

Review

# A Review of the Discriminant Analysis Methods for Food Quality Based on Near-Infrared Spectroscopy and Pattern Recognition

Jian Zeng<sup>1,2,3</sup>, Yuan Guo<sup>1,2</sup>, Yanqing Han<sup>4</sup>, Zhanming Li<sup>4</sup> , Zhixin Yang<sup>4</sup>, Qinqin Chai<sup>1,3,\*</sup>, Wu Wang<sup>1,3</sup>, Yuyu Zhang<sup>2,\*</sup>  and Caili Fu<sup>2,4,5</sup>

<sup>1</sup> College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China; zengjian\_2868@163.com (J.Z.); orianaguo@gmail.com (Y.G.); wangwu@fzu.edu.cn (W.W.)

<sup>2</sup> Beijing Key Laboratory of Flavor Chemistry, Beijing Technology and Business University (BTBU), Beijing 100048, China; caili.fu@nusri.cn

<sup>3</sup> Ministry of Education Key Laboratory of Medical Instrument and Pharmaceutical Technology, Fuzhou University, Fuzhou 350108, China

<sup>4</sup> National University of Singapore (Suzhou) Research Institute, 377 Lin Quan Street, Suzhou Industrial Park, Suzhou 215123, China; yanqing.han@nusri.cn (Y.H.); lizhanming@just.edu.cn (Z.L.); yang\_zhixin12138@163.com (Z.Y.)

<sup>5</sup> Key Laboratory of Eel Culture and Processing of Fujian Province, Fuzhou 350208, China

\* Correspondence: qq.chai@fzu.edu.cn (Q.C.); zhangyuyu@btbu.edu.cn (Y.Z.)



**Citation:** Zeng, J.; Guo, Y.; Han, Y.; Li, Z.; Yang, Z.; Chai, Q.; Wang, W.; Zhang, Y.; Fu, C. A Review of the Discriminant Analysis Methods for Food Quality Based on Near-Infrared Spectroscopy and Pattern Recognition. *Molecules* **2021**, *26*, 749. <https://doi.org/10.3390/molecules26030749>

Academic Editor:

Alfonso Martinez-Felipe

Received: 29 December 2020

Accepted: 26 January 2021

Published: 1 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

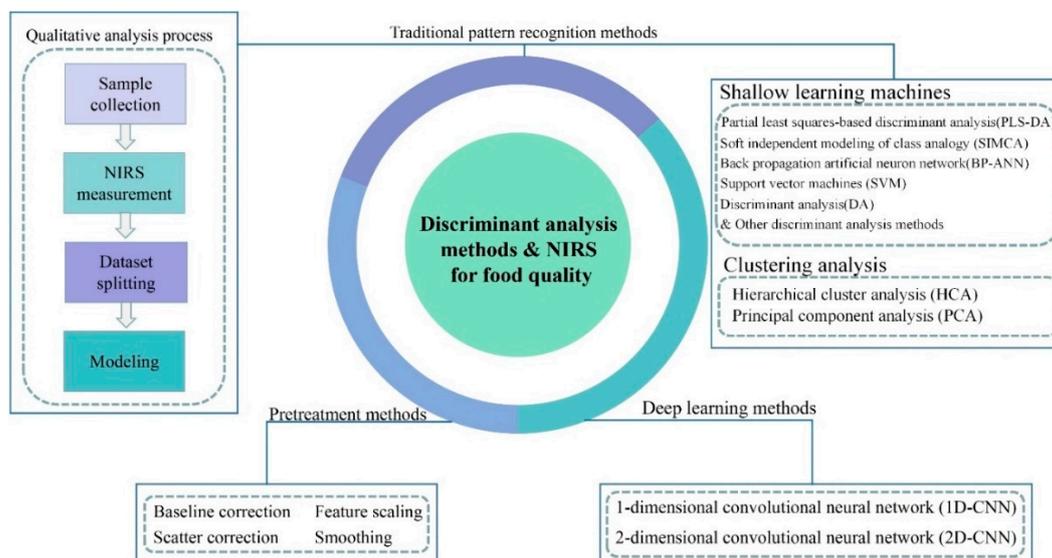
**Abstract:** Near-infrared spectroscopy (NIRS) combined with pattern recognition technique has become an important type of non-destructive discriminant method. This review first introduces the basic structure of the qualitative analysis process based on near-infrared spectroscopy. Then, the main pretreatment methods of NIRS data processing are investigated. Principles and recent developments of traditional pattern recognition methods based on NIRS are introduced, including some shallow learning machines and clustering analysis methods. Moreover, the newly developed deep learning methods and their applications of food quality analysis are surveyed, including convolutional neural network (CNN), one-dimensional CNN, and two-dimensional CNN. Finally, several applications of these pattern recognition techniques based on NIRS are compared. The deficiencies of the existing pattern recognition methods and future research directions are also reviewed.

**Keywords:** near-infrared spectroscopy; food quality; pattern recognition; deep learning; qualitative analysis

## 1. Introduction

With frequent exposure of the adulteration of food in the market, food safety issues have raised global health concerns. Adulterated food not only harms human health but also severely disrupts the order of the food market. Therefore, the detection and discrimination of food quality is of great importance. Traditional food quality detection methods are mainly based on high-performance liquid chromatography [1], mass spectrometry [2], or DNA technique [3]. However, the applications of these methods are limited by the following disadvantages: they are time-consuming, have low efficiency, have high measurement cost, require massive expertise, and cause damage to samples. Thus, it is necessary to seek a fast and efficient detection tool to facilitate the detection process and widen the applications. Near-infrared spectroscopy (NIRS) technology was developed nearly half a century ago as a detection tool. It obtains feature information of the hydrogen-containing groups (O-H, N-H, C-H) in organic substances by scanning the near-infrared spectrum of the samples [4] and possesses many advantages including being faster, having more accuracy, and being non-destructive. Recently, near infrared hyperspectral imaging (NIRHSI) technology was developed to obtain the chemical spectral information and the spatial

distribution of chemical substances in heterogeneous samples simultaneously [5]. It should be noted that by extracting the region of interest (ROI) from the hyperspectral image of the sample and calculating the average spectrum of each band within the ROI, the NIRHSI data can be reduced to near-infrared spectral data [6]. Thus, in this paper, we focus on the discrimination methods based on near-infrared spectroscopy (NIRS). NIRS is an indirect analysis technique that requires the establishment of empirical models through suitable pattern recognition methods. Although discrimination methods based on NIRS have been widely used for food quality discrimination, there are still some issues to be solved. In this review, the current developed NIRS-based food discrimination methods are compiled. The organizational structure of this article is shown in Figure 1, including pretreatment methods, traditional pattern recognition methods, and deep learning methods. The basic procedures of qualitative analysis based on NIRS is introduced in Section 2. Section 3 describes the main pretreatment methods used in NIRS. Section 4 presents different traditional pattern recognition methods combined with NIRS and their applications in food discrimination. Section 5 shows the potential of combining deep learning methods with NIRS to applied to food quality analysis. Finally, Section 6 summarizes the whole review and some future research directions are discussed.



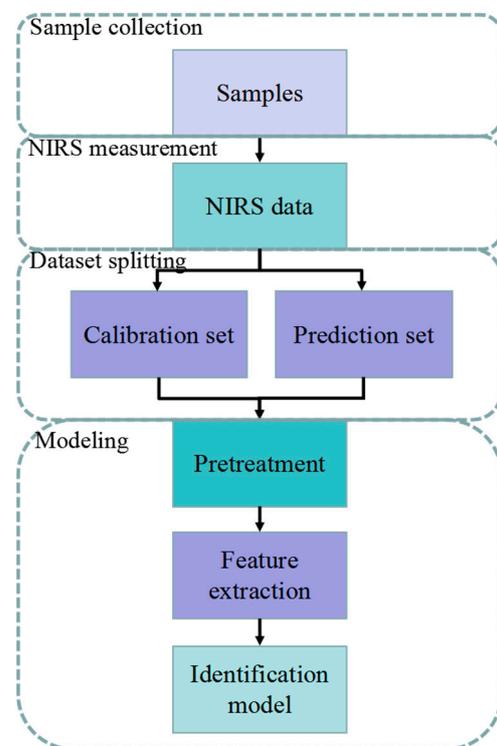
**Figure 1.** The organizational structure of the article.

## 2. Basic Structure of Qualitative Analysis Process Based on NIRS

The qualitative analysis process includes NIRS measurement and chemometric method for spectral analysis, which is shown in Figure 2.

Firstly, in order to obtain NIRS data, the samples are collected, stored, and specially processed if necessary. Then, each sample is measured with a near-infrared spectrum instrument. The obtained spectrum range is usually between 4000 and 10,000  $\text{cm}^{-1}$ . Although the near-infrared spectrometer is equipped with a high-performance detection device, the collected data can still be affected by the noise of the measuring instrument, thus affecting the effect of modeling [7]. Therefore, the spectral data must be pretreated before using chemometric methods. Secondly, the spectral peaks of different samples are usually similar in shape and position. The peak overlaps and interaction in some segments can be severe, which makes it difficult to distinguish the samples directly. Therefore, it is necessary to use chemometrics methods to achieve spectrum analysis and discrimination. Thirdly, the spectrum obtained has thousands of dimensions, which brings challenges to chemometrics methods, such as the problem of “the curse of dimensionality”. Therefore,

feature extraction methods usually are adopted to reduce the dimensions of the spectrum, and the discrimination is achieved by the establishment of pattern recognition methods.



**Figure 2.** Process of the qualitative analysis by near infrared spectroscopy (NIRS).

Clearly, preprocessing methods, feature extraction methods, and pattern recognition methods are three key procedures of qualitative analysis. Thus, preprocessing methods, feature extraction methods, and pattern recognition methods for food discrimination are reviewed in the following sections.

### 3. Pretreatment Methods

When analyzing complex samples, the final qualitative analysis results are often affected by interference factors such as stray light, noise, and baseline drift [8]. Therefore, in order to increase the stability and robustness of the qualitative analysis model, it is necessary to preprocess the original spectrum to eliminate the influence of uncertain factors and useless information variables before modeling [9]. There are many kinds of spectral preprocessing methods. According to their effect on preprocessing, they can be divided into four categories: baseline correction, scatter correction, smoothing treatment, and characteristic scaling [10,11]. Baseline correction is used to eliminate the influence of instrument background or drift. This method includes first derivative (D1), second derivative (D2), and continuous wavelet transformation. Scattering correction methods, such as multiplicative scatter correction (MSC) and standard normal variables (SNV), are used to avoid the influence of scattering caused by uneven particle distribution or different particle sizes to enhance the absorbance information that most related to the spectral content in the spectrum. Savitzky-Golay smoothing is a low-pass filter. It can effectively filter out the high-order signal interference and remove the effects of random noise. Feature scaling can eliminate the adverse effects caused by large-scale differences. This method includes mean centering, min-max scaling, and standardization.

### 4. Traditional Pattern Recognition Method

Traditional pattern recognition methods can be divided into supervised learning and unsupervised learning. The main difference between these two categories is that supervised

learning requires the tag information of the samples when training the models, whereas this is not necessary for unsupervised learning. Supervised learning mainly applies some shallow learning machines to achieve classification, while unsupervised learning mainly focuses on the clustering analysis. The following subsections describe them in detail.

#### 4.1. Shallow Learning Machines

Shallow learning machines lack effective feature expression ability. This means that the performance of these models highly depends on valid feature extraction. By combining feature extraction methods with shallow learning machines, the problem of curse of dimensionality can be avoided and the performance of shallow learning machines can be greatly improved [12]. The commonly used feature extraction methods include principal component analysis (PCA) [13–16], successive projections algorithm (SPA) [17,18], uninformative variables elimination (UAE) [19], and intelligent optimization algorithm [4,20,21]. The commonly used shallow learning machines include partial least squares-based discriminant analysis (PLS-DA) [22], soft independent modeling of class analogy (SIMCA), support vector machines (SVM), back propagation artificial neuron network (BP-ANN), and other discriminant analysis methods. The following parts will introduce some shallow learning machines according to their types.

##### 4.1.1. PLS-DA

PLS-DA is a supervised qualitative analysis method based on PLS regression analysis [23,24]. In order to achieve classification, the output matrix of PLS-DA is a one-hot encoding, while the input matrix and output matrix of PLS regression are continuous variables. For example, the PLS-DA output matrix  $Y = [y_1, y_2, y_3]$  of a three-category task is shown in Table 1. The threshold  $\theta_i$  is set to determine the category. If  $y_i = \max(y_1, \dots, y_n) > \theta_i$ , the predicted label is the  $i$ -th category.

**Table 1.** The partial least squares-based discriminant analysis (PLS-DA) output matrix of a three-category task.

Category	Y		
A	1	0	0
B	0	1	0
C	0	0	1

The PLS-DA model is widely applied in qualitative analysis of food NIRS, such as the identification of geographical origins and varieties. In the field of food geographical origin identification, in [15,25], PCA and PLS-DA binary classification models were built to discriminate the geographical origins of Huajiao and French olive oils. The accuracy rates of the model were 97.6% and 100%, respectively. In Reference [26], a PLS-DA classification model was established to trace the origin of Chilean flour and wheat grain. The flour possessed the best matrix for PLS-DA and the accuracy rate exceeded 90%. In Reference [27], PCA method combined with different preprocessing methods and different spectral bands was used to establish PLS-DA binary classification models to identify the origins of rice. The comparative experimental results showed that the model combining SNV and 4000–5500  $\text{cm}^{-1}$  spectral band was the most suitable, and the accuracy of the training set and testing set were 93.33% and 86.67%, respectively.

In the field of food varieties identification, in Reference [28], multiple PLS-DA models were established by combining different preprocessing methods to achieve the authentication of Iberian pork. The experiment showed that the model in combination with the SNV and D2 initialization methods achieved 93.2% and 93.4% accuracy in the training set and testing set, respectively, of the fresh meat and subcutaneous fat samples. In Reference [29], on the basis of visible NIRS, the authors constructed a PCA and PLS-DA model to identify varieties of Riesling and Chardonnay commercial wine. The recognition rate of Riesling and Chardonnay wines was up to 100% and 96%, respectively. Similarly, PLS-DA model

was established to identify oat and grain kernels through combining with NIRHSI technology. The results showed that the correct classification rates of oat and grain kernels were 97.1% and 99.0%, respectively, proving the effectivity of this method [30]. In Reference [31], on the basis of NIRHSI technology, the authors established SVM and PLS-DA models with a reduced set of only 54 wavelengths, resulting in excellent accuracy rates with 98.29% and 99.00% for the prediction sets, respectively. In Reference [32], on the basis of the spectral range of 5000–4000  $\text{cm}^{-1}$ , SNV method, and eight principal components, the PLS-DA model showed the best performance in the assessment of hard seed characteristics of legume seeds, with an accuracy of 100%. Furthermore, PLS-DA models were also applied to meat quality identification [33], obtaining good results.

#### 4.1.2. SIMCA

SIMCA is a supervised qualitative analysis method based on PCA algorithm [34,35]. The establishment of SIMCA model consists of three steps. First, PCA algorithm is used on the calibration set and the principal component regression model is established. Second, the similarities among the samples of different classes are calculated to obtain the critical value  $\theta_0$ . Finally, the similarity  $\theta_i$  among the tested sample and the samples of the given category in the calibration set is calculated. If  $\theta_i \leq \theta_0$ , the tested sample belongs to this category; otherwise, it does not belong to this category. In general, the classification performance of the SIMCA model is evaluated by

$$\text{Recognition rate}_i = \frac{n_c}{N_i} \quad (1)$$

$$\text{Rejection rate}_i = \frac{n_r}{N - N_i} \quad (2)$$

where  $N_i$  is the number of samples in the  $i$  category,  $n_i$  is the number of samples that correctly identify the category,  $n_r$  is the number of samples rejected from other categories, and  $N$  is the total number of samples.

In the field of food geographical origin identification, in References [36,37], SIMCA models with different numbers of principal components, different pretreatment methods, and different near-infrared spectral bands were established to identify wolfberry (*Lycium barbarum* L.) from eight different origins, achieving good results. In References [38,39], the SIMCA models were applied to identify the geographic origins of *Forsythia suspense* and mutton. The recognition rate of the geographic origin of forsythia was 90%, while the average recognition rate and rejection rate of the geographic origin of lamb were 93.75% and 100%, respectively.

In the field of qualitative analysis of food based on NIRS, the SIMCA model is also very popular. In Reference [40], full-band near-infrared spectroscopy data combined with SIMCA technology were used to identify three brands of milk powder—the recognition rate was 84.3% and the rejection rate was 91.6%. In Reference [41], combined with the SNV and D2 pretreatment methods, the SIMCA and PLS-DA models were established to identify the brand traceability of edible vinegar. The results showed that SIMCA was more competitive than PLS-DA, with accuracy rates of 84% and 94.7%, respectively. In Reference [42], the full spectrum band and the optimized characteristic band were used to establish the SIMCA models to classify four different types of wheat. The results showed that the recognition rate and rejection rate of the model were both 100%. In addition, the PLS-DA and SIMCA were also successfully applied to qualitative analysis of different navel orange varieties [43], with both recognition rates reaching 100%.

#### 4.1.3. SVM

SVM was firstly proposed by Vapnik in 1995 and is based on statistical learning theory to solve two-group classification problems [44]. It is then extended to solve multiple classification problems in Reference [45]. By using the principle of structural risk minimization and Vapnik-Chervonenkis (VC) dimension theory, SVM can handle the data

of high dimensions with good generalization and accuracy [46]. The VC dimension for the set of functions  $\{F\}$  is defined as the maximum number of training points that can be shattered by  $\{F\}$  [47]. It should be noted that SVM can only basically reflect the linear relations between the inputs and outputs. For nonlinear applications, the kernel function should be firstly used to map the data to high dimension, and then the optimal hyperplane should be constructed to classify the linearly separable data.

Since the mechanism of molecular vibration under different frequencies of light irradiation is complex, the obtained near-infrared spectral data contains nonlinear relations between the spectrum and its category. Thus, the SVM mentioned here is kernel SVM. The commonly used SVM kernel functions include linear kernel, polynomial kernel, radial basis function (RBF), and Gaussian kernel. Among these SVM kernels, RBF kernel is most widely used due to its few hyperparameters and fast approximation speed [48]. For SVM, the penalty parameter  $c$  and the RBF kernel function parameter  $g$  are two key parameters that influence the performance of the SVM model. The parameter  $c$  represents the tolerance to errors, which determines the generalization ability of the SVM model. The value of  $g$  influences the training and prediction speed of the SVM model [49]. However, in real applications, appropriate combination of these two parameters is difficult. Therefore, the way in which to optimize the two parameters is still a considerable issue. The commonly used methods for the parameter optimization of SVM model include trial and error method [50–52], grid search algorithms, and intelligence optimization algorithms.

The trial-and-error method-based SVM has many successful applications. For example, the RBF kernel function-based SVM models were established for qualitative analysis of apple origin and species [50], soybean oil color discrimination [51], and authenticity identification of tea [52], with accuracy rates of 100%, 100%, and 84.44%, respectively. However, the trial-and-error method relies on large numbers of experiments and expert experience. Moreover, the result searching may fall into local optimization, which will limit the performance of the model.

To avoid the drawbacks of the trial-and-error method, researchers have used the grid search method to select the  $c$  and  $g$  parameters. Established SVM models had been successfully applied to identify resin [53], adulterated peanut oil [54] origins, and varieties of maize kernels [14]. The model accuracy rates were 100%, 100%, and 95%, respectively. Although the grid search method is simple, universal, and easy to implement, the grid search algorithm is an exhaustive method. Thus, when the search range is relatively large, there will be too many invalid operations, which will result in a slow search speed.

Different from the grid search method, the intelligent optimization algorithm can select optimal parameter combination through automatic searching. The intelligent optimization algorithm has higher search efficacy and robustness as it is not easily affected by the initial value. It can also generally avoid falling into the local optimization. As a result, it has obtained more and more attention in the SVM training process in recent years. For example, as reported in Reference [55], genetic algorithm (GA) was used to select  $c$  and  $g$  parameters and then the SVM model combined with the SNV preprocessing method was established to identify different vegetable oils. Furthermore, the particle swarm optimization (PSO) was used to determine the  $c$  and  $g$  parameters for the establishment of the PCA-SVM model in order to identify the adulteration of hops [56] and Huajiao powder [57]. Apart from the abovementioned algorithms, the newly developed intelligent optimization algorithms, such as differential evolution algorithm, ant colony optimization algorithm, whale optimization algorithm, and wolf optimization algorithm can also be used for parameter optimization. In real applications, the algorithm complexity and the search efficiency should be taken into consideration.

Furthermore, there have been some improved SVM models used in the qualitative analysis of foods performed through NIRS. Least squares SVM (LS-SVM) is the most typical one. It converts inequality constraints into equality constraints on the basis of SVM. It can also transform the quadratic programming problem into the linear equation solution in SVM. Thus, compared with traditional SVM algorithm, the LS-SVM algorithm is simpler

and faster to realize and is more accurate [58]. In Reference [59], an LS-SVM model with GA optimization was established to identify whether the chestnut was moldy or not, with an accuracy of 93.56%. In Reference [60], an error-correcting output coding SVM model using NIRHSI technology for six different commercial tea products is presented, and an excellent accuracy rate of 97.41% was obtained. In References [61–63], PLS-DA, SIMCA, BP-ANN, SVM, and LS-SVM models were established separately to identify honey adulteration, lotus seed powder, and its adulteration and damage of Huping jujube. Comparison studies showed that the performance of the LS-SVM model was more competitive than other traditional pattern recognition methods in the application of NIRS in food qualitative analysis. Moreover, milk adulteration identification models were built by two improved and simplified *k*-nearest neighbors of non-linear supervised pattern recognition methods and improved SVM algorithms in Reference [64]. Experiments showed that the calculation burden of the improved SVM model was reduced by 90%, and good results were obtained.

#### 4.1.4. Discriminant Analysis Method

According to different discriminant principles, discriminant analysis methods include Fischer discriminant analysis, distance discriminant analysis, and Bayes discriminant analysis. Discriminant analysis methods are widely used in qualitative analysis of NIRS data of foods.

##### Fisher Discriminant Analysis

Fischer discriminant analysis, also known as linear discriminant analysis (LDA) method, is a projection-based method. By choosing appropriate projection axes for LDA, the projection points of similar samples are close to each other, while the projection points of different classes are as far as possible [65,66]. Then, the category of each sample is determined according to its distance from the projection points of other samples.

In the field of qualitative analysis of NIRS data, in Reference [13], PCA-LDA, SVM, and SIMCA methods were established to identify Chinese liquor, including 22 types, 10 famous brands, and 6 flavor types. Experiments showed that PCA-LDA was the best model. The average accuracy of the training set was 98.94%, and the average accuracy of the test set was 95.70%. In Reference [67], combined with the artificially selected near-infrared spectral bands, LDA and SIMCA were used to establish various discrimination methods of extra virgin olive oils. The experimental results showed that the average accuracy rates of the fivefold cross-validation were 90.5% and 80.5%, respectively. In this case, it was clear that LDA had better performance. In another case [68], PCA-LDA and GA-LDA were used to establish the identification model of macadamia cultivars. Experiments showed that, with Savitzky–Golay smoothing preprocessed spectra, the accuracy of GA-LDA was higher than 94.44%.

##### Distance Discriminant Analysis

Since the labels of the training data are known, the distance between the center of the unknown data and the training data can be calculated to determine which category the sample belongs to. Mahalanobis distance is usually used in distance discriminant analysis. The Mahalanobis distance discriminant analysis (MDDA) is used to find the Mahalanobis distance of the test sample from the center of each category. The category with the smallest distance is selected as the category identified by the MDDA model [69].

In the field of food quality evaluation and adulteration discrimination, the authors of [70] used the MDDA model to evaluate rice quality, and the accuracy of the model was 98.33%. Moreover, in References [71,72], MDDA combined with PCA was used to build qualitative models for peanut seed quality determination and surimi grade estimation—both of the models obtained good identification results. MDDA, SIMCA, and LS-SVM algorithms were compared in References [13,73] for estimating plum maturity levels and quality of Chinese liquors. The experimental results demonstrated that the MDDA and PCA-MDDA models had the best performance—the accuracy rates of testing set were

95.7% and 96.3%, respectively. In addition, in the tracing of food origin, the MDDA method is also involved. For example, the identification of sea cucumber geographical origin [74] and the discrimination of Chinese wolfberry from Zhongning [75]. The accuracy rates reached 100% in both cases.

#### Bayesian Discriminant Analysis

Bayesian estimation is to obtain the probability density function by combining the new evidence with the prior probability, so as to maximize the posterior probability. In Bayesian discriminant analysis methods, the probability density functions of each of the categories are estimated first, and then certain classification rules are set to determine the category for each test samples. Usually, the sample belongs to the category that has the greatest probability. Typical Bayesian discriminant analysis methods include the relevance vector machine (RVM) and the naive Bayesian classifier (NBC).

In the field of qualitative analysis of NIRS data, RVM and new clustering algorithm were used in Reference [76] to establish various identification models of adulteration of *Ganoderma lucidum* spore oil. The accuracy of the model was 100%. In Reference [77], on the basis of the competitive adaptive reweighted sampling (CARS) wavelength optimization, the authors compared RVM, SVM, and LS-SVM identification models of traditional Chinese medicine Sanyeqing and its similar species. The experimental results indicated that CARS-RVM achieved the best performance. The accuracy of the correction set and the verification set were all 100%. In addition, in Reference [78], an improved naive Bayesian classifier (INBC) combined with a fast search and find of density peaks (CFSFDP) clustering method was proposed to identify the origin of the traditional Chinese medicinal Sanyeqing. It should be pointed out that INBC-CFSFDP method can not only identify the known origin in the database accurately, but also identify the unknown origin. The recognition rate of unknown and known geographical origin in the database was 100%. This research opened up a new horizon for qualitative analysis of NIRS.

#### 4.1.5. Artificial Neural Network

Artificial neural network (ANN) is a mathematical model based on the mechanism of the processing of complex information by a biological neural network. ANN generally includes three layers: the input layer, hidden layer, and output layer. In the qualitative analysis of NIRS, the features extracted from the spectral data are inputted into the input layer. The number of input units is consistent with the input dimension of the spectral data. The hidden layer is to construct the relationships between the input features and the output categories. The outputs of the hidden layer are real numbers. The output layer is equipped with discriminant principles to generate the classified categories. The values of the output layer are the category labels. Generally, the more hidden layer units, the stronger nonlinear ability. However, too many hidden layers will increase the risk of overfitting. In the qualitative analysis of NIRS, the BP-ANN, RBF-ANN, and extreme learning machine (ELM) models are most widely used.

In terms of qualitative analysis of NIRS data, PCA combined with BP-ANN methods were applied in References [79,80] to establish the identification models of cultivated and wild *Erigeron breviscapus* and apple varieties. The recognition rate reached 100% in both cases. In Reference [81], in order to establish the BP-ANN model, NIRHSI technology was used to obtain the NIRS data. The model was used to identify soybean varieties. The training accuracy rate reached 97.50% and the average test accuracy rate was over 93.88%. The MDDA and BP-ANN models were compared in Reference [82] on the basis of NIR data of the browning of plum fruit. Experiments showed that both models can effectively identify the browning of plum fruit, and the recognition rates were all 100%. The PCA-BP-ANN and PLS-BP-ANN models were compared in Reference [83] on the basis of NIR data of different species of fish, and the accuracy of training set and testing set were 96.4% and 95.5%, respectively. The PLS-DA and BP-ANN models were compared in Reference [84] on the basis of NIR data of camellia oil from different geographical origins.

Experiments showed that the accuracies of MDDA and BP-ANN models were similar, but both were better than the PLS-DA model. The PCA-BP-ANN model had the best performance, with an accuracy rate of 95.83%. Combining the smoothing pretreatment method with the models can help to improve the models' performances. It is worth noting that the performance of ANN greatly depends on its parameters. Thus, how to optimize the methods' parameters is widely studied. For example, in Reference [85], BP-ANN was optimized by PSO method. The spectral band and parameters were optimized to establish the authenticity discrimination model of official rhubarb samples. The model accuracy reached 96.15%. Different from ANNs, the number of neurons in the hidden layer is the only parameter needed to be determined in ELM. Thus, ELM contains less parameters to be optimized, which means the model training process is simplified. The ELM combining with the CFSFDP clustering method was used to identify the adulterated grape seed oils in Reference [86], and a detection system was designed, with an accuracy rate of 100%.

#### 4.1.6. Other Qualitative Analysis Methods

For qualitative analysis of food by NIRS, in addition to the above-mentioned traditional pattern recognition methods, there are some other methods that are worth noting. For example, random forest (RF) and PCA were used to establish the identification model of green tea species on the basis of the joint information from NIRS and ultraviolet-visible spectroscopy [87]. The results show that the RF model can effectively identify five green tea varieties, and the accuracy rate of the model was 96%. Similarly, in Reference [88], Fourier transform mid-infrared spectroscopy and NIRS technologies were combined with single spectra analysis and multi-sensor information fusion strategy to establish the RF model for the origin identification of *Panax notoginseng*. The accuracy of RF identification model was 95.6%, suggesting an effective identification ability to identify *Panax notoginseng* from five different origins. In Reference [89], *k*-nearest neighbors (KNN) and LDA were used to establish the identification model of purple sweet potatoes, white sweet potatoes, and their adulterate samples. The prediction accuracy of LDA and KNN were 100% and 97.4%, respectively. In addition, HCA, KNN, and LDA were used to identify four production sources of tomato, and good results were obtained [90]. In Reference [91], NIRS data combined with PCA were used to establish SVM and learning vector quantization (LVQ) neural networks for coix seed variety identification, and the prediction accuracies were 100% and 90.91%, respectively. In Reference [92], PLS-DA, KNN, PCA, and classification and regression trees (CART) models were established to identify the quality of frying oil. The experimental results showed that the proposed models could effectively identify qualified and unqualified frying oil.

#### 4.2. Clustering Analysis

The cluster analysis method is an unsupervised pattern recognition method without labeled matrix. It can automatically realize classification by calculating the similarity of the samples. For NIRS data processing, the distance of the spectrum [93] is generally used to calculate the similarity of the data. However, for the processing of NIRS data of high-dimensions, distances among samples from different catalogs could be very close or the same. Usually, the feature extraction methods introduced in Section 4.1 are firstly used to eliminate overlapping parts of information and reduce the dimensions of NIRS. Then, the cluster analysis methods are applied.

The cluster analysis methods are widely used in qualitative analysis of NIRS data for food because they save the work of labeling data. For example, in Reference [94], with the 1100–1800nm band and the D2 and SNV pretreatment methods, the authors used the PCA method to establish the qualitative analysis model of shelf-life of Mopan persimmon. The accuracy of the model training set and testing set were 95.4% and 93.3%, respectively. In the PCA score plots, the visual classification can be presented directly, and samples with similar spectral characteristics can be clustered. In References [95–97], hierarchical cluster analysis (HCA) with different wavelength bands and different pretreatment methods were utilized

to analyze chicken origins, identify sesame oil and its adulterations, and discriminate the adulterated Cheng'an strawberry. Compared with PCA, HCA had better accuracy, with the accuracy rates of the HCA model being 100%, 100%, and 93.3%, respectively. In addition, other clustering methods, such as a fast search and find of density peaks clustering method, were also involved [78,86].

## 5. Deep Learning Methods

Deep learning has become a hot research direction in the field of artificial intelligence and has developed rapidly. Convolutional neural network (CNN) is one of the representative algorithms of deep learning. Like ANN, the purpose of CNN is also to simulate the learning process of the human brain. However, the structure of each layer in CNN is different from that of ANN. For example, the hidden layer of CNN involves several convolutional layers and pooling layers to automatically extract the essential features from the data. The fully connection layer realizes fitting or classification. In a word, the deep-learning-based model is an end-to-end self-adaptive model that only needs the input data and output data, and then the classified results can be obtained. The internal model can complete the feature extraction and classification process simultaneously.

Due to the excellent learning abilities for high-dimensional data, CNN has achieved great success in the food qualitative analysis with NIRS data. For example, a six-layer CNN model was designed in Reference [98] to achieve the quality identification of three types of macadamia nuts, with an accuracy of 100%. It provided a reliable technical support for the efficient and low-cost quality determination of macadamia nuts. In Reference [99], a 1D convolutional neural network (1D-CNN) was proposed for the discrimination of aristolochic acids and their analogues in traditional Chinese medicine. The validation set accuracy of 1D-CNN, PCA-SVM, T-distributed stochastic neighbor embedding support vector machine (*t*-SNE-SVM), SVM and BP-ANN were 100%, 91.41%, 90.63%, 87.89%, and 39.45%, respectively. The comparison results showed that traditional shallow learning machines heavily relied on the feature extraction algorithms such as PCA or *t*-SNE, while the accuracy of the 1D-CNN model was still very competitive without using any feature extraction algorithms, and was much better than BP-ANN. Similarly, SVM and self-adaptive evolutionary ELM and CNN models were established to identify the different geographical origins of *Tetrastigma hemsleyanum*. The accuracy rates were 90%, 98.3%, and 100%, respectively [100]. Experiments showed that CNN was faster and more accurate than traditional methods. CNN was also successfully used in the field of tobacco geographical origin identification. In Reference [101], the improved LeNet-5 network structure was used to identify the three major tobacco leaf production areas. The accuracy of the training set and testing set were 98.6% and 95.0%, respectively. In Reference [102], 1D-CNN, 2D-CNN, and PLS-DA were established to identify the geographical origin of tobacco leaves. Experiments showed that the performance of 1D-CNN and 2D-CNN was better than that of PLS-DA. The model accuracy rates were found to be 93.15%, 93.05%, and 81.25%, respectively. Although at present there are only a few studies on the application of CNN in the field of qualitative analysis of NIRS data of food, it is worthy of further research. Examples of all the mentioned applications of pattern recognition technology for quantitative analysis of food based on NIRS in this review are shown in Table 2.

Table 2. Application of pattern recognition technology for the qualitative analysis of food.

Pattern Recognition Technology	Applications	Objectives	Source	
PLS-DA	Geographic origin identification	Huajiao; French olive oils; rice; Chilean flour and wheat grain	[15,25–27]	
	Quality identification	Iberian pork; meat	[28,32]	
	Variety identification	Commercial wine; oat/grain kernels; hard seeds of legume plants	[29,30,33]	
SIMCA	Geographic origin identification	Wolfberry; <i>Forsythia suspense</i> ; mutton	[36–39]	
	Brand traceability	Milk powder; edible vinegar	[40,43]	
	Defective items classification	Wheat grain	[42]	
	Variety identification	Navel orange	[43]	
SVM	Origin and variety analysis	Apple; maize kernels; hops	[14,50,56]	
	Color discrimination	Soybean oil	[51]	
	Authenticity/adulteration identification	Tea; peanut oil; Huajiao powder; honey; lotus seed powder; milk	[52,54,57,61,62,64]	
	Defective items classification	Chestnut; Huping jujube	[59,63]	
Shallow learning machine	Classification	Tea	[60]	
	Identification of brand and type	Chinese liquor	[13]	
	Discriminant analysis	Variety identification	Extra virgin olive oils; Sanyeqing	[67,77]
		Nondestructive classification	Macadamia cultivars	[68]
		Geographic origin identification	Sea cucumber; Chinese wolfberry from Zhongning; Sanyeqing	[74,75,78]
	Quality and grade estimation	Peanut seed; surimi; plum	[71–73]	
	Adulteration identification	<i>Ganoderma lucidum</i> spore oil	[76]	
	Artificial neural network	Variety identification	<i>Erigeron breviscapus</i> ; soybean; apple; fish	[79–81,83]
		Color discrimination	Plum	[82]
		Geographical origin	Camellia oil	[84]
Authenticity/adulteration discrimination	Rhubarb; grape seed oils	[85,86]		
Others	Variety identification	Sweet potatoes; green tea; coix seed	[87,89,91]	
	Origin identification	Panax notoginseng; tomato	[88,90]	
	Quality assessment	Frying oil	[92]	

Table 2. Cont.

Pattern Recognition Technology	Applications	Objectives	Source
Clustering analysis	Shelf-life analysis	Mopan persimmon	[94]
	Geographic origin analysis	Chicken	[95]
	Adulteration identification	Sesame oil; Cheng'an strawberry	[96,97]
Deep learning methods	Quality identification	Macadamia nuts	[98]
	Discrimination	Aristolochic acids and their analogues in traditional Chinese medicine	[99]
	Geographical origin	Sanyeqing; tobacco leaves	[100–102]

## 6. Conclusions

Pattern recognition methods based on NIRS technique are widely used in the detection of food quality. The qualitative detection methods mainly focus on the early period shallow learning machines. However, they are not suitable for the feature learning of complex inputs when the number of samples is much smaller than the number of feature dimensions. Therefore, traditional NIRS technique-based discrimination methods usually employ feature extraction methods to eliminate redundant wavelengths before constructing shallow learning classifier. Still, the feature extraction methods depend on practical experience and professional knowledge. In addition, the selected features belonging to a certain kind of dataset may be not suitable for another dataset. This means that the features need to be re-extracted when the tested food products are replaced, which consumes a large amount of time. To avoid these drawbacks, deep learning theories such as CNN models are developed to achieve qualitative analysis. It can automatically extract important features from large collections of high-dimensional raw data with high recognition performance. Therefore, the CNN models based on NIRS techniques have great potential in the field of qualitative analysis of food.

The deep learning methods have excellent performance in the analysis of large datasets. If the dataset is small, it may lead to over fitting problem of the model. However, in the application of near infrared spectroscopy, a large number of samples are often difficult to obtain. For example, due to the long growth cycle and limited environmental conditions of some samples, the cost of obtaining a large number of NIRS data is too high. Therefore, the way in which to apply deep learning to the analysis of NIRS datasets with sample size far less than the sample dimension is a challenge in the future.

In addition, pattern recognition methods have developed from traditional shallow learning machines to deep learning methods. The novel methods have obtained effective results in food quality detection. However, the theories are limited to known categories that have been trained in the training stage. New unknown categories that have never merged in the dataset are less considered. Clearly, it means that the training set-based models cannot be applied to new unknown categories before training. In reality, as driven by interests, products of the same species but cultivated by new cultivation technology or new cultivation environment will continue to emerge, yet the qualities of these products may be unknown. Thus, enhancing the unknown categories recognition ability of the qualitative analysis models is a key direction for future research.

**Author Contributions:** Conceptualization, J.Z. and Q.C.; validation, Y.Z. and W.W.; formal analysis, Z.L.; investigation, J.Z. and Q.C.; resources, Q.C. and C.F.; writing—original draft preparation, J.Z.; writing—review and editing, Z.Y., Y.G. and Y.H.; visualization, Z.Y.; supervision, C.F.; project administration, Q.C.; funding acquisition, C.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the National Natural Science Foundation of China, grant number 61773124; the Open Project Program of Beijing Key Laboratory of Flavor Chemistry, grant number SPFW2019YB04; the Science and Technology Program of Fujian, grant number 2018S0001; the Science and Technology project of Dantu District, grant number NY2017003; the Scientific Research Project of Jinjiang Science and Education Park of Fuzhou University, grant number 2019-JJFDKY-48; the Science and Technology Program of Fuzhou, grant number 2018-N-9; and the Open Project Program of Beijing Key Laboratory of Flavor Chemistry, grant number SPFW2020YB10.

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

## References

1. Harp, B.P.; Miranda-Bermudez, E.; Baron, C.I.; Richard, G.I. Qualitative identification of permitted and non-permitted colour additives in food products. *Food Addit. Contam.* **2012**, *29*, 886–896. [[CrossRef](#)] [[PubMed](#)]
2. Self, R.L.; Wu, W.H. Rapid qualitative analysis of phthalates added to food and nutraceutical products by direct analysis in real time/orbitrap mass spectrometry. *Food Control* **2012**, *25*, 13–16. [[CrossRef](#)]
3. Lv, T.; Teng, R.; Shao, Q.; Wang, H.; Zhang, W.; Li, M.; Zhang, L. DNA barcodes for the identification of *Anoectochilus roxburghii* and its adulterants. *Planta Int. J. Plant Biol.* **2015**, *242*, 1167–1174. [[CrossRef](#)] [[PubMed](#)]
4. Li, Y.; Guo, Y.; Liu, C.; Wang, W.; Rao, P.; Fu, C.; Wang, S. SPA combined with swarm intelligence optimization algorithms for wavelength variable selection to rapidly discriminate the adulteration of apple juice. *Food Anal. Methods* **2017**, *10*, 1965–1971. [[CrossRef](#)]
5. Chandrasekaran, I.; Panigrahi, S.S.; Ravikanth, L.; Singh, C.B. Potential of near-Infrared (NIR) spectroscopy and hyperspectral imaging for quality and safety assessment of fruits: An overview. *Food Anal. Meth.* **2019**, *12*, 2438–2458. [[CrossRef](#)]
6. Feng, L.; Zhu, S.; Liu, F.; He, Y.; Bao, Y.; Zhang, C. Hyperspectral imaging for seed quality and safety inspection: A review. *Plant Methods* **2019**, *15*, 91. [[CrossRef](#)]
7. Oliveri, P.; Malegori, C.; Simonetti, R.; Casale, M. The impact of signal pre-processing on the final interpretation of analytical outcomes—A tutorial. *Anal. Chim. Acta* **2019**, *1058*, 9–17. [[CrossRef](#)]
8. Yu, C.; Chen, K.; Huang, B. The oretical hypothesis and verification of physical signal elimination in near infrared spectroscopy preprocessing. *Chin. J. Anal. Lab.* **2017**, *36*, 369–373.
9. Hui, G.; Sun, L.; Wang, J.; Wang, L.; Dai, C. Research on the pre-processing methods of wheat hardness prediction model based on visible-near infrared spectroscopy. *Spectrosc. Spectr. Anal.* **2016**, *36*, 2111–2116.
10. Rinnan, Å.; Van Den Berg, F.; Engelsen, S.B. Review of the most common pre-processing techniques for near-infrared spectra. *Trends Anal. Chem.* **2009**, *28*, 1201–1222. [[CrossRef](#)]
11. Jiao, Y.; Li, Z.; Chen, X.; Fei, S. Preprocessing methods for near-infrared spectrum calibration. *J. Chemometr.* **2020**, *34*, e3306. [[CrossRef](#)]
12. Yun, Y.H.; Li, H.D.; Deng, B.C.; Cao, D.S. An overview of variable selection methods in multivariate analysis of near-infrared spectra. *Trends Anal. Chem.* **2019**, *113*, 102–115. [[CrossRef](#)]
13. Li, Z.; Wang, P.; Huang, C.; Shang, H.; Pan, S.; Li, X. Application of Vis/NIR spectroscopy for Chinese liquor discrimination. *Food Anal. Methods* **2014**, *7*, 1337–1344. [[CrossRef](#)]
14. Han, Z.; Wan, J.; Zhang, H.; Deng, L.; Du, H.; Yang, J. Variety and origin identification of maize based on near infrared spectrum. *J. Chin. Cereals Oils Assoc.* **2014**, *29*, 21–25.
15. Wu, X.; Zhu, S.; Huang, H.; Xu, D.; Guo, Q. Near infrared spectroscopy for determination of the geographical origin of Huajiao. *Spectrosc. Spectr. Anal.* **2018**, *38*, 68–72.
16. Pu, D.D.; Duan, W.; Huang, Y.; Zhang, L.L.; Zhang, Y.Y.; Sun, B.G.; Ren, F.Z.; Zhang, H.Y.; Tang, Y.Z. Characterization of the dynamic texture perception and the impact factors on the bolus texture changes during oral processing. *Food Chem.* **2021**, *339*, 128078. [[CrossRef](#)]
17. Ye, S.; Wang, D.; Min, S. Successive projections algorithm combined with uninformative variable elimination for spectral variable selection. *Chemometr. Intell. Lab.* **2008**, *91*, 194–199. [[CrossRef](#)]
18. Soares, S.F.C.; Gomes, A.A.; Araujo, M.C.U.; Filho, A.R.G.; Galvão, R.K.H. The successive projections algorithm. *Trends Anal. Chem.* **2013**, *42*, 84–98. [[CrossRef](#)]
19. Centner, V.; Massart, D.; de Noord, O.E.; de Jong, S.; Vandeginste, B.M.; Sterna, C. Elimination of uninformative variables for multivariate calibration. *Anal. Chem.* **1996**, *68*, 3851–3858. [[CrossRef](#)]
20. Shi, J.Y.; Zou, X.B.; Wang, K.L.; Huang, X.W.; Sun, L.L.; Xia, R. Simulated annealing algorithm based wavenumber selection for total acid content analysis in vinegar by near infrared spectroscopy. *Food Sci.* **2011**, *32*, 120–123.

21. Lavine, B.K.; Mirjankar, N.; Delwiche, S. Classification of the waxy condition of durum wheat by near infrared reflectance spectroscopy using wavelets and a genetic algorithm. *Microchem. J.* **2014**, *117*, 178–182. [[CrossRef](#)]
22. Wang, L.; Kong, Q.; Li, G.; Zhang, L.; Yu, D.; Jiang, L. Characteristic band selection of near-infrared spectrum for determining peroxide value of oil based on iPLS. *Food Sci.* **2011**, *32*, 97–100.
23. Barker, M.; Rayens, W. Partial least squares for discrimination. *J. Chemom.* **2003**, *17*, 166–173. [[CrossRef](#)]
24. Brereton, R.G. Pattern recognition in chemometrics. *Chemom. Intell. Lab.* **2015**, *149*, 90–96. [[CrossRef](#)]
25. Galtier, O.; Dupuy, N.; Dréau, Y.L.; Ollivier, D.; Pinatel, C.; Kister, J.; Artaud, J. Geographic origins and compositions of virgin olive oils determined by chemometric analysis of NIR spectra. *Anal. Chem. Acta.* **2007**, *595*, 136–144. [[CrossRef](#)]
26. González-Martín, M.I.; Moncada, G.W.; González-Pérez, C.; Martín, N.Z.S.; López-González, F.; Ortega, I.L.; Hernández-Hierro, J.M. Chilean flour and wheat grain: Tracing their origin using near infrared spectroscopy and chemometrics. *Food Chem.* **2014**, *145*, 802–806. [[CrossRef](#)]
27. Gao, T.; Wu, J.; Lin, L.; Liu, Z.; Liu, C.; Yu, C. Chinese northeast rice origin traceability based on near-infrared spectroscopy and partial least square-discriminant analysis. *J. Chin. Cereals Oils Assoc.* **2019**, *124*, 114–117.
28. Horcada, A.; Valera, M.; Juárez, M.; Fernández-Cabanás, V.M. Authentication of Iberian pork official quality categories using a portable near infrared spectroscopy (NIRS) instrument. *Food Chem.* **2020**, *318*, 126471. [[CrossRef](#)]
29. Cozzolino, D.; Smyth, H.E.; Gishen, M. Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins. *J. Agric. Food Chem.* **2003**, *51*, 7703–7708. [[CrossRef](#)]
30. Serranti, S.; Cesare, D.; Marini, F.; Bonifazi, G. Classification of oat and groat kernels using NIR hyperspectral imaging. *Talanta* **2013**, *103*, 276–284. [[CrossRef](#)]
31. Feng, X.; Zhao, Y.; Zhang, C.; Cheng, P.; He, Y. Discrimination of transgenic maize kernel using NIR hyperspectral imaging and multivariate data analysis. *Sensors* **2017**, *17*, 1894. [[CrossRef](#)] [[PubMed](#)]
32. Zhu, L.; Huang, Y.; Wang, Q.; Ma, H.; Sun, B.; Sun, Q. Nondestructive identification of hard seeds of three legume plants using near infrared spectroscopy. *Trans. Chin. Soc. Agric. Eng.* **2012**, *28*, 237–242.
33. Yan, G.; Tang, X.; Wang, M.; Tao, R.; Mao, X. Rapid and nondestructive identification of chicken breeds by near infrared spectroscopy. *Food Sci.* **2015**, *36*, 148–152.
34. Svante, W. Pattern recognition by means of disjoint principal components models. *Pattern Recognition* **1976**, *8*, 127–139.
35. Kvalheim, O.M.; Karstang, T.V. Chapter 7 SIMCA—Classification by means of disjoint cross validated principal components models. *Data Handl. Sci. Technol.* **1992**, *9*, 209–248.
36. Lei, J.; Liu, D. Optimization of wolfberry (*Lycium barbarum L.*) traceability model for different geographic origins using near infrared spectroscopy combined with SIMCA. *Food Sci.* **2013**, *34*, 148–152.
37. Tang, L.; Liu, D. Tracing the geographic origin of Chinese wolfberry by near infrared spectroscopy. *Food Sci.* **2011**, *32*, 175–178.
38. Zhang, X.; Liu, J. Identification of *Forsythia suspense* from different habitats by NIR spectra. *Laser Infrared* **2008**, *4*, 342–344.
39. Zhang, N.; Zhang, D.; Li, S.; Li, Q. Preliminary study on origin traceability of mutton by near infrared reflectance spectroscopy coupled with SIMCA method. *Trans. Chin. Soc. Agric. Eng.* **2008**, *24*, 309–312.
40. Guan, X.; Gu, F.; Liu, J.; Yang, Y. Studies on the brand traceability of milk powder based on NIR spectroscopy technology. *Spectrosc. Spectr. Anal.* **2013**, *33*, 2621.
41. Guan, X.; Gu, F.; Liu, J.; Yang, Y. Development and optimization of the brand traceability model of vinegar based on near-infrared spectroscopy. *Mod. Food Sci. Technol.* **2014**, *30*, 200–203, 212.
42. Guan, E.; Cui, G.; Bian, K.; Zheng, Z. SIMCA identification model establishment of gibberellic disease wheat grain based on near infrared spectrum characteristics. *J. Chin. Cereals Oils Assoc.* **2016**, *31*, 124–129.
43. Hao, Y.; Sun, X.; Gao, R.; Pan, Y.; Liu, Y. Application of visible and near infrared spectroscopy to identification of navel orange varieties using SIMCA and PLS-DA methods. *Trans. Chin. Soc. Agric. Eng.* **2010**, *26*, 373–377.
44. Cortes, C.; Vapnik, V.N. Support Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
45. Chang, C.; Lin, C. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27. [[CrossRef](#)]
46. Vapnik, V.N. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **1999**, *10*, 988–999. [[CrossRef](#)]
47. Vapnik, V.N.; Chervonenkis, A.Y. On the uniform convergence of relative frequencies of events to their probabilities. Measures of complexity. In *Measures of Complexity*; Springer: Cham, Switzerland, 2015; pp. 11–30.
48. Bai, J.; Li, J.; Zou, H.; Tian, H.; Liu, F.; Li, W.; Wang, H.; Zhang, Z.; Wang, S. Qualitative and Quantitative Detection of Pork in Adulterated Beef Patties Based on Near Infrared Spectroscopy. *Food Sci.* **2019**, *40*, 287–292.
49. Chen, X.; Chai, Q.; Li, X.; Huang, J.; Wang, W. A Rapidly method for the discrimination of aristolochic acid and its analogues using SVM and PCA. In Proceedings of the 38th Chinese Control Conference, Guangzhou, China, 27 July 2019.
50. Zhao, J.; Hu, H.; Zou, X. Application of support vector machine to apple classification with near-infrared spectroscopy. *Trans. Chin. Soc. Agric. Eng.* **2007**, *4*, 149–152.
51. Wang, L.; Cui, Y.; Zhang, H.; Liu, F.; Li, G.; Ge, H.; Yu, D. Near-infrared spectroscopy analysis of soybean oil color based on SVM. *J. Chin. Cereals Oils Assoc.* **2015**, *30*, 115–119.
52. Chen, Q.; Zhao, J.; Zhang, H.; Wang, X. Identification of authenticity of tea with near infrared spectroscopy based on support vector machine. *Acta Opt. Sin.* **2006**, *6*, 933–937.

53. Wei, C.; Lei, F.; Ai, W.; Feng, J.; Zheng, H.; Ma, D.; Shi, X. Rapid identification of 6 kinds of traditional Chinese medicines containing resins and other components based on near infrared reflectance spectroscopy and PCA-SVM algorithm. *Chin. J. Exper. Tradit. Med. Formul.* **2017**, *23*, 17–23.
54. Zeng, L.; Tu, B.; Yin, C.; Zheng, X.; Song, Z.; He, D.; Qi, P. Qualitative and quantitative analysis of peanut oil adulteration by laser near infrared spectroscopy with SVM. *J. Chin. Cereals Oils Assoc.* **2016**, *31*, 126–137.
55. Tu, B.; Song, Z.; Zheng, X.; Zeng, L.; Yin, C.; He, D.; Qi, P. Effect of sample temperature on near-infrared qualitative analysis models of vegetable oil. *J. Chin. Cereals Oils Assoc.* **2016**, *31*, 133–137.
56. Guo, Y.; Chen, L.; Li, X.; Wang, G.; Jiang, Y.; Xin, H.; Jia, X. Rapid identification of domestic and imported hops based on NIRS technology and PCA-SVM algorithm. *J. Pharm. Pract.* **2019**, *37*, 322–331.
57. Wu, X.; Zhu, S.; Wang, Q.; Long, Y.; Xu, D.; Tang, C. Qualitative identification of adulterated Huajiao powder using near infrared spectroscopy based on DPLS and SVM. *Spectrosc. Spectr. Anal.* **2018**, *8*, 2369–2373.
58. Zhang, Y.; Tan, L.; Cao, F.; He, Y. Study on brand discrimination of fish oil based on multiple spectroscopy techniques. *Mod. Food Sci. Technol.* **2014**, *30*, 263–267.
59. Zhou, Z.; Li, X.; Li, P.; Gao, Y.; Liu, J. Near-infrared spectral detection of moldy chestnut based on GA-LSSVM and FFT. *Trans. Chin. Soc. Agric. Eng.* **2011**, *27*, 331–335.
60. Mishra, P.; Nordon, A.; Tschannerl, J.; Lian, G.; Redfern, S.; Marshall, S. Near-infrared hyperspectral imaging for non-destructive classification of commercial tea products. *J. Food Eng.* **2018**, *238*, 70–77. [[CrossRef](#)]
61. Tu, Z.; Zhu, D.; Ji, B.; Chen, H.; Qing, Z. Adulteration detection of honey based on near-infrared spectroscopy. *Trans. Chin. Soc. Agric. Eng.* **2011**, *27*, 382–387.
62. Fu, C.; Li, Y.; Chen, L.; Wang, S.; Wang, W. Rapid detection of lotus seed powder based on near infrared spectrum technology. *Spectrosc. Spectr. Anal.* **2018**, *38*, 424–429.
63. Wang, B.; Zhang, S.; Zhao, C. Dynamic discriminant models for the detection of subtle bruising in Huping Jujube constructed based on their visible/near-infrared spectral data. *Mod. Food Sci. Technol.* **2015**, *31*, 258–263.
64. Zhang, L.; Zhang, X.; Ni, L. Rapid identification of adulterated cow milk by non-linear pattern recognition methods based on near infrared spectroscopy. *Food Chem.* **2014**, *145*, 342–348. [[CrossRef](#)] [[PubMed](#)]
65. Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D.J. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE T. Pattern Anal.* **1997**, *19*, 711–720.
66. Fisher, R.A. The use of multiple measurements in taxonomic problems. *Ann. Eugenics* **1936**, *7*, 179–188. [[CrossRef](#)]
67. Sinelli, N.; Casale, M.; Egidio, V.D.; Oliveri, D.; Bassi, D.; Tura, D.; Casiraghi, E. Varietal discrimination of extra virgin olive oils by near and mid infrared spectroscopy. *Food Res. Int.* **2010**, *43*, 2126–2131. [[CrossRef](#)]
68. Carvalho, L.C.; Morais, C.L.M.; Lima, K.M.G.; Leite, G.W.P.; Oliveira, G.S.; Casagrande, I.P.; Neto, J.S.; Teixeira, G. Using intact nuts and near infrared spectroscopy to classify Macadamia cultivars. *Food Anal. Method.* **2018**, *11*, 1857–1866. [[CrossRef](#)]
69. Meng, Y.; Zhang, Y.; Wang, J.; Bi, R.; Wang, G. Rapid identification of meat of different animal origins based on near infrared spectroscopy. *Food Sci.* **2014**, *35*, 156–158.
70. Chen, K.; Gong, H.; Li, Y. Instantaneous determination of milled rice grade using NIR spectroscopy and discrimination analysis. *J. Chin. Cereals Oils Assoc.* **2011**, *26*, 105–109.
71. Zheng, T.; Sun, T.; Cao, Z.; Zhang, J. Quality analysis of peanut seed by visible/near-infrared spectra. *Spectrosc. Spectr. Anal.* **2015**, *35*, 622.
72. Wu, H.; Chen, W.; Wang, X.; Liu, Y. Applied research in grade estimation of surimi by near infrared spectroscopy. *Spectrosc. Spectr. Anal.* **2015**, *35*, 1239–1242.
73. Niu, X.; Gong, D.; Wang, Y.; Lu, W.; Liang, H.; Zhao, Z.; Ren, R. Determination of plum maturity levels by using NIR and chemometrics. *Mod. Food Sci. Technol.* **2014**, *30*, 230–234.
74. Guo, X.; Cai, R.; Wang, S.; Tang, B.; Li, Y.; Zhao, W. Non-destructive geographical traceability of sea cucumber (*Apostichopus japonicus*) using near infrared spectroscopy combined with chemometric methods. *R Soc. Open Sci.* **2018**, *5*, 170714. [[CrossRef](#)] [[PubMed](#)]
75. Xu, C.; Zhang, R.; Yu, X.; Wang, N. Discrimination of Chinese wolfberry from Zhongning based on near infrared spectroscopy. *Food Sci.* **2014**, *35*, 164–167.
76. Wang, W.; Wang, J.; Li, Y.; Li, X.; Li, Y. Study on diversified adulteration of *Ganoderma lucidum* spore oil by RVM and new clustering algorithms. *Spectrosc. Spectr. Anal.* **2017**, *37*, 1064–1068.
77. Fu, C.; Li, Y.; Wang, W.; Qiu, B.; Lin, Z.; Wang, S.; Wang, S.; Asiri, A.M.; Alarmy, K.A. Use of Fourier transform near-infrared spectroscopy combined with a relevance vector machine to discriminate *Tetrastigma hemsleyanum* (Sanyeqing) from other related species. *Anal. Methods* **2017**. [[CrossRef](#)]
78. Lin, W.; Chai, Q.; Wang, W.; Li, Y.; Qiu, B.; Peng, X.; Lai, T.; Huang, J.; Lin, Z. A novel method for geographical origin identification of *Tetrastigma hemsleyanum* (Sanyeqing) by near-infrared spectroscopy. *Anal. Methods* **2018**, *1*, 2980–2988. [[CrossRef](#)]
79. Li, G.; Zhang, L.; Yang, J.; Wang, D.; Liu, F.; Zhao, L. Identification Model of Cultivated and Wild Chinese Medical Herbs *Erigeron Breviscapus* with Near-Infrared Spectroscopy. *Spectrosc. Spectr. Anal.* **2007**, *27*, 1959–1961.
80. He, Y.; Li, X.; Shao, Y. Discrimination of varieties of apple using near infrared spectra based on principal component analysis and artificial neural network model. *Spectrosc. Spectr. Anal.* **2006**, *26*, 850–853.

81. Tan, K.; Chai, Y.; Song, W.; Cao, X. Identification of soybean seed varieties based on hyperspectral image. *Trans. Chin. Soc. Agric. Eng.* **2014**, *30*, 235–242.
82. Zhao, Z.; Wang, Y.; Gong, D.; Niu, X.; Cheng, W.; Gu, Y. Discrimination of plum browning with near infrared spectroscopy. *Spectrosc. Spectr. Anal.* **2016**, *36*, 2089–2093.
83. Xu, W.; Liu, R.; Hong, X.; Xiong, S. Discrimination of freshwater fish varieties based on near-infrared spectra. *Trans. Chin. Soc. Agric. Eng.* **2014**, *30*, 253–261.
84. Wen, T.; Zheng, L.; Gong, Z.; Li, L.; Xie, J.; Ma, Q. Rapid identification of geographical origin of camellia oil based on near infrared spectroscopy technology. *Trans. Chin. Soc. Agric. Eng.* **2016**, *32*, 293–299.
85. Tang, Y.; Zhang, Z.; Fan, G.; Zhu, H.; Wang, X. Identification of official rhubarb samples based on IR spectra and neural networks. *Spectrosc. Spectr. Anal.* **2005**, *25*, 715–718.
86. Tang, Y.; Chai, Q.; Lin, X.; Huang, J.; Li, Y.; Wang, W. Study on detection system of grape seed oil adulteration based on visible/near infrared spectroscopy. *Spectrosc. Spectr. Anal.* **2020**, *40*, 202–208.
87. Wang, X.; Huang, J.; Fan, W.; Lu, H. Identification of green tea varieties and fast quantification of total polyphenols by near-infrared spectroscopy and ultraviolet-visible spectroscopy with chemometric algorithms. *Anal. Methods* **2015**, *7*, 787–792. [[CrossRef](#)]
88. Zhou, Y.; Zuo, Z.; Xu, F.; Wang, Y. Origin identification of Panax notoginseng by multi-sensor information fusion strategy of infrared spectra combined with random forest. *Spectrochim. Acta Mol. Biomol. Spectro.* **2020**, *226*, 117619. [[CrossRef](#)] [[PubMed](#)]
89. Ding, X.; Ni, Y.; Kokot, S. NIR spectroscopy and chemometrics for the discrimination of pure, powdered, purple sweet potatoes and their samples adulterated with the white sweet potato flour. *Chemom. Intell. Lab.* **2015**, *144*, 17–23. [[CrossRef](#)]
90. Ding, X.; Guo, Y.; Ni, Y.; Kokot, S. A novel NIR spectroscopic method for rapid analyses of lycopene, total acid, sugar, phenols and antioxidant activity in dehydrated tomato samples. *Vib. Spectrosc.* **2016**, *82*, 1–9. [[CrossRef](#)]
91. Liu, X.; Mao, D.; Wang, Z.; Yang, Y. Rapid identification of Coix seed varieties by near infrared spectroscopy. *Spectrosc. Spectr. Anal.* **2014**, *34*, 1259–1263.
92. Zhu, X.; Li, G.; Qin, Y.; Su, D.; Liu, W.; Shan, Y. Rapid detection for the quality of frying oil using near infrared spectroscopy and pattern discrimination techniques. *Spectrosc. Spectr. Anal.* **2014**, *34*, 1259–1263.
93. Sun, X.; Shi, Y. Determining the geographic origin of cooked chicken based on near-infrared spectroscopy. *Mod. Food Sci. Technol.* **2015**, *31*, 315–321.
94. Wang, D.; Lu, X.; Zhang, P.; Chen, S.; Li, J. Qualitative identification of mopan persimmon shelf-life by near infrared spectroscopy. *Food Sci.* **2013**, *34*, 185–188.
95. Shi, Y.; Zhao, T.; Chen, H.; Xu, K. Traceability of chicken origin based on near-infrared spectroscopy. *J. Chin. Food Sci. Technol.* **2014**, *14*, 198–204.
96. Liu, J. Study on fast detection method of sesame oil authenticity. *J. Chin. Cereals Oils Assoc.* **2012**, *27*, 116–121.
97. Pang, Y.; Liu, K.; Yan, J.; Li, P. Rapid identification of Cheng'an strawberry with near infrared spectroscopy. *Mod. Food Sci. Technol.* **2013**, *29*, 1160–1162.
98. Du, J.; Hu, B.; Liu, Y.; Wei, C.; Zhang, G.; Tang, X. Study on quality identification of macadamia nut based on convolutional neural networks and spectral features. *Spectrosc. Spectr. Anal.* **2018**, *38*, 1514–1519.
99. Chen, X.; Chai, Q.; Lin, N.; Li, X.; Wang, W. 1D convolutional neural network for the discrimination of aristolochic acids and their analogues based on near-infrared spectroscopy. *Anal. Methods* **2019**, *11*, 5118–5125. [[CrossRef](#)]
100. Zhou, D.; Yu, Y.; Hu, R.; Li, Z. Discrimination of *Tetrastigma hemsleyanum* according to geographical origin by near-infrared spectroscopy combined with a deep learning approach. *Spectrochim. Acta A* **2020**, *238*, 118380. [[CrossRef](#)]
101. Lu, M.; Yang, K.; Song, P.; Shu, R.; Zhang, Y. The study of classification modeling method for near infrared spectroscopy of tobacco leaves based on convolution neural network. *Spectrosc. Spectr. Anal.* **2018**, *38*, 3724–3728.
102. Zhang, L.; Ding, X.; Hou, R. Classification modeling method for near-infrared spectroscopy of tobacco based on multimodal convolution neural networks. *J. Anal. Methods Chem.* **2020**, *22*, 1–13. [[CrossRef](#)]