

Article

Qualitative Analysis of Lambda-Cyhalothrin on Chinese Cabbage Using Mid-Infrared Spectroscopy Combined with Fuzzy Feature Extraction Algorithms

Yanjun Shen ¹, Xiaohong Wu ^{2,3,*} , Bin Wu ⁴, Yang Tan ¹ and Jinmao Liu ¹

¹ Institute of Talented Engineering Students, Jiangsu University, Zhenjiang 212013, China; jsu_syj@163.com (Y.S.); TaNvYaN@126.com (Y.T.); ljm3180301167@163.com (J.L.)

² School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China

³ High-Tech Key Laboratory of Agricultural Equipment and Intelligence of Jiangsu Province, Jiangsu University, Zhenjiang 212013, China

⁴ Department of Information Engineering, Chuzhou Polytechnic, Chuzhou 239000, China; wubin2003@163.com

* Correspondence: wxh_www@163.com

Abstract: Excess pesticide residues on cabbage are harmful to humans. In this study, we propose an innovative strategy for a quick and nondestructive qualitative test of lambda-cyhalothrin residues on Chinese cabbage. Spectral profiles of Chinese cabbage leaf samples with different concentrations of surface residues of lambda-cyhalothrin were collected with an Agilent Cary 630 FTIR Spectrometer. Standard normal variate (SNV), multiplicative scatter correlation (MSC), and principle component analysis (PCA) were utilized to preprocess the spectra. Then, fuzzy Foley-Sammon transformation (FFST), fuzzy linear discriminant analysis (FLDA), and fuzzy uncorrelated discriminant transformation (FUDT) were employed to extract features from the spectra data. Finally, *k*-nearest neighbor (*k*NN) was applied to classify samples according to the concentration of lambda-cyhalothrin residue. The highest identification accuracy rates of FFST, FLDA, and FUDT were 100%, 97.22%, and 100%, respectively. FUDT performed the best considering the combination of accuracy rate and required computing time. We believe that mid-infrared spectroscopy combined with fuzzy uncorrelated discriminant analysis is an effective method to accurately and quickly conduct qualitative analyses of lambda-cyhalothrin residues on Chinese cabbages. This method may have applications in other crops and other pesticide residues.

Keywords: Chinese cabbage; lambda-cyhalothrin; mid-infrared spectroscopy; fuzzy linear discriminant analysis; fuzzy Foley-Sammon transformation; fuzzy uncorrelated discriminant analysis; *k*NN



Citation: Shen, Y.; Wu, X.; Wu, B.; Tan, Y.; Liu, J. Qualitative Analysis of Lambda-Cyhalothrin on Chinese Cabbage Using Mid-Infrared Spectroscopy Combined with Fuzzy Feature Extraction Algorithms.

Agriculture **2021**, *11*, 275. <https://doi.org/10.3390/agriculture11030275>

Academic Editor: Azucena González Coloma

Received: 16 March 2021

Accepted: 19 March 2021

Published: 23 March 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/>).

1. Introduction

Chinese cabbage (*Brassica rapa*, *Chinensis* group), featuring high nutritional quality and mild flavor, has long been popular in China [1,2]. To protect the crop from pests and to increase the yield, farmers have long been using pesticides. In the past century, pyrethroids, including lambda-cyhalothrin, have been common pesticides to raise cabbage and other crops' production [3,4]. However, excess pyrethroid residue is harmful to mammals [4]. Researchers have found that there is a higher risk for children, ages 3–11, to develop neurological and developmental diseases if they consumed foods with excess pyrethroid residues frequently [4,5].

Detection of pesticide residues in food has long been in demand [6]. Traditionally, gas chromatography [7], liquid chromatography and its modified version, combined with another extraction method [8], were used. For higher sensitivity, tandem mass spectrometry is preferred [9]. Solid phase microextraction (SPME) was employed to determine 10 triazole fungicides in grapes [10]. Gas chromatography–mass spectrometry (GC-MS) was employed to quantitatively detect pesticides in propolis, and triadimefon was

found in most samples [11]. Although these methods are relatively sensitive and accurate, they require multistep sample preparation, and the processes are time-consuming as well as expensive [12].

In order for foods to be widely tested for pesticide residues, it is important that the methods used are fast and economical. At present, enzymatic inhibition methods and spectroscopy technologies are most often used. Based on an enzyme inhibition and heat generation method, a multilayer paper chip was proposed to detect organophosphorus and carbamate residues on vegetables and fruits [13]. Spectroscopy technologies have great potential in the field of food pesticide residue detection due to their low cost and ease of quantification [14]. Each pesticide has unique spectra in the near-infrared (NIR) and mid-infrared (MIR) wavelength ranges. The unique spectra are due to the pesticides' molecular structure induced vibration and rotation modes. Mid-infrared spectroscopy or Raman spectroscopy are often used to identify specific chemical components [15]. The information collected by the spectrometer can be used for qualitative and quantitative analyses [16]. Near infrared spectroscopy, transmission spectroscopy, polarization spectral detection technology, chlorophyll fluorescence spectra, and hyperspectral technique were demonstrated to detect triadimefon in propolis [11], fenvalerate and triazoline on lettuce leaves [17], five different types of organophosphorus pesticides on lettuce leaves [18], dimethoate on lettuce [19] and fenvalerate as well as triazoline on lettuce [20], respectively. To process the near infrared hyperspectral image information, researchers employed chemical molecular structure coupled with wavelet transform to find the most influential wavelength [21]. Then, these authors were able to detect five kinds of pesticides (dimethoate, acephate, phoxim, dichlorvos, and avermectin) [21]. Mid-infrared spectroscopy has also been applied to identify tea varieties, and the accuracy rate was 96.3% because of its high specificity [16]. In the following, we apply MIR spectroscopy, for the first time, to the detection of lambda-cyhalothrin residues on Chinese cabbages.

Fuzzy Foley-Sammon transformation (FFST), fuzzy linear discriminant analysis (FLDA), and fuzzy uncorrelated discriminant transformation (FUDT) are used in processing data and resolving component information of samples. FFST combined with an E-nose system has been employed to classify Chinese vinegar and reached an identification accuracy rate as high as 96.92% [22]. FLDA was efficient in classifying some Romanian and German mineral waters with an accuracy as high as 88% [23]. FUDT was applied to extract the feature information in images of synthetic aperture radar (SAR) [24]. In this study, FLDA, FFST, and FUDT were applied to detect lambda-cyhalothrin on Chinese cabbages and the performance of these different algorithms was compared.

In this work, the MIR technique, combined with the pattern recognition algorithms, was evaluated for the qualitative analysis of lambda-cyhalothrin residues on Chinese cabbages. The specific objectives are as follows: (1) to establish the feature extraction model FLDA, FFST, and FUDT and the supervised classification model *k*NN of the spectroscopic analysis; (2) to compare the performance of the three different pattern recognition algorithms, FLDA, FFST, and FUDT in detecting lambda-cyhalothrin residues on Chinese cabbage.

2. Materials and Methods

2.1. Sample Preparation

Fresh Chinese cabbage samples were bought from a supermarket in Zhenjiang, China. All samples were first washed with water (45 °C) to remove pesticide residues left on the sample surfaces. This was done to prevent the data from being contaminated, and to eliminate the possible presence of pesticides other than lambda-cyhalothrin on the Chinese cabbage. The samples were divided into four groups.

Lambda-cyhalothrin pesticide (5% EC, Shandong Shenda Crop Science Co. Ltd., Shouguang, China) was used. The lambda-cyhalothrin was mixed with water at ratios of 1:500, 1:100, and 1:20. Lambda-cyhalothrin pesticide mixtures at each dilution ratio were sprayed onto the surfaces of one group of Chinese cabbage leaves. One group that did

not have any pesticide sprayed on it served as the reference group. Samples that were sprayed with the 1:500 mixture were called the mildly contaminated group. Similarly, the samples sprayed with the 1:100 mixture and the 1:20 mixture were called the moderately contaminated and the seriously contaminated groups, respectively.

Each group of samples was measured with a spectrometer.

2.2. Spectral Data Acquisition

An Agilent Cary 630 FTIR Spectrometer with data collection software (Microlab PC and Resolutions Pro) was used in this study. The prepared samples were well ventilated and cooled for 24 h so that the influence of water on the spectra could be minimized. Afterwards, the samples of Chinese cabbage leaves were cut into pieces of about 2 mm × 2 mm. Each group had about 40 pieces (total 160 pieces) for spectra collection.

The spectrometer was set to ATR (attenuated total reflection) mode and was calibrated with the scanning times at 64. The wavenumber range was between 590–4289 cm⁻¹ with a resolution of 8 cm⁻¹. The diamond crystal surface was cleaned with anhydrous alcohol. The background spectra were collected first. The samples' spectra were collected after subtracting the background spectra. We collected 40 spectra from 40 pieces of cabbage leaves in each group. Each spectrum had 971 points. Among the spectra collected, 22 spectra in each group were picked to train the supervised model (total 88 spectra) and another 18 spectra were picked to test the model's performance of feature extraction and classification (total 72 spectra).

2.3. Spectral Processing Method

By analyzing the spectra of the Chinese cabbage samples, we can resolve the component information. However, the original spectra can be easily influenced by the physical properties of the samples. Because of scatter, the baseline may drift and this exerts a negative influence on the performance of classification results [25]. In order to eliminate physical light scatter and raise the signal to noise ratio, multiplicative scatter correction (MSC) was used in previous work [26]. As in this prior research, we used MSC to correct each experimental spectrum by subtracting the reference spectrum. This minimized the influence of baseline drift.

2.4. Principle Component Analysis

Principle component analysis (PCA) is an unsupervised method used to decrease the dimensions of data and obtain primary features. Although some information may be lost in the reduction, the most important information can be retained and noise can be eliminated. Meanwhile, the curse of dimensionality can be avoided, and the processing speed can be increased because of the fewer dimensions. When PCA is conducted on the data, it is common practice to select principle components that make larger contributions to the spectra. The larger the contribution is, the more component information can be reserved [22]. Considering the advantages of PCA, we employed it to process the spectral data after conducting SNV and MSC.

2.5. Feature Extraction Algorithms

In this study, three feature extraction algorithms (FFST, FLDA, and FUDT) combined with fuzzy theory were used to obtain the component information in the samples' spectral data. We introduce the concept of fuzzy membership value to represent the degree that one sample belongs to a certain class. Although one sample can be related to two or more classes because its fuzzy membership values of the related classes do not equal zero, the class of the sample is defined by the highest fuzzy membership value it possesses. Let

$X = \{x_i\}, i = 1, 2, \dots, n$, be a set of q -dimensional samples. The fuzzy membership value can be calculated as follows:

$$u_{ik} = \left[\sum_{j=1}^c \left(\frac{\|x_k - v_i\|^2}{\|x_k - v_j\|^2} \right) \right]^{-1} \quad (1)$$

In this equation, u_{ik} is the fuzzy membership value of the k -th sample x_k belonging to the i -th class and c is the number of classes. v_i here is the mean value of samples in class i .

2.5.1. Fuzzy Linear Discriminant Analysis

FLDA aims to minimize the distances between samples in the same class while maximizing the distances between samples in different classes. Computers can achieve this by projecting the spectra data to a lower dimension through the equations below [23].

The Fisher optimal discriminant function is:

$$\max J(\psi) = \frac{\psi^T S_{fB} \psi}{\psi^T S_{fW} \psi} \quad (2)$$

Equation (2) can be solved as an eigenvalue problem:

$$S_{fW}^{-1} S_{fB} \psi = \lambda \psi \quad (3)$$

where S_{fB} is the fuzzy between-class scatter matrix:

$$S_{fB} = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m (v_i - \bar{x})(v_i - \bar{x})^T \quad (4)$$

S_{fW} is the fuzzy within-class scatter matrix:

$$S_{fW} = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m (x_k - v_i)(x_k - v_i)^T \quad (5)$$

In Equations (4) and (5), n is the number of training samples and m ($m > 1$) is the weight index, while \bar{x} is the mean of the training samples:

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j \quad (6)$$

Based on the calculations above, the discriminant values and their corresponding discriminant vectors can be obtained. The bigger the discriminant value is, the better it can represent the whole data. Hence, with the data projected to the selected discriminant vectors, most information can be reserved for further processing.

Fuzzy linear discriminant analysis can be described in the following steps:

1. Fuzzify the data and determine membership values with Equation (1).
2. Obtain the discriminant values and their corresponding discriminant vectors, and then sort them in descending order $\psi_1, \psi_2, \dots, \psi_p$. The following linear transformation projects the data from R^q to R^p :

$$y = [\psi_1 \psi_2 \dots \psi_p]^T X \quad (7)$$

2.5.2. Fuzzy Foley-Sammon Transformation

In contrast to LDA, orthogonal and unitization constrains were added when maximizing the Fisher optimal discriminant function [27]. Using the method of Lagrange multipliers, a recursive expression can be obtained. Discriminant vectors on which the

data are projected can be calculated from the recursive expression. Combined with fuzzy theory, FFST was developed based on FST [22].

In fuzzy linear discriminant analysis, the vector ψ can be obtained by solving the following equation:

$$\max J(\psi) = \frac{\psi^T S_{fB} \psi}{\psi^T S_{fT} \psi} \tag{8}$$

Equation (8) can be solved as an eigenvalue problem:

$$S_{fT}^{-1} S_{fB} \psi = \lambda \psi \tag{9}$$

where S_{fT} is the fuzzy total class scatter matrix:

$$S_{fT} = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m (x_k - \bar{x})(x_k - \bar{x})^T \tag{10}$$

ψ_1 is the eigenvector corresponding to maximum eigenvalue λ_1 . Let the unit vector ψ_1 be the first vector of the set of fuzzy Foley-Sammon optimal discriminant vectors. Suppose the first r optimal discriminant vectors of FFST $\psi_1, \psi_2, \dots, \psi_r$ ($r \geq 1$) are obtained. The next vector ψ_{r+1} which maximizes Equation (8), can be computed with the following orthogonality constraint and recursive expression:

$$\psi_{r+1}^T \psi_i = 0, i = 1, 2, \dots, r \tag{11}$$

$$\psi_{r+1}^T S_{fT} \psi_{r+1} = 1 \tag{12}$$

$$Q S_{fB} \psi_{r+1} = \beta S_{fT} \psi_{r+1}, Q = I - \psi^T (\psi S_{fT}^{-1} \psi^T)^{-1} \psi S_{fT}^{-1} \tag{13}$$

where $\psi = [\psi_1, \psi_2, \dots, \psi_r]^T$. I is a unit matrix. ψ_{r+1} and β are the eigenvector and corresponding eigenvalue, respectively.

Fuzzy Foley-Sammon transformation can be described in the following steps:

1. Fuzzify the data and determine membership values with Equation (1).
2. Obtain the FLDA vector ψ_1 corresponding to the maximum discriminant value λ_1 of Equation (9). This is the first vector in the set of fuzzy Foley-Sammon optimal discriminant vectors.
3. Suppose the first r optimal discriminant vectors of FFST $\psi_1, \psi_2, \dots, \psi_r$ ($r \geq 1$) are obtained. The next step is to calculate the following vector ψ_{r+1} according to Equation (13).
4. Based on the above calculations, the optimal discriminant vectors $\psi_1, \psi_2, \dots, \psi_p$ can be obtained. The following linear transformation projects the data from R^q to R^p :

$$y = [\psi_1 \psi_2 \dots \psi_p]^T X \tag{14}$$

2.5.3. Fuzzy Uncorrelated Discriminant Analysis

After the data are projected onto the directions of the discriminant vector set using FST, it is proven that any two features are statistically correlated [28]. In order to obtain uncorrelated features after projection (better classification results can be obtained using uncorrelated features), extra constraints were proposed in uncorrelated discriminant transformation (UDT) [28]. It was found that UDT was more effective than FST during classification [29]. In light of the previous work, FUDT was used to extract the fuzzy features of synthetic aperture radar images [24].

Fuzzy discriminant analysis is also used to find the vector ψ by solving Equation (8) as an eigenvalue problem. ψ_1 is the eigenvector corresponding to maximum eigenvalue λ_1 . Let the unit vector ψ_1 be the first vector of the set of fuzzy uncorrelated optimal discriminant vectors. Suppose the first r optimal discriminant vectors of FUDT $\psi_1, \psi_2, \dots, \psi_r$ ($r \geq 1$) are

obtained. The next vector ψ_{r+1} , which maximizes Equation (8), can be computed with the following constraints and the recursive expression:

$$\psi_{r+1}^T S_{fT} \psi_i = 0, i = 1, 2, \dots, r \quad (15)$$

$$\psi_{r+1}^T S_{fT} \psi_{r+1} = 1 \quad (16)$$

$$Q S_{fB} \psi_{r+1} = \beta S_{fT} \psi_{r+1}, Q = I - S_{fT} \psi^T (\psi S_{fT} \psi^T)^{-1} \psi \quad (17)$$

where $\psi = [\psi_1, \psi_2, \dots, \psi_r]^T$. I is a unit matrix. ψ_{r+1} and β are the eigenvector and corresponding eigenvalues, respectively.

Fuzzy uncorrelated discriminant transformation (FUDDT) can be described in the following steps:

1. Fuzzify the data and determine membership values and centroids with Equation (1).
2. Obtain the FLDA vector ψ_1 corresponding to the maximum eigenvalue λ_1 of Equation (9). This is the first vector in the set of fuzzy uncorrelated optimal discriminant vectors.
3. Suppose the first r optimal discriminant vectors of FUDDT $\psi_1, \psi_2, \dots, \psi_r$ ($r \geq 1$) are obtained. The next vector ψ_{r+1} can be calculated in light of Equation (17).
4. Based on the above calculations, the optimal discriminant vectors $\psi_1, \psi_2, \dots, \psi_p$ can be obtained. The following linear transformation projects the data from R^q to R^p :

$$y = [\psi_1 \psi_2 \dots \psi_p]^T X \quad (18)$$

2.6. Classification Algorithm

In this study, k -nearest neighbor (k NN) was employed as the classifier because of its simplicity and good performance. Given a data space in which all the data points are labeled, the class of an unknown sample is determined by the classes of its k nearest neighbors. The Euclidian distance is used to define the distances between the unknown sample and the labeled data points. The class of the unknown sample depends on the majority class of its nearest neighbors, while k is always set to odd to avoid the same votes from different classes [30].

3. Results

3.1. Spectral Data Processing

In this study, the wavenumber range of collected spectra was 590–4289 cm^{-1} and the whole spectra data were used. In Figure 1, the first group of samples was the control, which were not sprayed with lambda-cyhalothrin. The other three groups of samples were sprayed with lambda-cyhalothrin/water ratios of 1:500, 1:100, and 1:20. Figure 1a shows the average spectra of these four groups of samples. They show the most differences among themselves in the range of wavelengths from 1000 cm^{-1} to 1700 cm^{-1} , and from 3000 cm^{-1} to 3500 cm^{-1} . To improve visualization, we enlarged the average spectra in the range of wavelengths from 1000 cm^{-1} to 1700 cm^{-1} in Figure 1b. The differences in the spectral data made it possible to classify samples with different concentrations of lambda-cyhalothrin residues.

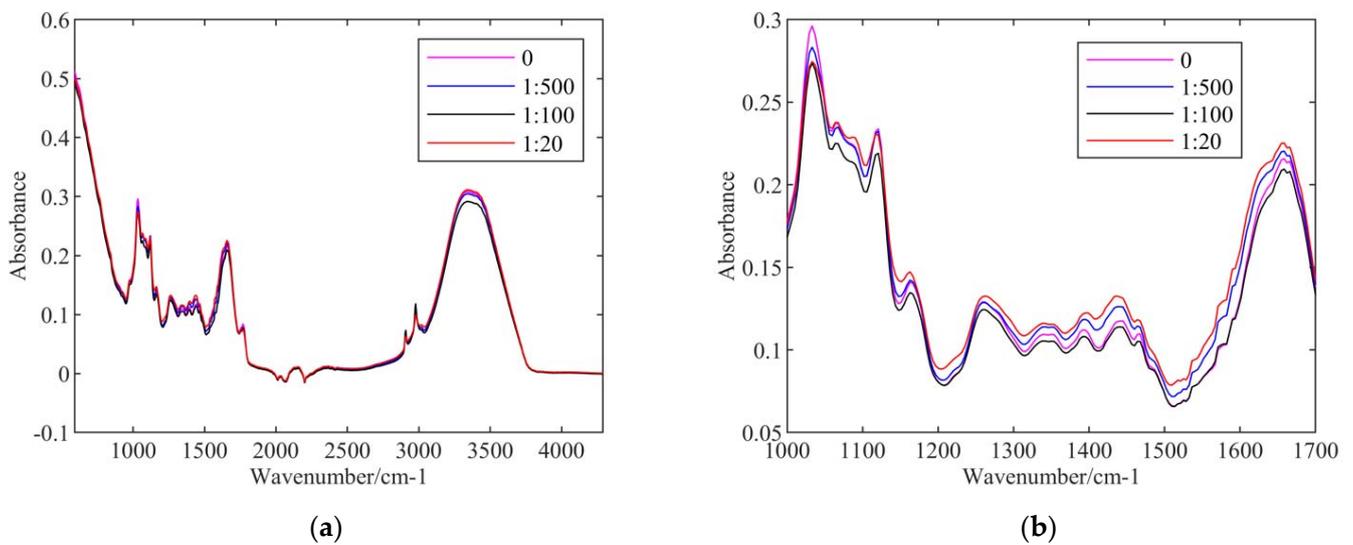


Figure 1. (a) Average spectra of four groups of samples; (b) Partially enlarged area of the average spectra.

Figure 2 compares the original data and the data processed by MSC.

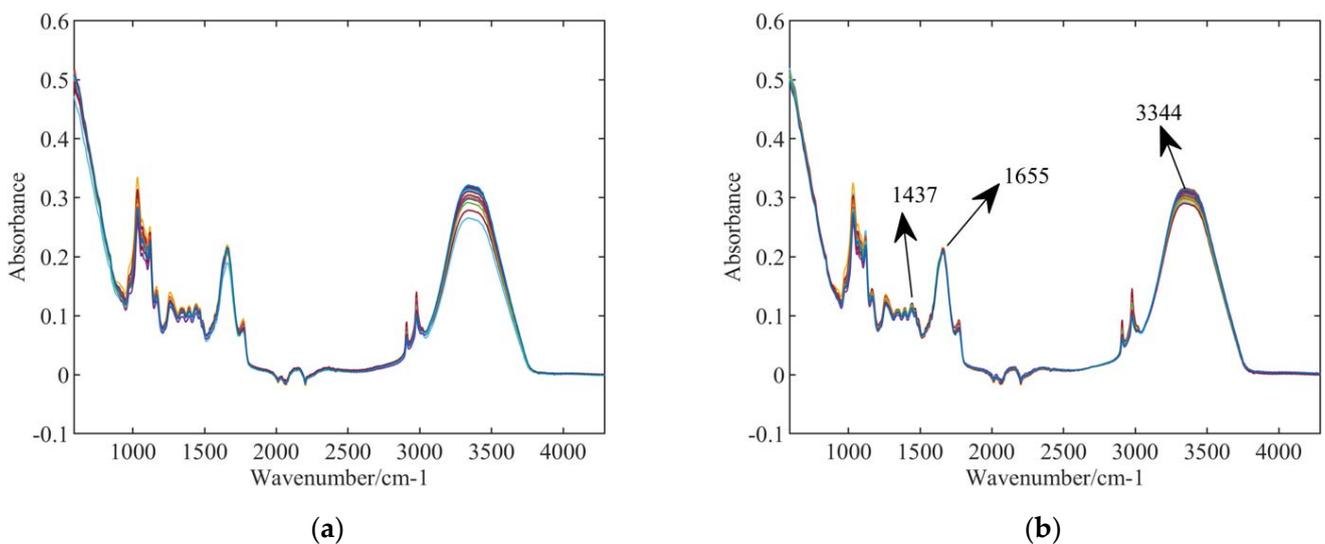


Figure 2. (a) Raw spectra of samples with no pesticide; (b) MSC processed spectra of samples with no pesticide.

Figure 2a,b show the spectra of samples in the first group (samples without pesticide) without and with MSC processing method. Compared to the raw spectra, the within-class variance was decreased, especially at the locations of the absorption peaks (1437 cm^{-1} , 1655 cm^{-1} , 3344 cm^{-1} , etc.). This improves the recognition rate of the spectra.

After MSC processing, standard normal variate (SNV) was used. The data were first subtracted by the mean value of the whole data, and then divided by their standard deviations, which makes the mean value zero and moves the central point of the data to the original point. This process was used in preparation for PCA, in order to avoid the disparity between magnitudes and ensure the significance of principle components.

3.2. Principle Component Analysis

To carry out the analysis, eigenvalues and corresponding eigenvectors of the covariance matrix were calculated. After that, eigenvalues were sorted in descending order. According to the input number of eigenvalues, the eigenvectors corresponding to the selected eigenvalues became the direction in which the training and test data were pro-

jected. In this study, the number of eigenvalues was 27 and the data dimensions were compressed from 971 to 27. The total contribution of the 27 principle components is 99.67%. This indicates that almost all of the spectra information is included in the compressed data. Meanwhile, this compression increases the computing speed. In Figure 3, the first principle component (PC1), the second principle component (PC2), and the third principle component (PC3) explained 57.55%, 22.81%, and 10.98% of the total variance. These three accounted for 91.34% of the total variance.

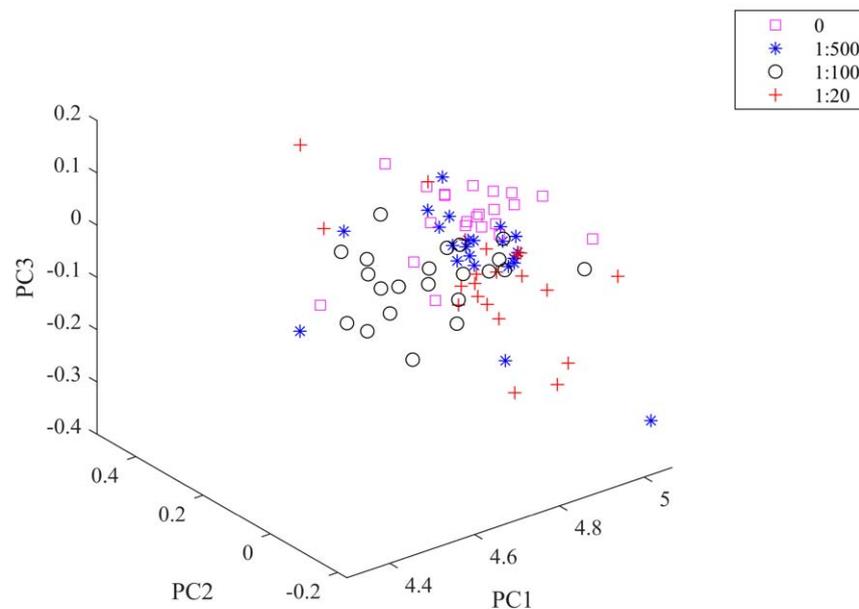


Figure 3. Top three principle components (PCs) for training set samples with no pesticide contamination, mild contamination (pesticide water ratio 1:500), moderate contamination (pesticide water ratio 1:100), and severe contamination (pesticide water ratio 1:20).

3.3. Calculation of Fuzzy Membership

Before applying FFST, FLDA, and FUDT, the fuzzy membership values were calculated according to Equation (1) and plotted in Figure 4.

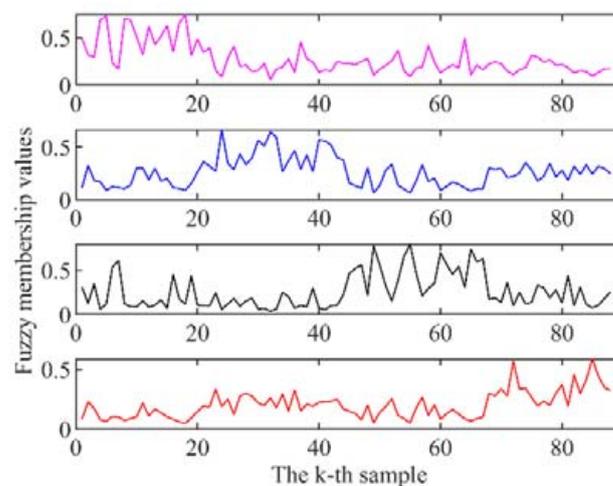


Figure 4. The fuzzy membership values of samples.

From top to bottom, the four subplots represent the samples' membership values belonging to the four classes. The top (control group) does not have any pesticide residues on the cabbage leaf sample. The ratio of lambda-cyhalothrin and water is 1:500, 1:100, and

1:20, respectively, from the second to the fourth rows. The abscissa is the k -th sample in the training set and the ordinate is the fuzzy membership value of samples. The class of a sample depends on its maximum membership value. If the maximum membership value of a sample belongs to the i -th class, this sample belongs to the i -th class. Thus, the most important differences among the plots is in which interval on the abscissa they have the highest value compared to the other three. This is important for the accurate classification of each sample. In Figure 4, it is shown that most samples' fuzzy membership values match with the class they belong to correctly. This can be helpful when conducting the classification algorithms FFST, FLDA, and FUDT.

3.4. Feature Extraction and Data Classification

FFST, FLDA, and FUDT were used to extract the features in the fuzzified training data. The basic idea of these algorithms is to maximize the between-class distances and to minimize the within-class distances. Through mathematical transformation, the training data were projected to a new space. In this space, the distances between samples with the same amount of pesticide were smaller, while the space between samples with different amounts of pesticide were larger. Then, the data in the test set were projected to the same space. Finally, k NN was applied to determine the amount of pesticide on samples in the test set. For an unknown sample, its amount of pesticide residue depends on that of its k nearest known training set samples. The performance of FLDA- k NN, FFST- k NN, and FUDT- k NN models were evaluated by calculating the proportion of samples in the test set (72 spectra) that were correctly classified. To establish the FLDA, FFST, and FUDT models and the k NN model, several parameters have to be determined. These parameters are the number of eigenvectors, the weight exponent m when computing the center of clustering, the number of discriminant vectors, and the value of k . The number of eigenvectors when applying PCA influences the amount of information retained and the same is true with the number of discriminant vectors in FFST and FUDT. The weight exponent m affects the center of clustering. Usually, the number of discriminant vectors is set to $c-1$, while c is the number of sample classes. Since the calculation was not complex, the enumeration method was employed. Fixing two parameters, m and k , we changed the number of eigenvectors to learn how the classification accuracy rate changes. Then the same method was used with m and k . To reduce the computational load, we created a plot to explore how the classification accuracy rates change when the parameters are changed.

Figure 5 illustrates the performance of FFST, FUDT, and FLDA combined with k NN on the test set. Although the lines fluctuate, the graph shows that the accuracy rates keep rising rapidly before reaching a plateau. This can be explained by the principle of PCA. Since several eigenvectors associated with bigger eigenvalues make a higher contribution, and usually the space formed by such eigenvectors can represent the data very well, those computed from smaller eigenvalues are meaningless and have low influence on the results of classification. Thus, the figure implies that only the first 30 eigenvectors are meaningful in the enumeration process.

When it comes to the influence of m and k on the results, the accuracy rates decrease when m and k increase. Figure 6 demonstrates how the classification accuracy rates change with the rise of the value of m . All three feature extraction algorithms perform the best when the value of m nears 1. The accuracy rates using FUDT and FLDA decline relatively slowly as m increases. However, the accuracy rates of FFST fall sharply after the value of m reaches 3. Figure 7 illustrates the changing trend of classification accuracy rates when k differs. Different from the results shown in Figure 6, the accuracy rates of the three algorithms change similarly, and their accuracy drops quickly after k reaches 9.

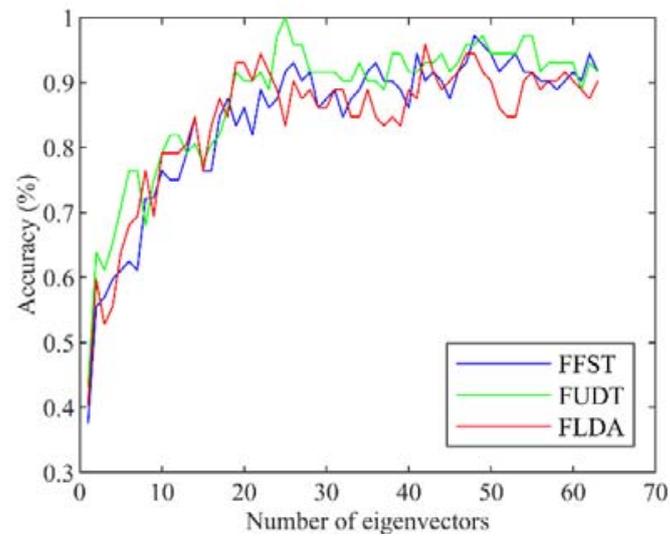


Figure 5. The accuracy rates on the test set when changing the number of eigenvectors.

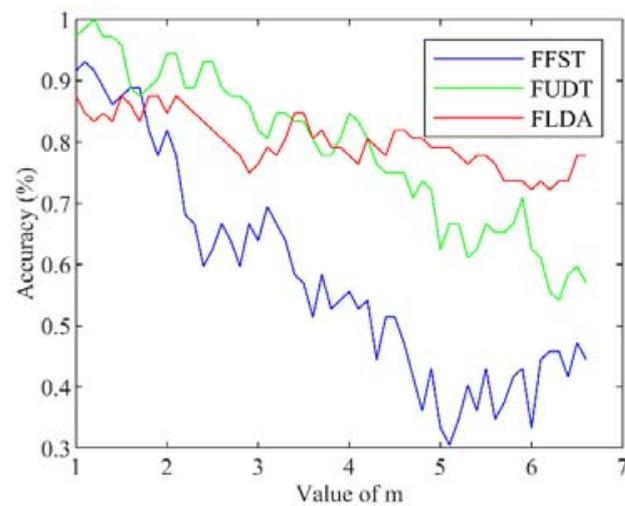


Figure 6. The accuracy rates on the test set when changing the value of weight exponent m .

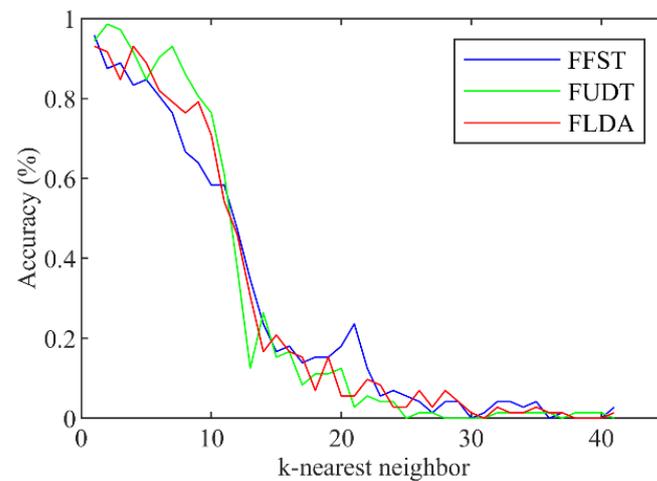


Figure 7. The accuracy rates on the test set when changing the value of k .

In light of the three above figures, the search range for optimizing the algorithms' use for pesticide residue detection can be narrowed. The search range for the number of eigenvectors was set from 3 to 65, and that of m was set from 1 to 3, with the step size set at 0.1, while the range of odd number k was set from 1 to 9. We used 22 spectra data in each group (88 total spectra from samples in the training set) to build the training set and 18 spectra in each group to test the performance of the feature extraction and classification model. Using these parameters, the final highest classification accuracy rates of FFST, FUDT, and FLDA are 100%, 100%, and 97.22%, respectively. The parameter settings and computer running times at the accuracy rate peaks are shown in Table 1.

Table 1. The accuracy rates, parameters, and running time using FFST, FUDT, and FLDA.

Method	Accuracy Rate	Parameters			Running Time (s)
		Number of Eigenvectors	M	k	
FFST	91.67%	27	1.2	7	0.015029
	97.22%	51	1	1	0.013869
	100%	49	1.1	1	0.014187
FUDT	100%	27	1.2	7	0.0040336
	95.83%	51	1	1	0.0033832
	95.83%	49	1.1	1	0.0035083
FLDA	83.33%	27	1.2	7	0.0019712
	97.22%	51	1	1	0.0012611
	94.44%	49	1.1	1	0.0016125

According to the results, using these three pattern recognition algorithms to extract the component information and k NN as the classifier is effective when processing the mid-infrared spectra data of Chinese cabbage leaf sample. Among these feature extraction methods, the time cost of FFST is about four times that of FUDT and about ten times that of FLDA. Compared to FLDA, FUDT performs better in terms of classification accuracy, even though FUDT requires twice the amount of computer time as that used by FLDA to complete the calculation. To conclude, FUDT performs the best on the test set of Chinese cabbage samples, considering both the classification accuracy rate and the computational time required.

4. Discussion

This study explored the performance of FLDA- k NN, FFST- k NN, and FUDT- k NN to qualitatively detect lambda-cyhalothrin on cabbage leaves. Some researchers used k NN to classify different kinds or concentrations of pesticide residues on vegetables after feature extraction [17,18]. In both of these publications, k NN and support vector machine (SVM) were employed as the classifier after extracting the feature. When classifying different concentrations of fenvalerate and triazoline on lettuce surfaces, accuracy rates on the test set using deep brief network- k NN (DBN- k NN) and DBN-SVM were 76.66% and 95%, respectively [17]. Using SVM as the classifier also outperformed k NN when classifying different kinds of pesticide residues. Accuracy rates on the test set using Competitive adaptive reweight sampling- k NN (CARS- k NN) and CARS-SVM were 84.44% and 97.78%, respectively [18]. k NN, as the classifier, performed unsatisfactorily in these publications. This might be related to the feature extraction algorithm, and could be studied in future research. In some cases, k NN could perform satisfactorily. When classifying four kinds of Chinese vinegar based on the electronic nose system, the highest identification rate using FFST- k NN was 96.92%. [22]. Additionally, it is easy to build the k NN model [30]. We can add samples with new labels into the training set directly, then an unknown sample is decided by its k nearest neighbors. The SVM model needs to be retrained when new kinds of samples appear. However, k NN is not suitable when the training set is large since the distances between the unknown sample and all objects in the training set are calculated [30].

Future work could address the balance between the computing time required and the size of the training set.

The method proposed in this article can be used for on-site and nondestructive detection of pesticide residues. Even though some of the traditional methods might be more sensitive [7–11], they require multistep sample preparation, are time-consuming and can only be conducted in the laboratory [12]. The spectroscopy method that we used could realize on-site detection of pesticide residues with portable spectrometers [31]. Moreover, there is no requirement for the use of chemical reagents for detection, which reduces the cost and effect on the environment. This method might also be utilized to inspect pesticide residues on other vegetables, or even be used for large scale residue detection on an automated production line. If such a system were placed in the factory, the mid-infrared spectrum of each product might be collected by a portable spectrometer. These data could then be sent for computation and classification using computers. Thus, products with excess pesticide residue could be sent for further processing. As more products were detected before going to the market, the pesticide concentration on the food could be maintained under the safety threshold. The public would be better protected from food poisoning by pesticide. This will also incentivize farmers to follow the standard of pesticide usage during the growth of vegetables.

Future work could be done to better distinguish lambda-cyhalothrin and other pyrethroid products (cypermethrin, deltamethrin), since they have very similar molecular structures. In comparison to cypermethrin and deltamethrin, lambda-cyhalothrin has a unique group (C-F bond) that has a unique absorption peak in the mid-infrared spectrum. By modifying the feature extraction algorithm to the unique absorption peak, the technology can be made more accurate in lambda-cyhalothrin detection. Additionally, full spectra were used for classification in our work. This might lead to interference and longer computational times. He et al. applied interval partial least squares (*i*PLS) to select a range of wavelengths containing the most useful information in the laser-induced breakdown spectroscopy (LIBS) spectra [32]. If combined with a region selection method, our method might be faster and more sensitive when interference exists.

5. Conclusions

In this study, mid-infrared spectroscopy combined with SNV, MSC, PCA, *k*NN, and three fuzzy feature extraction methods were successfully used to qualitatively analyze lambda-cyhalothrin residues on Chinese cabbage leaf samples. Under certain parameter settings, we achieved a classification accuracy of 100%.

When processing the spectra, SNV and MSC were used to eliminate the noise and prepare for PCA. Conducting PCA reduced the dimensionality of the data for faster processing speed. Using fuzzy theory, FFST, FLDA, and FUDT were employed to find spectra features. When classifying Chinese cabbages leaf samples with different amounts of lambda-cyhalothrin residues, the results showed that the highest classification accuracy rates of FFST, FUDT, and FLDA were 100%, 100%, and 97.22%, respectively. Among these methods, FUDT performed the best, considering the accuracy rate and computational time required.

This study describes a promising approach to realize fast and nondestructive detection of lambda-cyhalothrin residues on Chinese cabbage leaves using mid-infrared spectroscopy and fuzzy uncorrelated discriminant transformation (FUDT). This method can be applied to qualitatively detect different levels of pesticide residues in other vegetables as well.

Author Contributions: Conceptualization, X.W.; methodology, B.W.; software, B.W.; validation, X.W.; data curation, X.W.; writing—original draft preparation, Y.S.; writing—review and editing, X.W., Y.S., Y.T., and J.L.; visualization, X.W. and Y.S.; supervision, X.W.; project administration, X.W. and B.W.; funding acquisition, X.W. and B.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (31471413), the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), the Talent Program of Chuzhou Polytechnic (YG2019026 and YG2019024), the Undergraduate Innovation and Entrepreneurship Training Program of Jiangsu Province (202010299244Y) and the Natural Science Foundation of the Anhui Higher Education Institution of China (KJ2019A1129) for funding this study.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the authors.

Acknowledgments: We would like to thank Min Li (School of Electronics and Materials Engineering, Leshan Normal University, Leshan, China) for providing the spectral data.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Alenyorege, E.A.; Ma, H.; Aheto, J.H.; Agyekum, A.A.; Zhou, C. Effect of sequential multi-frequency ultrasound washing processes on quality attributes and volatile compounds profiling of fresh-cut Chinese cabbage. *LWT Food Sci. Technol.* **2020**, *117*, 108666. [[CrossRef](#)]
2. Kim, M.J.; Chiu, Y.C.; Kim, N.K.; Park, H.M.; Lee, C.H.; Juvik, J.A.; Ku, K.M. Cultivar-specific changes in primary and secondary metabolites in pak choi (*Brassica rapa*, *Chinensis* group) by methyl jasmonate. *Int. J. Mol. Sci.* **2017**, *18*, 1004. [[CrossRef](#)] [[PubMed](#)]
3. Spurlock, F.; Lee, M. Synthetic pyrethroid use patterns, properties, and environmental effects. In *Synthetic Pyrethroids, Proceedings of the Symposium on Synthetic Pyrethroids and Surface Water Quality held at the 232nd ACS National Meeting, San Francisco, CA, USA, 10–14 September 2006*; Gan, J., Spurlock, F., Hendley, P., Weston, D.P., Eds.; AMER CHEMICAL SOC: Washington, DC, USA, 2008; Volume 991, pp. 3–25.
4. Zhu, Q.; Yang, Y.; Zhong, Y.; Lao, Z.; O'Neill, P.; Hong, D.; Zhang, K.; Zhao, S. Synthesis, insecticidal activity, resistance, photodegradation and toxicity of pyrethroids (A review). *Chemosphere* **2020**, *254*, 126779. [[CrossRef](#)] [[PubMed](#)]
5. Lu, C.; Schenck, F.J.; Pearson, M.A.; Wong, J.W. Assessing children's dietary pesticide exposure: Direct measurement of pesticide residues in 24-hr duplicate food samples. *Environ. Health Persp.* **2010**, *118*, 1625–1630. [[CrossRef](#)]
6. Samsidar, A.; Siddiquee, S.; Shaarani, S.M. A review of extraction, analytical and advanced methods for determination of pesticides in environment and foodstuffs. *Trends Food Sci. Tech.* **2018**, *71*, 188–201. [[CrossRef](#)]
7. Pihlström, T.; Blomkvist, G.; Friman, P.; Pagard, U.; Österdahl, B.G. Analysis of pesticide residues in fruit and vegetables with ethyl acetate extraction using gas and liquid chromatography with tandem mass spectrometric detection. *Anal. Bioanal. Chem.* **2007**, *389*, 1773–1789. [[CrossRef](#)]
8. Bakirci, G.T.; Hişil, Y. Fast and simple extraction of pesticide residues in selected fruits and vegetables using tetrafluoroethane and toluene followed by ultrahigh-performance liquid chromatography/tandem mass spectrometry. *Food Chem.* **2012**, *135*, 1901–1913. [[CrossRef](#)] [[PubMed](#)]
9. Hiemstra, M.; de Kok, A. Comprehensive multi-residue method for the target analysis of pesticides in crops using liquid chromatography-tandem mass spectrometry. *J. Chromatogr. A* **2007**, *1154*, 3–25. [[CrossRef](#)]
10. Souza-Silva, E.; Lopez-Avila, V.; Pawliszyn, J. Fast and robust direct immersion solid phase microextraction coupled with gas chromatography-time-of-flight mass spectrometry method employing a matrix compatible fiber for determination of triazole fungicides in fruits. *J. Chromatogr. A* **2013**, *1313*, 139–146. [[CrossRef](#)] [[PubMed](#)]
11. González-Martín, M.I.; Revilla, I.; Vivar-Quintana, A.M.; Betances Salcedo, E.V. Pesticide residues in propolis from Spain and Chile. An approach using near infrared spectroscopy. *Talanta* **2017**, *165*, 533–539. [[CrossRef](#)]
12. Wang, J.; Gao, L.; Han, D.; Pan, J.; Qiu, H.; Li, H.; Wei, X.; Dai, J.; Yang, J.; Yao, H.; et al. Optical detection of λ -cyhalothrin by core-shell fluorescent molecularly imprinted polymers in Chinese spirits. *J. Agric. Food Chem.* **2015**, *63*, 2392–2399. [[CrossRef](#)]
13. Yang, N.; Wang, P.; Xue, C.; Sun, J.; Mao, H.; Oppong, P.K. A portable detection method for organophosphorus and carbamates pesticide residues based on multilayer paper chip. *J. Food Process Eng.* **2018**, *41*, e12867. [[CrossRef](#)]
14. Lin, X.; Sun, D. Recent developments in vibrational spectroscopic techniques for tea quality and safety analyses. *Trends Food Sci. Tech.* **2020**, *104*, 163–176. [[CrossRef](#)]
15. Hu, X.; Shi, J.; Zhang, F.; Zou, X.; Holmes, M.; Zhang, W.; Huang, X.; Cui, X.; Xue, J. Determination of retrogradation degree in starch by mid-infrared and raman spectroscopy during storage. *Food Anal. Method* **2017**, *10*, 3694–3705. [[CrossRef](#)]
16. Wu, X.; Zhou, J.; Wu, B.; Sun, J.; Dai, C. Identification of tea varieties by mid-infrared diffuse reflectance spectroscopy coupled with a possibilistic fuzzy c-means clustering with a fuzzy covariance matrix. *J. Food Process Eng.* **2019**, *42*, e13298. [[CrossRef](#)]
17. Wu, M.; Sun, J.; Lu, B.; Ge, X.; Zhou, X.; Zou, M. Application of deep brief network in transmission spectroscopy detection of pesticide residues in lettuce leaves. *J. Food Process Eng.* **2019**, *42*, e13005. [[CrossRef](#)]

18. Zhou, X.; Sun, J.; Lu, B.; Wu, X.; Dai, C.; Yang, N. Study on pesticide residues classification of lettuce leaves based on polarization spectroscopy. *J. Food Process Eng.* **2018**, *41*, e12903.
19. Sun, J.; Zhou, X.; Mao, H.; Wu, X.; Zhang, X.; Gao, H. Identification of pesticide residue level in lettuce based on hyperspectra and chlorophyll fluorescence spectra. *Int. J. Agric. Biol. Eng.* **2016**, *9*, 231–239.
20. Sun, J.; Cong, S.; Mao, H.; Wu, X.; Yang, N. Quantitative detection of mixed pesticide residue of lettuce leaves based on hyperspectral technique. *J. Food Process Eng.* **2017**, *41*, e12654. [[CrossRef](#)]
21. Sun, J.; Zhou, X.; Mao, H.; Wu, X.; Zhang, X.; Li, Q. Discrimination of pesticide residues in lettuce based on chemical molecular structure coupled with wavelet transform and near infrared hyperspectra. *J. Food Process Eng.* **2016**, *40*, e12509. [[CrossRef](#)]
22. Wu, X.; Zhu, J.; Wu, B.; Huang, D.; Sun, J.; Dai, C. Classification of Chinese vinegar varieties using electronic nose and fuzzy Foley–Sammon transformation. *J. Food Sci. Tech.* **2020**, *57*, 1310–1319. [[CrossRef](#)]
23. Guidea, A.; Găceanu, R.D.; Pop, H.F.; Sârbu, C. Mineral waters classification using fuzzy linear discriminant analysis. *Studia Univ. Babeş Bolyai Chem.* **2020**, *65*, 45–56. [[CrossRef](#)]
24. Wu, X.; Wu, B.; Zhou, J. Fuzzy uncorrelated discriminant transformation and its application. *J. Image Graph.* **2009**, *14*, 1832–1836.
25. Lu, Y.; Qu, Y.; Song, M. Research on the correlation chart of near infrared spectra by using multiple scatter correlation technique. *Spectrosc. Spect. Anal.* **2007**, *27*, 877–880.
26. Shi, Z.; Chow, C.W.K.; Fabris, R.; Liu, J.; Jin, B. Alternative particle compensation techniques for online water quality monitoring using UV-Vis spectrophotometer. *Chemom. Intell. Lab. Syst.* **2020**, *204*, 104074. [[CrossRef](#)]
27. Foley, D.H.; Sammon, J.W. An optimal set of discriminant vectors. *IEEE Trans. Comput.* **1975**, *24*, 281–289. [[CrossRef](#)]
28. Jin, Z.; Yang, J.; Hu, Z.; Lou, Z. Face recognition based on the uncorrelated discriminant transformation. *Pattern Recogn.* **2001**, *34*, 1405–1416. [[CrossRef](#)]
29. Xu, Y.; Yang, J.; Zhong, J. Theory analysis on FSLDA and ULDA. *Pattern Recogn.* **2003**, *36*, 3031–3033. [[CrossRef](#)]
30. Wu, X.; Kumar, V.; Quinlan, J.R.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.; Liu, B.; Yu, P.; et al. Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [[CrossRef](#)]
31. Yang, Y.; Zhang, X.; Yin, J.; Yu, X. Rapid and nondestructive on-site classification method for consumer-grade plastics based on portable NIR spectrometer and machine learning. *J. Spectrosc.* **2020**, *2020*, 6631234. [[CrossRef](#)]
32. He, Y.; Zhao, Y.; Zhang, C.; Li, Y.; Bao, Y.; Liu, F. Discrimination of grape seeds using laser-induced breakdown spectroscopy in combination with region selection and supervised classification methods. *Foods* **2020**, *9*, 199. [[CrossRef](#)] [[PubMed](#)]