltd., Oulu, Finland) coupled with an InGaAs camera (Xeva 992, Xenics Infrared Solutions, Leuven, Belgium). The spectral range of the hyperspectral imaging system was 874–1734 nm, the spectral resolution 5 nm, and the number of wavebands 256. The lens for the camera was OLES22 (Spectral Imaging Ltd., Oulu, Finland). The illumination module had a 3900 light source (Illumination Technologies Inc., New York, NY, USA). The sample motion module was formed by an IRCP0076 electric displacement table (Isuzu Optics Corp., Taiwan, China) and samples were placed in the motion platform for line-scan. The software module was used to control the image acquisition and motion platform. The structure of the acquired hyperspectral image was able to be expressed as 320 pixels ×

L pixels × 256 (wavebands), where 320 pixels was the width of the image, the number 256 was the number of wavebands, and L pixels was the length of the image. L was manually determined during the image acquisition to ensure all samples in one plate were covered in one image.

The image quality, which was determined by the distance between the sample and the lens, the moving speed of the motion platform, and the camera exposure time, was determined by setting these parameters as 12.6 cm, 11 mm/s, and 3000 μs, respectively. In this study, intact narrow-leaved oleaster fruits were placed separately on a black plate for image acquisition. For each image, a random number of fruits was placed there (as shown in Figure 1), and there were at least twenty fruits in an image. During image acquisition, the imaging conditions and system parameters always remained. After image acquisition, the raw hyperspectral images were corrected into reflectance images according to the equation

$$I_c = \frac{I_r - I_d}{I_w - I_d}, \quad (1)$$

where I_c is the corrected image, I_r is the raw original image, I_d is the dark reference image and I_w is the white reference image.

2.3. Spectral Data Extraction

After image correction, spectral data were extracted from each narrow-leaved oleaster fruit. The hyperspectral imaging system collected reflectance spectra of the samples, and reflectance spectra were used for analysis in this study. Each single narrow-leaved oleaster fruit was defined as a region of interest (ROI). A binary image was formed of each hyperspectral image by binarizing the gray-scale image at 1119 nm, in which the narrow-leaved oleaster fruits region was '1' and the background region was '0'. The binary image was then applied to the gray-scale images at each gray-scale image to remove background information. Considering that obvious noises existed at the beginning and end of the spectra, only spectra in the range 975–1646 nm (waveband numbers 31 to 230) were studied, resulting in 200 wavelength variables in the spectral range. Pixel-wise spectra were preprocessed by wavelet transform (wavelet function Daubechies 6 with decomposition level 3) to reduce random noise and area normalization to reduce the influence of sample shape. Pixel-wise spectra within one narrow-leaved oleaster fruit were averaged to represent the sample.

2.4. Data Analysis Methods

2.4.1. Principal Component Analysis

Principal component analysis (PCA) is a widely used qualitative analysis and feature extraction method for spectral data analysis. PCA projects the original spectral data to some new principal component variables (PCs) through linear transformation. Each principal component is linearly combined with the original data. The PCs are ranked by the explained variance. The first PC (PC1) explains the largest of the total variance, followed by PC2 and PC3 and so on. In general, the first few PCs could explain most of the total variance and these few principal components with the largest variance could reflect the data information. In general, the scores of scatter plots which are obtained by projecting scores of one PC onto another PC are used to explore clusters of samples from different classes. In this study, PCA was used to explore qualitative discrimination of narrow-leaved oleaster fruit samples from Gansu, Ningxia, and Xinjiang.

2.4.2. Partial Least Squares Discriminant Analysis

The partial least squares discriminant analysis (PLS-DA) algorithm is based on the PLS regression model to discriminate the target, where the variables in the X block (spectral data) are related to the category values corresponding to the classes contained in the Y vector [29–35]. The integer values are assigned to each class. The category values can be assigned as real integer numbers or they can be

formed by dummy variables (0 and 1). PLS regression is firstly conducted on X and Y and the decimal prediction results are transformed into category values according to certain rules.

2.4.3. Support Vector Machine

The support vector machine (SVM) system has been widely applied in statistics, especially for classification. The main idea of SVM is to find the most distinguishable hyperplane by maximizing the margin between the closest points in each class [34–38]. By choosing and optimizing parameters such as penalty factor and kernel function, the discriminant model established by small data samples can still produce small errors for independent test sets. In this paper, the parameter penalty coefficient C of SVM model was searched, and the optimum range was 10^{-8} to 10^{8}. The kernel function was a radial basis function (RBF) and the searching range of the width of the kernel function (g) was 10^{-8} to 10^{8}.

2.4.4. Convolutional Neural Network

The convolutional neural network has been proved as a data processing method with high efficiency and high performance for hyperspectral data analysis due to its ability to aid automatic feature learning [39]. In this study, a simplified CNN architecture based on the model proposed in [40] was designed for narrow-leaved oleaster fruit discrimination.

Figure 2 shows the CNN architecture used in this research. It consisted of two main parts. The first part included two one-dimensional convolution layers (Conv1D, represented by a box with a green background), each of which having been followed by a ReLU activation (yellow box), a one-dimension MaxPooling layer (MaxPool1D, blue box) and a batch-normalization (white box) process. The other part included a fully connected network which was constructed by three Dense layers (light red box) and a SoftMax layer (gray box). The numbers of kernels in the convolution layers were 64 and 32, respectively, with a kernel size of 3 and stride of 1 without padding. MaxPooling layers were configured with a pool size of 2 and stride of 2. The numbers of neurons in the Dense layers were defined as 512, 128, and 3, in order. The first two Dense layers were activated by the ReLU function and followed by a batch-normalization process.

The training procedure was implemented by minimizing the SoftMax Cross Entropy Loss using a stochastic gradient descent (SGD) algorithm. The learning rate was optimized and set as 0.0005. The batch size was set as 400. The train epoch was defined as 400.

2.4.5. Optimal Wavelength Selection

Extracted spectra data contain redundant and collinear information, and some of the wavelengths are uninformative. These uninformative wavelengths may result in unstable calibrations. Moreover, a large number of wavelengths for calibration may result in a complex model structure. Selecting the most informative wavelengths is an important step for further multivariate analysis.

In this study, second derivative spectra were used to select the optimal wavelengths for narrow-leaved oleaster fruits. The second derivative is a widely used spectral preprocessing method which can highlight spectral peaks and suppress background information. In second derivative spectra, the background information is quite small and close to zero, and the positive and negative peaks with greater differences among different categories of samples are manually selected as optimal wavelengths [41].

2.5. Software and Model Evaluation

In this study, PCA, PLS-DA, and SVM were executed on a Matlab R2014b (The Math Works, Natick, MA, USA), the second derivate was conducted on Unscrambler 10.1 (CAMO AS, Oslo, Norway), and the CNN model was performed on Python 3 and MXNET framework (Amazon, Seattle, WA, USA). PCA and PLS-DA was computed using leave-one-out cross validation, SVM was computed using five-fold cross validation, and CNN was computed using an independent validation set. Model

performances were evaluated by their classification accuracy, which was calculated as the ratio of the number of correctly classified samples to the total number of samples.

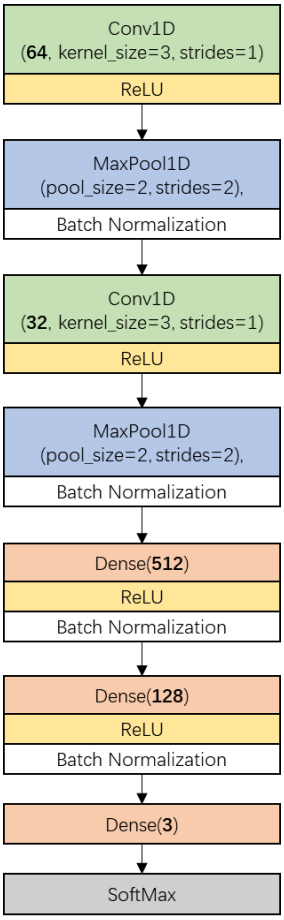

Figure 2. The proposed convolutional neural network (CNN) architecture for narrow-leaved oleaster fruit identification. Conv1D denotes 1-dimension convolution layer, ReLU (Rectified Linear Unit) is the activation function, MaxPool1D denotes 1-dimension max pooling layer, Dense denotes densely-connected neural network layer. The parameter of Conv1D which is defined as 'Channels' is the number of the kernels or filters. The parameter of Dense which is defined as 'units' is the number of the neurons.

3. Results

3.1. Spectral Profiles and Effective Wavelength Identification

Figure 3 shows the average spectra with standard deviation of each wavelength of narrow-leaved oleaster fruits from Gansu, Ningxia, and Xinjiang. Slight differences in reflectance values exist in the average spectra. The differences exist across the whole spectral ranges. However, the overlaps can be observed according to the standard deviation in Figure 3. With these overlaps, the samples from different geographical origins cannot simply be identified by observing their spectral differences. Figure 4 shows the second derivative spectra of the average spectra of narrow-leaved oleaster fruit samples from Gansu, Ningxia and Xinjiang. There are wavelengths with differences. Wavelengths corresponding to the peaks and valleys with greater differences were manually identified. As shown in Figure 4, a total of 22 wavelengths can be identified: 995, 1022, 1032, 1042, 1056, 1072, 1089, 1136, 1190, 1244, 1274, 1284, 1315, 1352, 1365, 1375, 1402, 1433, 1456, 1487, 1500, and 1632 nm. These wavelengths were selected as the effective wavelengths for geographical identification. In this study, the full spectra were used to conduct PCA for qualitative analysis of the sample cluster within one geographical origin

and sample separability among different geographical origins. The full spectra were also used to build machine learning models to quantitatively assess the sample separability among different geographical origins. To reduce redundant and collinear information which are informative in full spectra, simplify the models and improve model robustness, the selected effective wavelengths were used to build machine learning models for comparison with the full-spectra-based models.

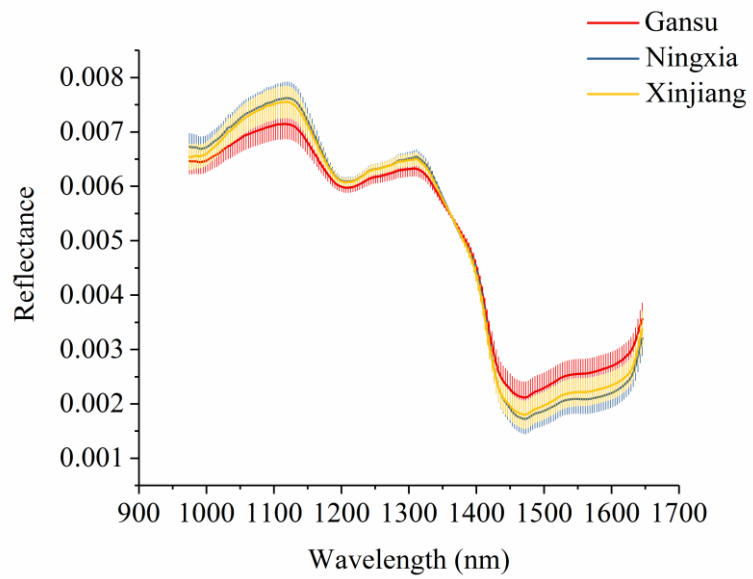

Figure 3. Average spectra with standard deviation of each wavelength of narrow-leaved oleaster fruits from Gansu, Ningxia, and Xinjiang.

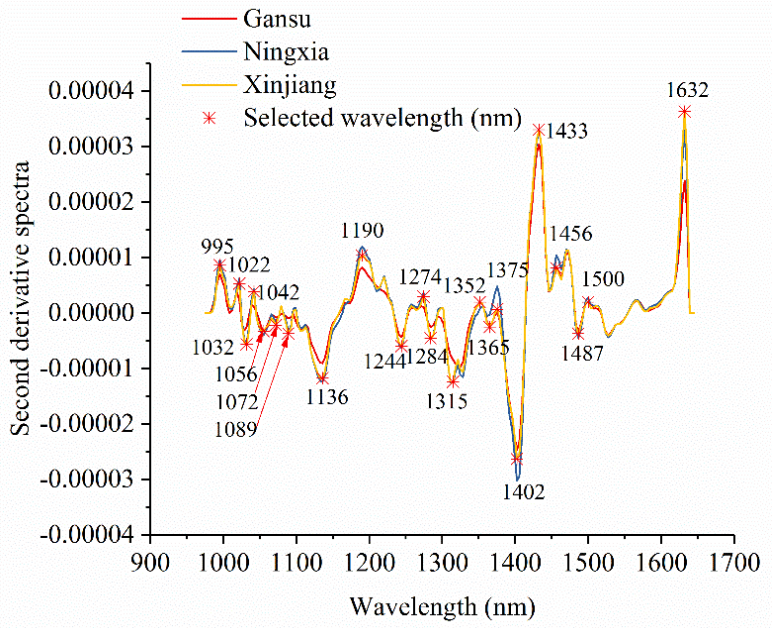

Figure 4. Effective wavelength selection using the second derivative spectra of average spectra of the samples from Gansu, Ningxia, and Xinjiang.

3.2. Principal Component Analysis

PCA was conducted to qualitatively cluster the samples in the scoring spaces. PCA was conducted on the full spectra of the calibration set, and the spectral data were centered for PCA analysis. The first three PCs explain most of the total variance, which was over 99% (PC1: 97.34%, PC2: 1.24%, PC3: 0.63%). Score scatter plots of two different PCs are shown in Figure 5. Samples from the same

geographical origins are marked with the same color, as well as the confidence ellipse (confidence level at 0.95). As shown in the score scatter plot of PC1 versus PC2, samples from each geographical origin are able to cluster well. Overlaps exist among the samples from Gansu, Ningxia, and Xinjiang. In the score scatter plot of PC1 versus PC3, samples from each geographical origin are able to cluster well. Samples from Gansu show greater overlaps with samples from the other geographical origins, and samples from Ningxia and Xinjiang are able to separate well. In the score scatter plot of PC2 versus PC3, samples from each geographical origin are able to cluster well. Samples from Gansu show greater overlaps with samples from the other geographical origins, and samples from Ningxia and Xinjiang are able to separate well. The score scatter plots in Figure 5 showed that the samples from different geographical origins are able to be well clustered and that they have great potential to be correctly identified.

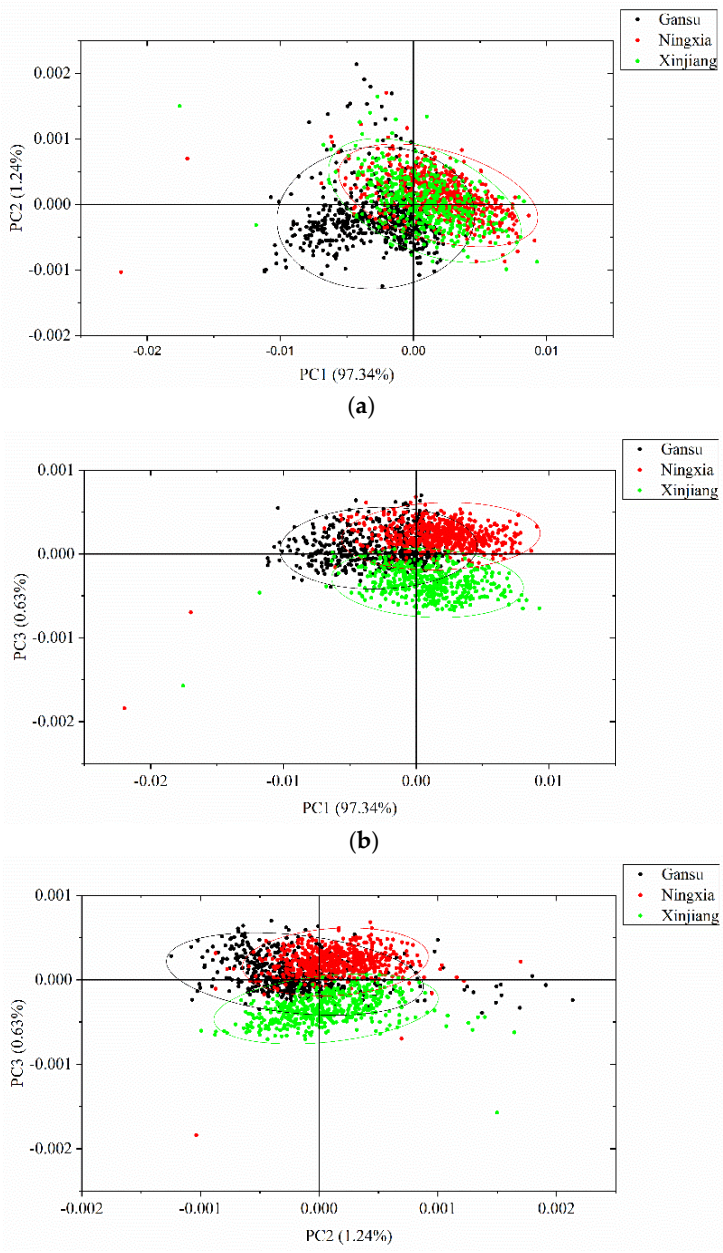

Figure 5. Principal component analysis (PCA) score scatter plots of (**a**) PC1 versus PC2; (**b**) PC1 versus PC3; and (**c**) PC2 versus PC3. The ellipse is the confidence ellipse (confidence level at 0.95).

3.3. Classification Models Using Full Spectra

PLS-DA, SVM, and CNN models were built using the full spectra. For the PLS-DA models, the category values of the samples from Gansu, Ningxia, and Xinjiang were labelled 001, 010, and 100. For the SVM and CNN models, the category values of the samples from Gansu, Ningxia, and Xinjiang were labelled 0, 1, and 2.

The classification results of the three different models are shown in Table 1. All discriminant models obtained good performances, with the classification accuracy of the calibration, validation, and prediction sets all over 90%. For the PLS-DA model, the optimal number of latent variables (LVs) was 12, and good classification performance was obtained. Classification accuracies of the calibration, validation, and prediction sets were all over 99%. For the SVM model, the model parameters (C, g) were optimized as (100, 10,000). The classification accuracy of the calibration set was 100%, while the classification accuracy of the validation and prediction sets was found to be lower. For the CNN model, the classification accuracy of the calibration, validation, and prediction sets were determined to be all over 97%. With regard to all three models, the PLS-DA model performed the best, the CNN model obtained results quite close to and slightly worse than those for PLS-DA, and the SVM model performed the worst.

Table 1. Confusion matrix of the partial least squares discriminant analysis (PLS-DA), support vector machine (SVM) and convolutional neural network (CNN) models using full spectra.

Model	Category Values	Calibration				Validation				Prediction			
		0	**1**	**2**	**Total (%)**	**0**	**1**	**2**	**Total (%)**	**0**	**1**	**2**	**Total (%)**
PLS	**0** *	539	0	0		291	0	0		268	0	7	
	1	0	601	1		0	303	0		0	299	1	
	2	0	0	481		0	0	241		0	0	240	
	Total (%)				99.94				100				99.02
SVM	**0**	539	0	0		289	0	2		224	0	51	
	1	0	602	0		0	303	0		0	300	0	
	2	0	0	481		0	0	241		0	0	240	
	Total (%)				100				99.76				93.74
CNN	**0**	539	0	0		289	0	2		253	0	22	
	1	1	601	0		0	303	0		0	300	0	
	2	6	0	475		4	0	237		0	0	240	
	Total (%)				99.57				99.28				97.30

* 0, 1, and 2 are the assigned category values of the samples from Gansu, Ningxia, and Xinjiang, respectively.

When using the PLS-DA model, samples from Ningxia were misclassified as samples from Xinjiang and samples from Gansu were misclassified as samples from Xinjiang; when using the SVM model, samples from Gansu were misclassified as samples from Xinjiang; and when using the CNN model, samples from Gansu and Xinjiang were misclassified as each other. The overall classification results indicated good separability among the samples from the three geographical origins. Samples from Gansu and Xinjiang were more likely to be misclassified, due to the results of the three discriminant models.

3.4. Classification Models Using Optimal Wavelengths

After effective wavelength selection, the PLS-DA, SVM, and CNN models were built using the selected effective wavelengths. The results of the three discriminant models are shown in Table 2. Good performances were obtained by the three models, with the classification accuracy of the calibration, validation, and prediction sets all over 95%. For the PLS-DA model, the optimal number of LVs was found to be 17. The classification accuracies of the calibration, validation, and prediction sets were all over 99%. For the SVM model, the model parameters (C, g) were optimized as (100, 108). The classification accuracies of the calibration, validation, and prediction sets were all over 95%. For the

CNN model, the classification accuracies of the calibration, validation, and prediction sets were all over 97%.

Table 2. Confusion matrices of the PLS-DA, SVM, and CNN models using effective wavelengths.

Model	Category Values	Calibration				Validation				Prediction			
		0	**1**	**2**	**Total (%)**	**0**	**1**	**2**	**Total (%)**	**0**	**1**	**2**	**Total (%)**
PLS	**0** *	538	0	1		291	0	0		272	0	3	
	1	1	601	0		0	303	0		0	300	0	
	2	1	0	480		0	0	241		0	0	240	
	Total (%)				99.92				100				99.63
SVM	**0**	539	0	0		271	0	20		238	0	37	
	1	0	602	0		0	303	0		0	300	0	
	2	2	0	479		1	0	240		1	0	239	
	Total (%)				99.88				97.49				95.34
CNN	**0**	539	0	0		287	0	4		263	0	12	
	1	0	602	0		0	303	0		0	299	1	
	2	4	0	477		8	0	233		5	0	235	
	Total (%)				99.75				98.56				97.79

* 0, 1, and 2 are the assigned category values of the samples from Gansu, Ningxia, and Xinjiang, respectively.

When using the PLS-DA model, samples from Gansu and Xinjiang were misclassified as each other, and one sample from Ningxia was misclassified as a sample from Gansu. When using the SVM model, it was observed that samples from Gansu and Xinjiang were misclassified as each other. When using the CNN model, samples from Gansu and Xinjiang were misclassified as each other, and one sample from Ningxia was misclassified as a sample from Xinjiang. The confusion matrices of the three models illustrate that samples from Gansu and Xinjiang were more likely to be misclassified.

The PLS-DA, SVM, and CNN models using effective wavelengths obtained similar results to those using effective wavelengths, illustrating the effectiveness of effective wavelength selection. The overall classification accuracy of all models indicates that there are great differences existing in narrow-leaved oleaster fruits from the three different geographical origins considered. As shown in Tables 1 and 2, the PLS-DA models performed slightly better than the CNN models, and the CNN models performed slightly better than the SVM models. Although differences existed in these model performances, the differences were quite small. The results illustrate that CNN models could be used for narrow-leaved oleaster fruit geographical origin identification. Moreover, the results of the discriminant models using full spectra and effective wavelengths all showed that samples from Gansu and Xinjiang were more likely to be misclassified.

4. Conclusions

In this work, near-infrared hyperspectral imaging was successfully used to identify the geographical origins of narrow-leaved oleaster fruits from Gansu, Ningxia, and Xinjiang. PCA score scatter plots showed the separability of the samples from the three geographical origins. PLS-DA, SVM, and CNN models were established using full spectra and effective wavelengths selected by second derivative spectra. The high classification accuracy, which was over 90% for models using full spectra and effective wavelengths, illustrates that the proposed method can effectively distinguish narrow-leaved oleaster fruits from different geographical origins. The performances of the models using effective wavelengths were similar to those using full spectra. Moreover, deep CNN models obtained close results to the PLS-DA and SVM models, showing good performances of deep learning for narrow-leaved oleaster fruit geographical origin detection. According to the discriminant models, samples from Gansu and Xinjiang were more likely to be misclassified. These results indicate that it would be possible to develop online systems for narrow-leaved oleaster fruit origin detection using near-infrared hyperspectral imaging and machine learning methods.

Author Contributions: Conceptualization, P.G.; data curation, P.G. and C.Z.; formal analysis, W.X.; funding acquisition, W.X.; investigation, C.Z.; methodology, C.Z., X.L., and Y.H.; project administration, P.G.; resources, T.Y.; software, T.Y.; supervision, Y.H.; validation, X.L.; visualization, T.Y.; writing—original draft, P.G., W.X., and C.Z.; writing—review and editing, X.L. and Y.H.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 61965014, and the Special Project for Scientific and Technological Innovation, grant number CXFZ201906.

Acknowledgments: The authors want to thank L.Z., a Ph.D candidate in College of Biosystems Engineering and Food Science, Zhejiang University, China, for providing help on data analysis.

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

References

1. Zhang, L.; Alvarez, L.V.; Bonthond, G.; Tian, C.; Fan, X. Cytospora elaeagnicola sp. nov. Associated with Narrow-leaved oleaster fruits Canker Disease in China. *Mycobiology* **2019**, *47*, 1–10. [CrossRef] [PubMed]
2. Zhanga, X.; Lia, G.; Sheng, D. Simulating the potential distribution of *Elaeagnus angustifolia* L. based on climatic constraints in China. *Ecol. Eng.* **2018**, *113*, 27–34. [CrossRef]
3. Lin, J.; Li, J.P.; Yuan, F.; Yang, Z.; Wang, B.S.; Chen, M. Transcriptome profiling of genes involved in photosynthesis in *Elaeagnus angustifolia* L. under salt stress. *Photosynthetica* **2018**, *56*, 1–12. [CrossRef]
4. Chen, X.; Yushuang, L.; Guangying, C.; Chi, G.; Shengge, L.; Huiming, H.; Tao, Y. Angustifolinoid A, a macrocyclic flavonoid glycoside from *Elaeagnus angustifolia* flowers. *Tetrahedron Lett.* **2018**, *59*, 2610–2613. [CrossRef]
5. Du, H.; Chen, J.; Tian, S.; Gu, H.; Li, N.; Sun, Y.; Ru, J.; Wang, J. Extraction optimization, preliminary characterization and immunological activities in vitro of polysaccharides from *Elaeagnus angustifolia* L. pulp. *Carbohydr. Polym.* **2016**, *151*, 348–357. [CrossRef]
6. Mcshane, R.; Auerbach, D.; Friedman, J.M.; Auble, G.T.; Shafroth, P.B.; Merigliano, M.; Scott, M.; Poff, N. Distribution of invasive and native riparian woody plants across the western USA in relation to climate, river flow, floodplain geometry and patterns of introduction. *Ecography* **2016**, *38*, 1254–1265. [CrossRef]
7. Collette, L.K.D.; Pither, J. Insect assemblages associated with the exotic riparian shrub Russian olive (Elaeagnaceae), and co-occurring native shrubs in British Columbia, Canada. *Can. Entomol.* **2016**, *148*, 316–328. [CrossRef]
8. Tredick, C.A.; Kelly, M.J.; Vaughan, M.R. Impacts of large-scale restoration efforts on black bear habitat use in Canyon de Chelly National Monument, Arizona, United States. *J. Mammal.* **2016**, *97*, gyw060. [CrossRef]
9. Khamzina, A.; Lamers, J.P.A.; Martius, C. Above- and belowground litter stocks and decay at a multi-species afforestation site on arid, saline soil. *Nutr. Cycl. Agroecosyst.* **2016**, *104*, 187–199. [CrossRef]
10. Singh, A.; Singh, N.B.; Hussain, I.; Singh, H.; Yadav, V.; Singh, S.C. Green synthesis of nano zinc oxide and evaluation of its impact on germination and metabolic activity of *Solanum lycopersicum*. *J. Biol.* **2016**, *233*, 84–94. [CrossRef]
11. Singh, A.; Singh, N.B.; Afzal, S.; Singh, T.; Hussain, I. Zinc oxide nanoparticles: A review of their biological synthesis, antimicrobial activity, uptake, translocation and biotransformation in plants. *J. Mater. Sci.* **2017**, *53*, 185–201. [CrossRef]
12. Hassanzadeh, Z.; Hassanpour, H. Evaluation of physicochemical characteristics and antioxidant properties of *Elaeagnus angustifolia* L. *Sci. Hortic.* **2018**, *238*, 83–90. [CrossRef]
13. Waili, A.; Yili, A.; Maksimov, V.V.; Mijiti, Y.; Atamuratov, F.N.; Ziyavitdinov, Z.F.; Mamadrakhimov, A.; Asia, H.A.; Salikhov, S.I. Erratum to: Isolation of Biologically Active Constituents from Fruit of *Elaeagnus angustifolia*. *Chem. Nat. Compd.* **2016**, *52*, 776. [CrossRef]
14. Wei, Q.; Wei, Y.; Wu, H.; Yang, X.; Zhang, H. Chemical Composition, Anti-oxidant, and Antimicrobial Activities of Four Saline-Tolerant Plant Seed Oils Extracted by SFC. *J. Am. Oil Chem. Soc.* **2016**, *93*, 1–10. [CrossRef]
15. Morehart, A.L. Phomopsis canker and dieback of *Elaeagnus angustifolia*. *Plant Dis.* **2015**, *64*, 66. [CrossRef]
16. Marena, M. Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chem. Soc. Rev.* **2014**, *43*, 8200–8214.

17. Ru, C.; Li, Z.; Tang, R. A Hyperspectral Imaging Approach for Classifying Geographical Origins of Rhizoma Atractylodis Macrocephalae Using the Fusion of Spectrum-Image in VNIR and SWIR Ranges (VNIR-SWIR-FuSI). *Sensors* **2019**, *19*, 2045. [CrossRef]
18. Noviyanto, A.; Abdulla, W.H. Honey botanical origin classification using hyperspectral imaging and machine learning. *J. Food Eng.* **2019**, *265*, 109684. [CrossRef]
19. Minaei, S.; Shafiee, S.; Polder, G.; Moghadam-Charkari, N.; Van Ruth, S.; Barzegar, M.; Zahiri, J.; Alewijn, M.; Kuś, P.M. VIS/NIR imaging application for honey floral origin determination. *Infrared Phys. Technol.* **2017**, *86*, 218–225. [CrossRef]
20. Puneet, M.; Alison, N.; Julius, T.; Guoping, L.; Sally, R.; Stephen, M. Near-infrared hyperspectral imaging for non-destructive classification of commercial tea products. *J. Food Eng.* **2018**, *238*, 70–77. [CrossRef]
21. Zhang, C.; Liu, F.; He, Y. Identification of coffee bean varieties using hyperspectral imaging: Influence of preprocessing methods and pixel-wise spectra analysis. *Sci. Rep.* **2018**, *8*, 2166. [CrossRef] [PubMed]
22. Yin, W.; Zhang, C.; Zhu, H.; Zhao, Y.; He, Y. Application of near-infrared hyperspectral imaging to discriminate different geographical origins of chinese wolfberries. *PLoS ONE* **2017**, *12*, e0180534. [CrossRef] [PubMed]
23. Zhu, S.; Zhou, L.; Gao, P.; Bao, Y.; He, Y.; Feng, L. Near-Infrared Hyperspectral Imaging Combined with Deep Learning to Identify Cotton Seed Varieties. *Molecules* **2019**, *24*, 3268. [CrossRef] [PubMed]
24. Mahesh, S.; Jayas, D.S.; Paliwal, J.; White, N.D.G. Comparison of partial least squares regression (plsr) and principal components regression (pcr) methods for protein and hardness predictions using the near-infrared (nir) hyperspectral images of bulk samples of Canadian wheat. *Food Bioprocess Technol.* **2015**, *8*, 31–40. [CrossRef]
25. Weinstock, B.A.; Janni, J.; Hagen, L.; Wright, S. Prediction of oil and oleic acid concentrations in individual corn (*Zea mays* L.) kernels using near-infrared reflectance hyperspectral imaging and multivariate analysis. *Appl. Spectrosc.* **2006**, *60*, 9. [CrossRef]
26. Sun, J.; Lu, X.; Mao, H.; Wu, X.; Gao, H. Quantitative determination of rice moisture based on hyperspectral imaging technology and bcc-ls-svr algorithm. *J. Food Process Eng.* **2016**, *40*. [CrossRef]
27. Zhang, N.; Liu, X.; Jin, X.; Li, C.; Wu, X.; Yang, S.; Ning, J.; Yanne, P. Determination of total iron-reactive phenolics, anthocyanins and tannins in wine grapes of skins and seeds based on near-infrared hyperspectral imaging. *Food Chem.* **2017**, *237*, 811–817. [CrossRef]
28. Rodríguez-Pulido, F.J.; Hernández-Hierro, J.M.; Nogales-Bueno, J.; Gordillo, B.; González-Miret, M.L.; Heredia, F.J. A novel method for evaluating flavanols in grape seeds by near infrared hyperspectral imaging. *Talanta* **2014**, *122*, 145–150.
29. Mazivila, S. Discrimination of the type of biodiesel/diesel blend (B5) using mid-infrared spectroscopy and PLS-DA. *Fuel* **2015**, *142*, 222–246. [CrossRef]
30. Botelho, B.G.; Reis, N.; Oliveira, L.S.; Sena, M.M. Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. *Food Chem.* **2015**, *181*, 31–37. [CrossRef]
31. Balage, J.M.; Amigo, J.M.; Antonelo, D.S.; Mazon, M.R.; e Silva, S.D. Shear force analysis by core location in Longissimus steaks from Nellore cattle using hyperspectral images—A feasibility study. *Meat Sci.* **2018**, *143*, 30–38. [CrossRef] [PubMed]
32. Melucci, D.; Bendini, A.; Tesini, F.; Barbieri, S.; Zappi, A.; Vichi, S.; Conte, L.; Gallina, T.T. Rapid direct analysis to discriminate geographic origin of extra virgin olive oils by flash gas chromatography electronic nose and chemometrics. *Food Chem.* **2016**, *204*, 263–273. [CrossRef] [PubMed]
33. Da Costa, G.B.; Fernandes, D.D.S.; Gomes, A.A.; De Almeida, V.E.; Veras, G. Using near infrared spectroscopy to classify soybean oil according to expiration date. *Food Chem.* **2016**, *196*, 539–543. [CrossRef]
34. Du, L.; Lu, W.; Cai, Z.J.; Bao, L.; Hartmann, C.; Gao, B.; Yu, L.L. Rapid detection of milk adulteration using intact protein flow injection mass spectrometric fingerprints combined with chemometrics. *Food Chem.* **2017**, *240*, 573–578. [CrossRef]
35. Schmutzler, M.; Beganovic, A.; Böhler, G.; Huck, C.W. Methods for detection of pork adulteration in veal product based on FT-NIR spectroscopy for laboratory, industrial and on-site analysis. *Food Control* **2015**, *57*, 258–267. [CrossRef]
36. Yang, H.X.; Fu, H.B.; Wang, H.D.; Jia, J.W.; Sigrist, M.W.; Dong, F.Z. Laser-induced breakdown spectroscopy applied to the characterization of rock by support vector machine combined with principal component analysis. *Chin. Phys. B* **2016**, *25*, 065201. [CrossRef]

37. Li, J.L.; Sun, D.W.; Pu, H.; Jayas, D.S. Determination of trace thiophanate-methyl and its metabolite carbendazim with teratogenic risk in red bell pepper (*Capsicum annuum* L.) by surface-enhanced Raman imaging technique. *Food Chem.* **2017**, *218*, 543–552. [CrossRef]
38. Ropodi, A.I.; Panagou, E.Z.; Nychas, G.J.E. Multispectral imaging (MSI): A promising method for the detection of minced beef adulteration with horsemeat. *Food Control* **2017**, *73*, 57–63. [CrossRef]
39. Wu, N.; Zhang, C.; Bai, X.; Du, X.; He, Y. Discrimination of Chrysanthemum Varieties Using Hyperspectral Imaging Combined with a Deep Convolutional Neural Network. *Moleclues* **2018**, *23*, 2831. [CrossRef]
40. Qiu, Z.; Chen, J.; Zhao, Y.; Zhu, S.; He, Y.; Zhang, C. Variety Identification of Single Rice Seed Using Hyperspectral Imaging Combined with Convolutional Neural Network. *Appl. Sci.* **2018**, *8*, 212. [CrossRef]
41. Zhang, C.; Feng, X.; Wang, J.; Liu, F.; He, Y.; Zhou, W. Mid-infrared spectroscopy combined with chemometrics to detect Sclerotinia stem rot on oilseed rape (*Brassica napus* L.) leaves. *Plant Methods* **2017**, *13*, 39. [CrossRef] [PubMed]