Original Article
Food Science and Technology

ISSN 0101-2061 (Print)
ISSN 1678-457X (Online)

DOI: https://doi.org/10.1590/fst.55822

Identification of pesticide residues on black tea by fluorescence hyperspectral technology combined with machine learning

Jie SUN[1], Yan HU[1], Yulin ZOU[1], Jinping GENG[1], Youli WU[1], Rongsheng FAN[1], Zhiliang KANG[1]*

Abstract

Black tea has a long history in China, but in export trade, pesticide residues often exceed the standard. To obtain a rapid, accurate, and non-destructive identification method of pesticide residues on black tea, the fluorescence hyperspectral data of dry black tea sprayed with distilled water and six pesticides were collected in this study. The spectra were preprocessed by multiplicative scatter correction (MSC) and standard normal variate (SNV). Then the uninformative variable elimination (UVE), successive projections algorithm (SPA), competitive adaptive re-weighted sampling (CARS), UVE-SPA, and CARS-SPA were used to feature extraction. This study proposes a machine learning model composed of a one-dimensional convolutional neural network backbone (1D CNN backbone) and a random forest classifier (RF classifier) to identify pesticide residues on black tea, and the 1D CNN-RF model was compared with three other machine learning models (support vector machine, RF, and 1D CNN). The results show that MSC-CARS-SPA-1D CNN-RF is the best model for identifying pesticide residues on black tea in which the accuracy of the test set is 99.05%. Combined with fluorescence hyperspectral technology, the proposed 1D CNN-RF model has great potential in the non-destructive identification of pesticide residues on black tea.

Keywords: black tea; pesticide residues; fluorescence hyperspectral; convolutional neural network; random forest; non-destructive identification.

Practical Application: Non-destructive identification of pesticide residues on tea.

1 Introduction

As a kind of functional health drink, tea is loved by people all over the world (Graham, 1992). China is a traditional tea production country, which has a large number of tea varieties and production ranks first in the world (He et al., 2021a). Black tea is an important part of China's tea industry, its yield accounts for 9.2%. Properly drinking black tea is beneficial to the human body, which can enhance human immunity, anti-aging, anti-oxidation, and even prevent cancer (Sun et al., 2007; Higdon & Frei, 2003). The tea economy is an important economic pillar in Ya'an. As of 2021, the area of tea gardens in Ya'an reached 6600 ha, accounting for 1/5 of the total tea planting area in Sichuan Province. The comprehensive output value of tea exceeded 19 billion RMB, and the number of tea farmers reached 400,000. Because of the long-term occurrence of tea plant diseases and insect pests, some organic pesticides such as bifenthrin (BFT), diafenthiuron (DFT), tolfenpyrad (TFP), and imidacloprid (IDC) are widely used to ensure the high yield of tea (Xu et al., 2021; Xin-Zhong et al., 2011; Bai et al., 2021; Chen et al., 2020). At the same time, in the process of spraying pesticides, tea farmers often mix BFT and DFT or TFP and IDC to enhance the effect. However, due to the nonstandard operation, misuse, and abuse of pesticides occurring in the process of pesticide use, tea is often detected with pesticide residues or even exceeds the limit standard (Zhu et al., 2021). The large use of the above pesticides will inevitably adhere pesticide residues to the surface of tea leaves and it is difficult to remove them by conventional cleaning (Chen et al., 2013). Long-term drinking of tea containing pesticide residues will cause certain damage to the body (Sun et al., 2021).

As food safety has attracted more and more attention from consumers, countries or regions such as Morocco and the European Union have formulated strict standards for pesticide residues in tea (Le et al., 2021). At present, the identification methods of pesticide residues in crops are mainly liquid chromatography-mass spectrometry (Fan et al., 2021; Xing et al., 2022), gas chromatography-mass spectrometry (Oyekunle et al., 2021), and immunoassay (Garcia-Febrero et al., 2014). These identification methods have the advantages of high sensitivity and wide identification range, but some methods are too expensive, some are too complex and time-consuming, and some can cause irreversible damage to the test samples. Therefore, it is of great practical value to establish a rapid, accurate, and non-destructive method for the identification of pesticide residues on tea. At present, spectral analysis has become a hot spot in the identification of pesticide residues due to its excellent performance. The most widely used methods are near-infrared spectroscopy (He et al., 2021b; Li et al., 2021), Raman spectroscopy (Zhu et al., 2021; Zhang et al., 2021a), terahertz time-domain spectroscopy (Chen et al., 2015), laser-induced breakdown spectroscopy (Martino et al., 2021; Wu et al., 2019),

Received 02 May, 2022
Accepted 22 June, 2022
[1] *College of Mechanical and Electrical Engineering, Sichuan Agricultural University, Ya'an, China*
*Corresponding author: 12200@sicau.edu.cn

and hyperspectral imaging (Jia et al., 2018; Sun et al., 2018). Hyperspectral imaging technology combined with fluorescence emission technology has played an important role in the field of real-time online quality and safe non-destructive testing of agricultural products (Mahmudiono et al., 2022). The fluorescence hyperspectral imaging system was used to identify the pesticide residues on the surface of black tea in this study.

In recent years, machine learning is often been used in the field of food safety detection, such as the classification of different quality crops (Li et al., 2020; Chen & Yu, 2022; Hou et al., 2022), and the identification and classification of pesticide residues on the surface of crops (Zhu et al., 2021). A one-dimensional convolutional neural network (1D CNN) is a widely used machine learning algorithm, and its biggest advantage is that it can automatically extract features from the input without manual selection (Chatzidakis & Botton, 2019; Zhang et al., 2019). By adjusting and optimizing the parameters of the convolution layer, 1D CNN can better identify and classify the input data. In 2021, He et al. proposed a 1D CNN machine learning model to recognize pesticide residues on garlic chive leaves, which used the hyperspectral imaging system to collect the short-wave infrared hyperspectral images of garlic chive leaves sprayed with distilled water or pesticides, and the 1D CNN model was compared with other models. Finally, the accuracy rate of the 1D CNN model reached 98.5% (He et al., 2021b). However, due to the lack of input data, the 1D CNN model may have over-fitting, which is not explained.

This study proposes a new method for the identification of pesticide residues on the surface of Ya'an black tea based on fluorescence hyperspectral technology and machine learning. The newly proposed machine learning method consists of a 1D CNN backbone and an RF classifier, namely the 1D CNN-RF model. The backbone of the 1D CNN extracts features from the input data and then classifies it through the RF classifier. Comparing the 1D CNN-RF pesticide residue identification model with other models, the proposed 1D CNN-RF model has high accuracy and strong robustness. A fluorescence hyperspectral imaging system was used to explore the influence of pesticide residues on spectral data. The results show that the method proposed in this study is feasible for the identification of pesticide residues on black tea, which provides the application basis of fluorescence hyperspectral technology in the identification of pesticide residues on black tea, and provides a new idea for the non-destructive identification of pesticide residues on Ya'an black tea.

2 Materials and methods

2.1 Experimental materials

Enough wild black tea was collected outdoors to ensure that the samples were free of pesticide residues. The finished black tea was prepared through the tea production process for test use. The pesticides used in the study were bifenthrin (BFT, active ingredient content is 2.5%), diafenthiuron (DFT, active ingredient content is 25%), tolfenpyrad (TFP, active ingredient content is 10%), imidacloprid (IDC, active ingredient content is 1.5%), all purchased from regular sales channels.

First, configure the pesticide solution, the appropriate amount of distilled water was prepared by using an electrothermal distilled water device (YA·ZD·5, Shanghai Baolan Experimental Instrument Co., Ltd., China), and then BFT, DFT, the mixture of BFT and DFT, TFP, IDC, and the mixture of TFP and IDC were diluted with distilled water at a ratio of 1:500 (the recommended concentration for agriculture) by using a pipette (Hunan Lichen Instrument Technology Co., Ltd., China). The prepared different pesticide solutions were sprayed equally on the surface of black tea (one group of samples was sprayed with distilled water as the blank group). After spraying, the samples were naturally dried under laboratory conditions for 24 hours, and then data collection was carried out. The sample collection process is shown in Figure 1, and the experimental sample situation is shown in Table 1.

Table 1. Overview of experimental samples.

Category	Quantity	Solution
Blank	60	0
BFT	60	1:500
DFT	60	1:500
The mixture of BFT and DFT	60	1:500
TFP	60	1:500
IDC	60	1:500
The mixture of TFP and IDC	60	1:500

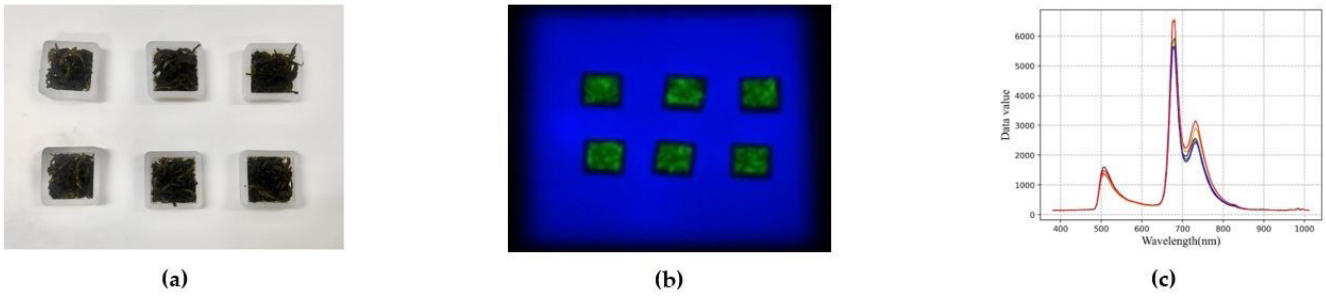

Figure 1. Collection of black tea samples image by fluorescence hyperspectral imaging system. (a) six black tea samples on each whiteboard; (b) fluorescence hyperspectral image of samples; (c) fluorescence hyperspectral data of samples.

2.2 Fluorescence hyperspectral data extraction

The fluorescence hyperspectral data of black tea samples were acquired using a GaiaFluo(/Pro)-VN-HR series of fluorescence hyperspectral test system manufactured by Sichuan Dualix Spectral Image Technology Co. Ltd. The schematic diagram of the fluorescence hyperspectral imaging system is shown in Figure 2. Different from most traditional optical imaging methods that can only capture gray or color images of samples, fluorescence hyperspectral can simultaneously obtain spatial characteristic imaging and spectral information of the tested samples. The spectral resolution of the hyperspectral camera is 2.8 nm, the sampling rate is 0.6 nm, the pixel size of the SCMOS detector is 6.5 um, and the size is 2048 × 2048. In this study, a xenon lamp was used as the excitation light source, and the detection range was 250 nm to 1100 nm (Hu et al., 2021), the 390 nm excitation filter and the 495 nm fluorescence filter were selected to act on the fluorescence hyperspectral imaging system, the exposure time was set to 800 ms, and the scanning speed was set at 0.06485 cm/s. Put the samples prepared in Section 2.1 into the sample table in batches. The fluorescence hyperspectral instrument was used to obtain a fluorescence image, the image contains 377 nm to 1011 nm spectral data. After obtaining the sample fluorescence hyperspectral image, ENVI 5.3 (Wang et al., 2020) software was used to obtain the sample-specific fluorescence hyperspectral data.

2.3 Spectral data preprocessing

Spectral data collected by the fluorescence hyperspectral system is easy to be affected by the surrounding environment and itself. To reduce the noise doped in the extracted spectral data and improve the sensitivity of spectral resolution, it is necessary to carry out reasonable spectral data preprocessing.

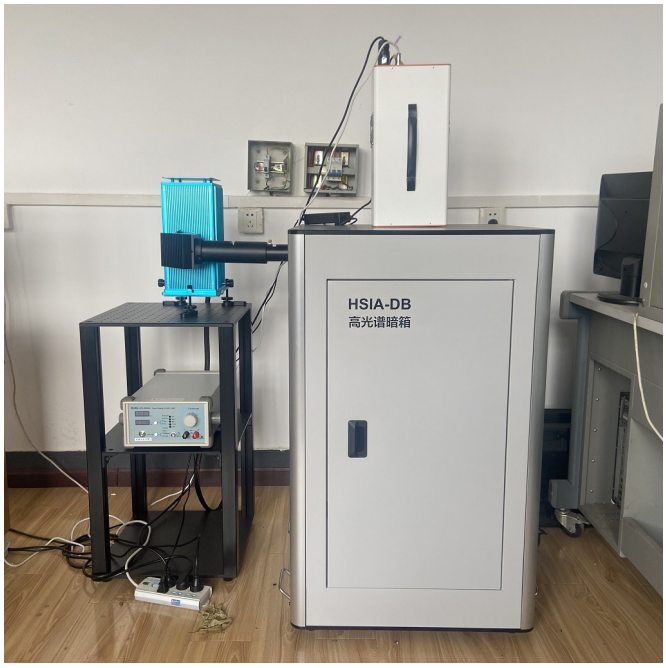

Figure 2. Fluorescence hyperspectral system.

In this study, two preprocessing algorithms, Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) were used to preprocess the spectrum (Zhang et al., 2021b). The basic principle of MSC is to correct the baseline shift and offset phenomenon of spectral data through an ideal spectrum, thereby effectively eliminating spectral differences caused by different scattering levels and enhancing the correlation between spectrum and data, and the basic principle of SNV is to standardize the original spectral data to correct the spectral error caused by light scattering. The modeling results of pesticide residues identification models of the two preprocessing algorithms were compared in Section 4.1, and the best one was selected as the preprocessing algorithm for the spectral data.

2.4 Feature extraction

Due to the large amount of full-spectrum data collected by the fluorescence hyperspectral imaging system, there is a certain amount of noise effect between each spectral segment and the linear correlation between the data, which will affect the performance of the pesticide residues identification model. Therefore, it is necessary to reduce the dimension of the raw spectral data. This study used the method of extracting the characteristic spectral segment to reduce the dimension.

Non-destructive identification of common crops based on the hyperspectral imaging system mostly uses single algorithms such as Uninformative Variable Elimination (UVE), Successive Projections Algorithm (SPA), or Competitive Adaptive Re-Weighted Sampling (CARS) to extract features and establish identification models (Yuan et al., 2021). However, the single algorithm lacks certain stability when extracting features. Therefore, this study used the SPA algorithm to perform quadratic dimension reduction on the feature segment variables extracted by UVE and CARS algorithms to eliminate the collinearity between variables and improve the stability of the algorithm. The preprocessed fluorescence hyperspectral data of non-pesticides samples and six samples with different types of pesticides were used as input data sets for extracting characteristic spectral segments. Five algorithms including UVE, SPA, CARS, UVE-SPA, and CARS-SPA were used for feature extraction, and the extracted characteristic spectrum was used for the pesticide residues identification experiments.

2.5 Model selection

In this study, three classic machine learning classification models, namely SVM, RF, and 1D CNN, were selected for comparative analysis with the proposed 1D CNN-RF model. The performances of machine learning models were evaluated using accuracy, recall, precision, and F1 score. All algorithms in this study ran on CPU(Intel(R) Core (TM) i5-10400F CPU@ 1.60 GHz), and have been implemented on Python 3.8.

One-dimensional convolutional neural network (1D CNN)

Among the machine learning algorithms, CNN is the most common and widely used one and has achieved excellent results in image identification and processing (Liang et al., 2021), speech identification (Kwon, 2019), and other fields. CNN is a

feedforward neural network, which processes data in multiple matrices by extracting local and global features (LeCun et al., 2015). The core of 1D CNN is to use convolution operation, and the structure of its model can be changed, that is, the depth of the model will change.

Random forest (RF)

RF is a classifier that uses multiple trees to train and predict samples, this classifier was first proposed by Leo Breiman (Farah et al., 2021). The basic unit of RF is the decision tree, the decision tree is composed of nodes and directed edges, and the nodes are divided into internal nodes and leaf nodes. The internal nodes represent a feature or attribute, and the leaf nodes represent a class. Each decision tree is a classifier. Many decision trees are randomly combined into a forest, and the parameters of each tree are different. Then, the predicted results of each tree are averaged, so that the working effectiveness of decision trees can be retained, and the risk of overfitting can be reduced.

Support vector machine (SVM)

SVM is a machine learning method based on the Vapnik-Chervonenkis dimension theory of statistical learning theory and the principle of structural risk minimization (Zagajewski et al., 2021). At the same time, SVM is a commonly used data mining method, which is suitable for regression problems and pattern identification and is also a training learning method suitable for a small number of samples. When the sample input set of the SVM model is large, SVM classification is slow and inefficient, and because SVM does not use the classical probability theory, the whole training process is relatively simple, which may lead to the training efficiency is not ideal.

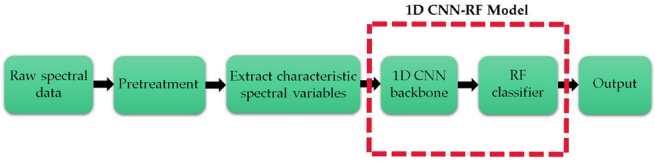

Figure 3. 1D CNN-RF structure diagram.

The proposed method

Based on the excellent performance of the 1D CNN model in extracting the features of one-dimensional data such as fluorescence spectral data, the 1D CNN model was selected to mine the deep features of the input data information in this study. However, when the amount of the input set data is too small, the 1D CNN model is prone to overfitting. The RF algorithm reduces the variance of the model through the collection of different decision trees, so it has good generalization ability and anti-overfitting ability. After the advantages and disadvantages of 1D CNN and RF were weighed, the 1D CNN-RF framework was proposed, the specific 1D CNN-RF framework is shown in Figure 3. The 1D CNN model deeply mines the information features of pesticide residues on black tea from the fluorescence spectral data, and the RF model classifies the new spectral representative feature data generated by the 1D CNN model to achieve better results.

In the feature extraction part of the 1D CNN model, after repeated training and modification, the final 1D CNN structure was obtained. The model structure of 1D CNN is shown in Figure 4. The structure and parameters of the optimized 1D CNN model are shown in Table 2. The candidate values of the frame structure and parameters of the 1D CNN model proposed in this study are based on the research results of previous scholars (Wu et al., 2018; Qiu et al., 2018). The characteristic mapping value of the convolution layer output is nonlinearly transformed. In this study, the modified linear unit (Relu) function was selected as the activation function to activate the outputs of these layers and the extract features.

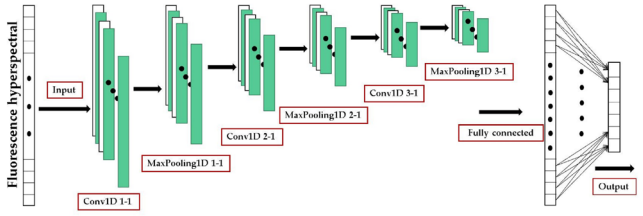

Figure 4. One-dimensional Convolutional Neural Network Architecture.

Table 2. The specifications of the optimized 1D CNN model show the obtained best parameter values.

Layer	Kernel	Feature Maps	Stride	Activation	Layer Output Shape
Input	-	-	-	-	1 × N
Con1D 1-1	1 × 3	8	1	ReLu	(N-2) × 8
MaxPooling1D 1-1	1 × 2	8	2	-	(N-2)/2 × 8
Con1D 2-1	1 × 3	16	1	ReLu	(N-2)/2-2 × 16
MaxPooling1D 2-1	1 × 2	16	2	-	((N-2)/2-2)/2 × 16
Con1D 3-1	1 × 3	32	1	ReLu	((N-2)/2-2)/2-2 × 32
MaxPooling1D 3-1	1 × 2	32	2	-	(((N-2)/2-2)/2-2)/2 × 32
Flatten	-	-	-	-	1 × ((((N-2)/2-2)/2-2)/2×32)
Dense	-	128	-	ReLu	1 × 128
Output	-	-	-	Softmax	1 × 2

N: the number of bands input by the model.

To get the best performance of the model, it is necessary to adjust the number of iterations of the 1D CNN model. The relationship between the accuracy and loss rate of the 1D CNN model and the number of iterations is shown in Figure 5. After 200 epochs, the accuracy and loss rate of the training set and validation set tended to be stable with small fluctuations, the loss rate tended to 0, and accuracy tended to 1, which means that the model training is completed. Therefore, this study set the epoch parameter of the 1D CNN model as 200. To achieve both efficiency and effect, the Adam optimizer in the Keras deep learning library was selected in this experiment to dynamically reduce the learning rate, and the learning rate parameter was 0.001. The Adam algorithm uses the first-order moment estimation and second-order moment estimation of the gradient to dynamically adjust the learning rate of each parameter.

The raw spectral data obtained by the fluorescence hyperspectral imaging system was preprocessed and the spectral data after extracting the feature spectral segment variables were used as the input of the 1D CNN framework in the 1D CNN-RF model. The 1D CNN model was used to mine the deep characteristic spectral data. After the feature extraction was completed, the classifier in the 1D CNN model was replaced by the RF classifier, and the classification was carried out by the RF classifier, namely, which is the 1D CNN-RF model proposed in this study.

3 Results and discussion

3.1 Fluorescence spectral data preprocessing

To eliminate the interference of unfavorable factors in the spectrum, it is necessary to preprocess the raw spectrum data. Two methods, MSC and SNV were selected to preprocess the spectral data of black tea. In the sample data of distilled water and six different pesticides with solutions of 1: 500, 3/4 were randomly selected as the training set, and the remaining 1/4 were selected as the test set. The corresponding SVM and 1D CNN identification models were established, and the optimal preprocessing method is selected through the model training results. The results of the SVM and 1D CNN identification model are shown in Table 3.

It can be seen from Table 3 that the accuracy of the test set obtained by SVM and 1D CNN for the raw data is less than 90%, respectively. The spectral data preprocessed by MSC or SNV are used as input sets of SVM and 1D CNN models, the accuracy of the test set is higher than or equal to that of the raw spectral data, which is because both can effectively eliminate the influence of scattering effect. Comparing MSC and SNV, when the data preprocessed by MSC were used as input, the accuracy of the SVM model increased from 80.95% to 84.76%, and the accuracy of the 1D CNN model increased from 89.52% to 92.38%, when the data preprocessed by SNV were used as input, the accuracy of SVM identification model increased from 80.95% to 82.86%, and the accuracy of 1D CNN identification model increased from 89.52% to 91.43%. The spectral data after MSC preprocessing are better than those after SNV preprocessing, so MSC is selected as the data preprocessing method for pesticide species identification in this study.

3.2 Characteristic spectral extraction

If the 125 preprocessed spectral data are used as input sets, the information redundancy may be generated, which makes it difficult for the identification model to find the spectral feature information in input data, thus leading to the low accuracy of identification models. Therefore, it is necessary to extract the characteristic spectral segments and reduce the amount of input data to improve the efficiency of the identification model.

In this study, five methods including UVE, SPA, CARS, UVE-SPA, and CARS-SPA were used to extract the characteristic spectral

Table 3. Identification results based on different models of various preprocessing methods.

Methods	Models of Identification	
	SVM	1D CNN
Raw	0.8095	0.8952
MSC	0.8476	0.9238
SNV	0.8286	0.9143

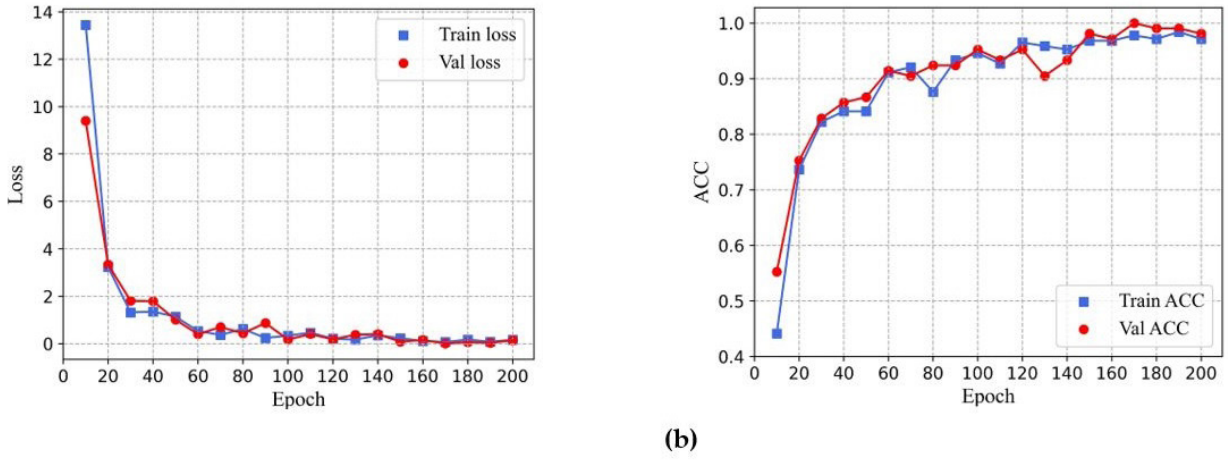

Figure 5. Training process of 1D CNN contains (a) the loss change and (b) the accuracy change on the train set and validation (Val) set.

segments. The fluorescence hyperspectral data preprocessed by MSC of distilled water samples and other six kinds of pesticide samples were used as the data set to extract the characteristic spectral segments. The distribution results of the characteristic spectral segments are shown in Figure 6.

Among them, 51 characteristic spectral bands were extracted by UVE, 37 by SPA, 57 by CARS, 27 by UVE-SPA, and 28 by CARS-SPA. The characteristic spectral segments extracted by the above five methods were used as the input of the pesticide identification models, optimal pesticide identification model was obtained based on model training results.

3.3 Comparison of the identification results of four models

In this study, four algorithms (SVM, RF, 1D CNN, and the proposed 1D CNN-RF) were selected to establish the identification model of pesticide residues on the surface of black tea, and the different characteristic spectrum segments extracted by the five methods selected in the previous section were used to model and identify. The classification performance of the four models was evaluated according to four evaluation indexes (accuracy, recall, precision, and F1 score), the obtained model training results are shown in Table 4.

It can be seen from Table 3 and Table 4 that the indexes of identification models were improved after MSC preprocessing

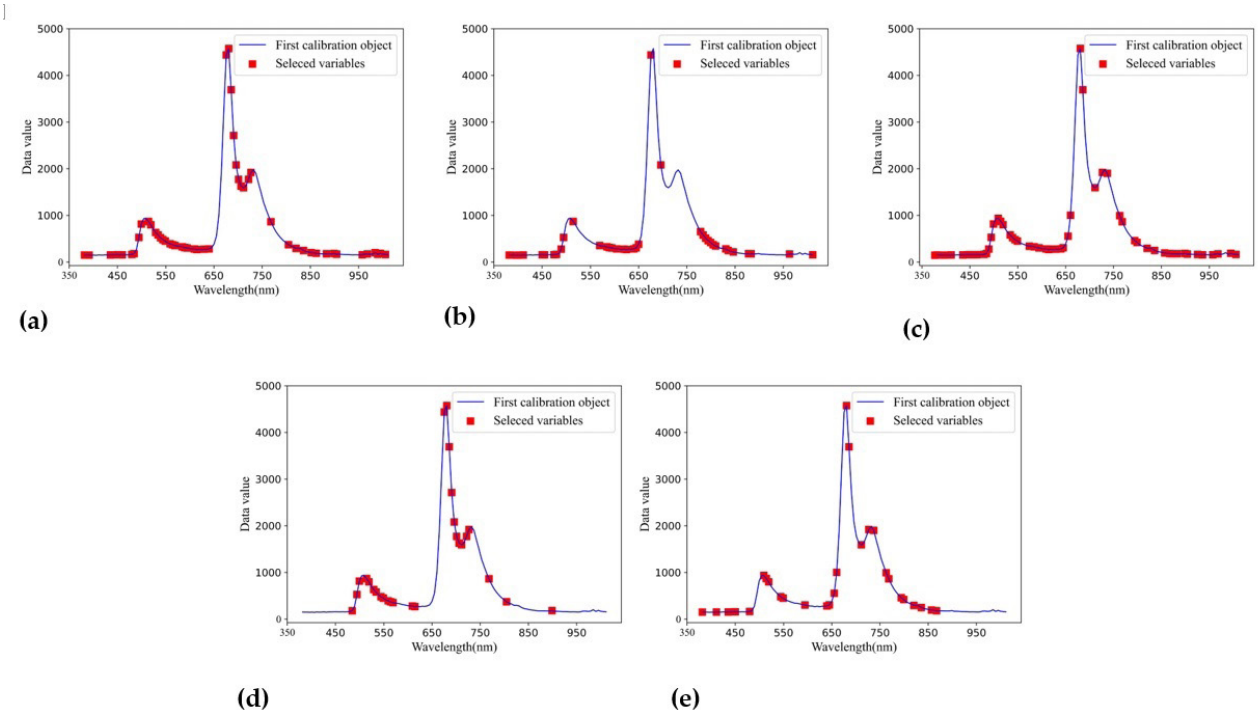

Figure 6. Characteristic spectrum distribution maps of five methods. (a) UVE; (b) SPA; (c) CARS; (d) UVE-SPA; (e) CARS-SPA.

Table 4. Identification results of test set samples by four models.

Model	Extraction Method	Accuracy	Recall	Precision	F1 score
SVM	UVE	0.8571	0.8571	0.8979	0.8111
	SPA	0.8571	0.8571	0.8890	0.8520
	CARS	0.8667	0.8667	0.9310	0.8754
	UVE-SPA	0.8857	0.8857	0.9087	0.8779
	CARS-SPA	0.9048	0.9048	0.9195	0.8955
RF	UVE	0.8476	0.8476	0.8832	0.8410
	SPA	0.8571	0.8571	0.8665	0.8528
	CARS	0.8762	0.8762	0.8961	0.8680
	UVE-SPA	0.8667	0.8667	0.9095	0.8563
	CARS-SPA	0.9048	0.9048	0.9071	0.9050
1D CNN	UVE	0.9238	0.9238	0.9373	0.9237
	SPA	0.9429	0.9429	0.9554	0.9444
	CARS	0.9714	0.9714	0.9743	0.9706
	UVE-SPA	0.9429	0.9429	0.9496	0.9435
	CARS-SPA	0.9714	0.9714	0.9726	0.9714
The proposed 1D CNN-RF	UVE	0.9429	0.9429	0.9459	0.9430
	SPA	0.9524	0.9524	0.9563	0.9521
	CARS	0.9810	0.9810	0.9815	0.9809
	UVE-SPA	0.9714	0.9714	0.9730	0.9714
	CARS-SPA	0.9905	0.9905	0.9911	0.9905

and five algorithms (UVE, SPA, CARS, UVE-SPA, CARS-SPA) extracting characteristic spectrum. The results show that the identification accuracy and precision of pesticide residues on black tea could be improved to some extent by extracting the effective information of the whole spectral data and reducing the influence of other irrelevant noise.

In terms of pesticide species identification, the accuracy of SVM and RF was lower than 91%, the identification results were not satisfactory. The overall performances of the identification models of 1D CNN and 1D CNN-RF were significantly better than that of SVM and RF models. As shown in Table 4, the model constructed by 1D CNN-RF showed the best performance on the test set, and the identification accuracy and precision of the 1D CNN-RF model were higher than those of other algorithms through longitudinal comparison. The identification rates of the 1D CNN-RF model based on five feature extraction algorithms were above 94%, compared with other methods, the CARS-SPA-1D CNN-RF model achieved the best performance on the test set, with an accuracy of 90.05%, a recall of 99.05%, a precision of 99.11%, and an F1 score of 99.05%. Based on the above results and analysis, the optimal model for identifying pesticide residues on black tea is the 1D CNN-RF model based on CARS-SPA. The identification distribution of pesticide residues on black tea by the 1D CNN-RF model is shown in Figure 7. It can be seen from Figure 7 that after the characteristic spectrum extracted by CARS-SPA was used as the input of the 1D CNN-RF model, the identification rates were significantly improved, and only one sample with the mixture of BFT and DFT was identified as one TFP sample. The results showed that the 1D CNN-RF model based on CARS-SPA can well identify whether black tea is sprayed with pesticides and which kind of pesticides is sprayed.

The above results showed that the combination of fluorescence hyperspectral technology with the 1D CNN-RF model could accurately distinguish whether the pesticide residues were on the surface of black tea, and could effectively identify the pesticide residues of BFT, DFT, mixtures of BFT and DFT, TFP, IDC, and mixtures of TFP and IDC on the surface of black tea, which provided a new method for the non-destructive identification of pesticide residues on Ya'an black tea.

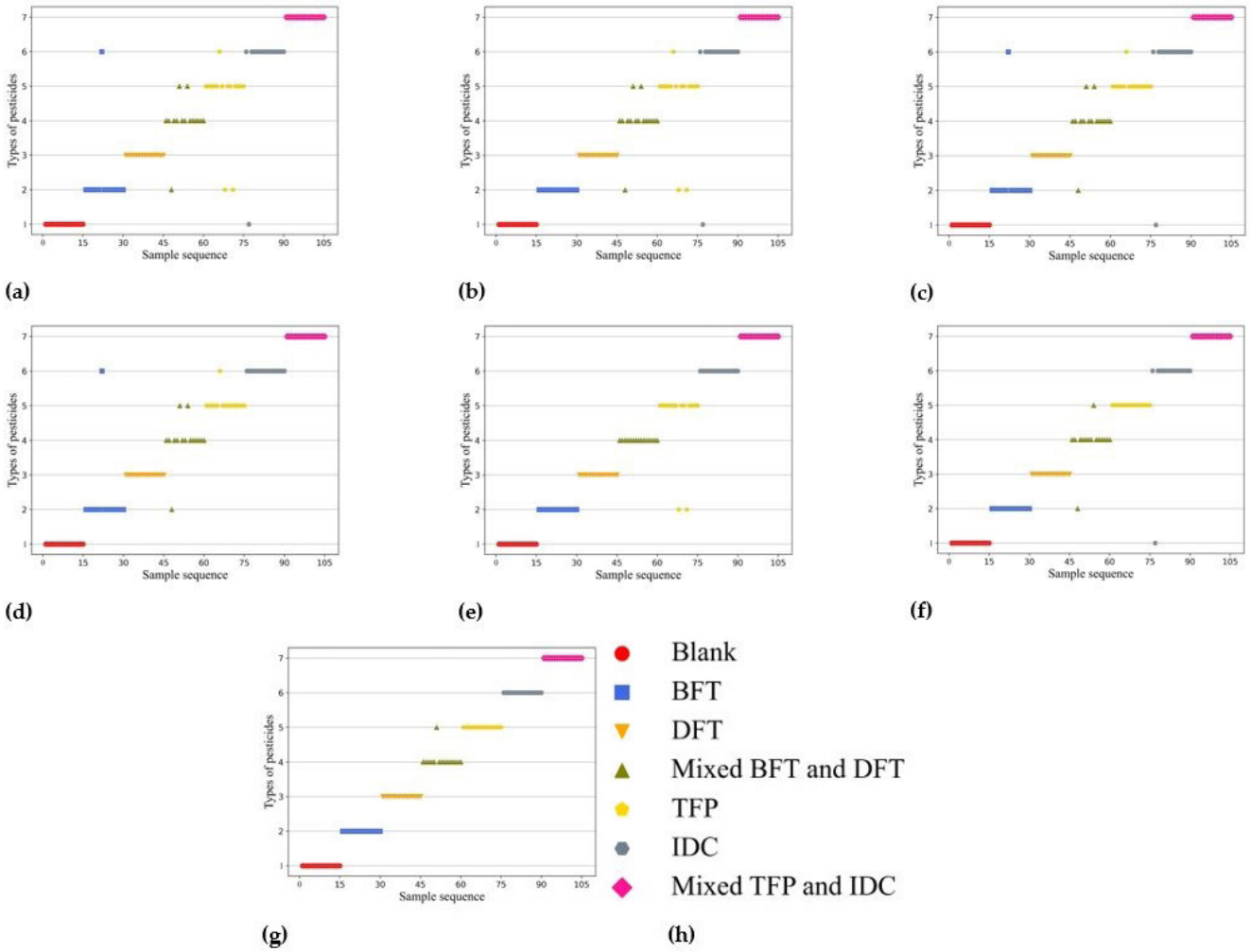

Figure 7. 1D CNN-RF pesticide residues identification model training results based on: (a) raw spectral data without extracting characteristic spectrum; (b) MSC; (c) UVE; (d) SPA; (e) CARS; (f) UVE-SPA; (g) CARS-SPA. (h) labels.

4 Conclusions

In this study, a new 1D CNN-RF framework was proposed for the non-destructive identification of pesticide residues on the surface of black tea based on fluorescence hyperspectral technology.

This study collected the fluorescence hyperspectral images of Ya'an black tea sprayed with distilled water, BFT, DFT, mixtures of BFT and DFT, TFP, IDC, and mixtures of TFP and IDC. The raw spectral data were preprocessed by MSC and SNV, and then the characteristic spectral segments were extracted by three one-dimensionality reduction methods (UVE, SPA, CARS), and two secondary combinations of dimensionality reduction methods (UVE-SPA, CARS-SPA). Based on the extracted characteristic spectral segment variables, four identification models of SVM, RF, 1D CNN, and 1D CNN-RF were established. The comparison results showed that the 1D CNN-RF identification model based on CARS-SPA had the best performance: accuracy of 99.05%, recall of 99.05%, precision of 99.11%, and F1 score of 99.05%. This was because the RF model made up for the overfitting of the 1D CNN model when the sample size was small. In conclusion, the proposed 1D CNN-RF model based on CARS-SPA has great potential in pesticide species identification, which provides a non-destructive, accurate, and fast pesticide residue identification method for food quality and safety.

The results show that fluorescence hyperspectral technology combined with machine learning is a feasible method for identifying pesticide residues. At the same time, combined with the analysis of experimental data, the proposed 1D CNN-RF model showed excellent performance in black tea pesticide residue identification, it can be obtained that this method can get better results than traditional methods. In the future, we will increase the amount of sample data and the number of pesticide species based on this experiment, and strive to develop a highly targeted and real-time online rapid non-destructive identification system for pesticide residues.

Funding:

This research was funded by the subject double support program of Sichuan Agricultural University (Grant No. 035-1921993093).

References

Bai, A., Chen, A., Chen, W., Liu, S., Luo, X., Liu, Y., & Zhang, D. (2021). Residue behavior, transfer and risk assessment of tolfenpyrad, dinotefuran and its metabolites during tea growing and tea brewing. *Journal of the Science of Food and Agriculture*, 101(14), 5992-6000. http://dx.doi.org/10.1002/jsfa.11253. PMid:33851415.

Chatzidakis, M., & Botton, G. A. (2019). Towards calibration-invariant spectroscopy using deep learning. *Scientific Reports*, 9(1), 2126. http://dx.doi.org/10.1038/s41598-019-38482-1. PMid:30765890.

Chen, J. Y., Lin, Y. J., & Kuo, W. C. (2013). Pesticide residue removal from vegetables by ozonation. *Journal of Food Engineering*, 114(3), 404-411. http://dx.doi.org/10.1016/j.jfoodeng.2012.08.033.

Chen, J., Huang, X., Wang, L., Ma, C., Wu, S., & Wang, H. (2020). Dummy template surface molecularly imprinted polymers based on silica gel for removing imidacloprid and acetamiprid in tea polyphenols. *Journal of Separation Science*, 43(12), 2467-2476. http://dx.doi.org/10.1002/jssc.201901268. PMid:32212382.

Chen, T., & Yu, S. (2022). Research on food safety sampling inspection system based on deep learning. *Food Science and Technology*, 42, e29121. http://dx.doi.org/10.1590/fst.29121.

Chen, Z., Zhang, Z., Zhu, R., Xiang, Y., Yang, Y., & Harrington, P. (2015). Application of terahertz time-domain spectroscopy combined with chemometrics to quantitative analysis of imidacloprid in rice samples. *Journal of Quantitative Spectroscopy & Radiative Transfer*, 167, 1-9. http://dx.doi.org/10.1016/j.jqsrt.2015.07.018.

Fan, S. F., Ma, J. M., Cao, M. R., Wang, J., Zhang, L. L., Zhang, Y., Li, Q., & Chen, J. (2021). Simultaneous determination of 15 pesticide residues in Chinese cabbage and cucumber by liquid chromatography-tandem mass spectrometry utilizing online turbulent flow chromatography. *Food Science and Human Wellness*, 10(1), 78-86. http://dx.doi.org/10.1016/j.fshw.2020.06.003.

Farah, J. S., Cavalcanti, R. N., Guimarães, J. T., Balthazar, C. F., Coimbra, P. T., Pimentel, T. C., Esmerino, E. A., Duarte, M. C. K. H., Freitas, M. Q., Granato, D., Neto, R. P. C., Tavares, M. I. B., Calado, V., Silva, M. C., & Cruz, A. G. (2021). Differential scanning calorimetry coupled with machine learning technique: An effective approach to determine the milk authenticity. *Food Control*, 121, 107585. http://dx.doi.org/10.1016/j.foodcont.2020.107585.

Garcia-Febrero, R., Salvador, J. P., Sanchez-Baeza, F., & Marco, M. P. (2014). Rapid method based on immunoassay for determination of paraquat residues in wheat, barley and potato. *Food Control*, 41, 193-201. http://dx.doi.org/10.1016/j.foodcont.2014.01.008.

Graham, H. N. (1992). Green tea composition, consumption, and polyphenol chemistry. *Preventive Medicine*, 21(3), 334-350. http://dx.doi.org/10.1016/0091-7435(92)90041-F. PMid:1614995.

He, J., Li, H., Zhang, L., Zhi, X., Li, X., Wang, X., Feng, Z., Shen, G., & Ding, X. (2021a). Silver microspheres aggregation-induced Raman enhanced scattering used for rapid detection of carbendazim in Chinese tea. *Food Chemistry*, 339, 128085. http://dx.doi.org/10.1016/j.foodchem.2020.128085. PMid:33152876.

He, W., He, H., Wang, F., Wang, S., & Lyu, R. (2021b). Non-destructive detection and recognition of pesticide residues on garlic chive (Allium tuberosum) leaves based on short wave infrared hyperspectral imaging and one-dimensional convolutional neural network. *Journal of Food Measurement and Characterization*, 15(5), 4497-4507. http://dx.doi.org/10.1007/s11694-021-01012-7.

Higdon, J. V., & Frei, B. (2003). Tea catechins and polyphenols: health effects, metabolism, and antioxidant functions. *Critical Reviews in Food Science and Nutrition*, 43(1), 89-143. http://dx.doi.org/10.1080/10408690390826464. PMid:12587987.

Hou, Y., Zhao, P., Zhang, F., Yang, S., Rady, A., Wijewardane, N. K., Huang, J., & Li, M. (2022). Fourier-transform infrared spectroscopy and machine learning to predict amino acid content of nine commercial insects. *Food Science and Technology*, 42, e100821. http://dx.doi.org/10.1590/fst.100821.

Hu, Y., Xu, L., Huang, P., Luo, X., Wang, P., & Kang, Z. (2021). Reliable identification of oolong tea species: nondestructive testing classification based on fluorescence hyperspectral technology and machine learning. *Agriculture*, 11(11), 1106. http://dx.doi.org/10.3390/agriculture11111106.

Jia, Y., He, J., Fu, H., Shao, X., & Li, Z. (2018). Apple surface pesticide residue detection method based on hyperspectral imaging. In *International Conference on Intelligent Science and Big Data Engineering* (pp. 539-556). Cham: Springer. http://dx.doi.org/10.1007/978-3-030-02698-1_47.

Kwon, S. (2019). A CNN-assisted enhanced audio signal processing for speech emotion recognition. *Sensors (Basel)*, 20(1), 183. PMid:31905692.

Le, V. S., Lesueur, D., Herrmann, L., Hudek, L., Quyen, L. N., & Brau, L. (2021). Sustainable tea production through agroecological management practices in vietnam: a review. *Environmental Sustainability*, 4(4), 589-604. http://dx.doi.org/10.1007/s42398-021-00182-w.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. http://dx.doi.org/10.1038/nature14539. PMid:26017442.

Li, L., Huang, J., Wang, Y., Jin, S., Li, M., Sun, Y., Ning, J., Chen, Q., & Zhang, Z. (2020). Intelligent evaluation of storage period of green tea based on VNIR hyperspectral imaging combined with chemometric analysis. *Infrared Physics & Technology*, 110, 103450. http://dx.doi.org/10.1016/j.infrared.2020.103450.

Li, M., Lu, L., & Zhang, X. (2021). Qualitative determination of pesticide residues in purple cabbage based on near infrared spectroscopy. *Journal of Physics: Conference Series*, 1884(1), 012015. http://dx.doi.org/10.1088/1742-6596/1884/1/012015.

Liang, S., Liu, H., Gu, Y., Guo, X., Li, H., Li, L., Wu, Z., Liu, M., & Tao, L. (2021). Fast automated detection of COVID-19 from medical images using convolutional neural networks. *Communications Biology*, 4(1), 35. http://dx.doi.org/10.1038/s42003-020-01535-7. PMid:33398067.

Mahmudiono, T., Saleh, R. O., Widjaja, G., Chen, T., Yasin, G., Thangavelu, L., Altimari, U. S., Chupradit, S., Kadhim, M. M., & Marhoon, H. A. (2022). A review on material analysis of food safety based on fluorescence spectrum combined with artificial neural network technology. *Food Science and Technology*, 42, e118721. http://dx.doi.org/10.1590/fst.118721.

Martino, L. J., D'Angelo, C. A., Marinelli, C., & Cepeda, R. (2021). Identification and detection of pesticide in chard samples by laser-induced breakdown spectroscopy using chemometric methods. *Spectrochimica Acta. Part B, Atomic Spectroscopy*, 177, 106031. http://dx.doi.org/10.1016/j.sab.2020.106031.

Oyekunle, J., Adekunle, A. S., Adewole, A. M., Elugoke, S. E., Durodola, S. S., & Oyebode, B. A. (2021). Determination of organochlorine pesticide residues in some evaporated milk samples in nigeria using gas chromatography-mass spectrometry. *Chemistry Africa*, 4(2), 349-366. http://dx.doi.org/10.1007/s42250-020-00224-6.

Qiu, Z., Chen, J., Zhao, Y., Zhu, S., He, Y., & Zhang, C. (2018). Variety identification of single rice seed using hyperspectral imaging combined with convolutional neural network. *Applied Sciences (Basel, Switzerland)*, 8(2), 212. http://dx.doi.org/10.3390/app8020212.

Sun, C. L., Yuan, J. M., Koh, W. P., Lee, H. P., & Yu, M. C. (2007). Green tea and black tea consumption in relation to colorectal cancer risk: the Singapore Chinese Health Study. *Carcinogenesis*, 28(10), 2143-2148. http://dx.doi.org/10.1093/carcin/bgm171. PMid:17724377.

Sun, J., Cong, S., Mao, H., Wu, X., & Yang, N. (2018). Quantitative detection of mixed pesticide residue of lettuce leaves based on hyperspectral technique. *Journal of Food Process Engineering*, 41(2), e12654. http://dx.doi.org/10.1111/jfpe.12654.

Sun, R., Yang, W., Li, Y., & Sun, C. (2021). Multi-residue analytical methods for pesticides in teas: a review. *European Food Research and Technology*, 247(8), 1839-1858. http://dx.doi.org/10.1007/s00217-021-03765-3.

Wang, Y., Jin, G., Li, L., Liu, Y., Kalkhajeh, Y. K., Ning, J., & Zhang, Z. (2020). NIR hyperspectral imaging coupled with chemometrics for nondestructive assessment of phosphorus and potassium contents in tea leaves. *Infrared Physics & Technology*, 108, 103365. http://dx.doi.org/10.1016/j.infrared.2020.103365.

Wu, D., Meng, L., Yang, L., Wang, J., Fu, X., Du, X., Li, S., He, Y., & Huang, L. (2019). Feasibility of laser-induced breakdown spectroscopy and hyperspectral imaging for rapid detection of thiophanate-methyl residue on mulberry fruit. *International Journal of Molecular Sciences*, 20(8), 2017. http://dx.doi.org/10.3390/ijms20082017. PMid:31022906.

Wu, N., Zhang, C., Bai, X., Du, X., & He, Y. (2018). Discrimination of Chrysanthemum varieties using hyperspectral imaging combined with a deep convolutional neural network. *Molecules (Basel, Switzerland)*, 23(11), 2831. http://dx.doi.org/10.3390/molecules23112831. PMid:30384477.

Xing, L., Wang, Y., Luo, R., Li, X., & Zou, L. (2022). Determination of 31 pesticide residues in wolfberry by LC-MS/MS and dietary risk assessment of wolfberry consumption. *Food Science and Technology*, 42, e61921. http://dx.doi.org/10.1590/fst.61921.

Xin-Zhong, Z., Feng-Jian, L., Guang-Ming, L., Zheng-Yun, L., & Zong-Mao, C. (2011). Determination of diafenthiuron and its metabolites in tea and soil by ultra performance liquid chromatography-electrospray ionization tandem mass spectrometry. *Chinese Journal of Analytical Chemistry*, 39(9), 1329-1335.

Xu, Z., Zhu, L., Yang, Y., Zhang, Y., Lu, M., Tao, L., & Xu, W. (2021). Bifenthrin induces DNA damage and autophagy in Spodoptera frugiperda (Sf9) insect cells. *In Vitro Cellular & Developmental Biology. Animal*, 57(3), 264-271. http://dx.doi.org/10.1007/s11626-021-00554-w. PMid:33689124.

Yuan, R., Liu, G., He, J., Wan, G., Fan, N., Li, Y., & Sun, Y. (2021). Classification of Lingwu long jujube internal bruise over time based on visible near-infrared hyperspectral imaging combined with partial least squares-discriminant analysis. *Computers and Electronics in Agriculture*, 182, 106043. http://dx.doi.org/10.1016/j.compag.2021.106043.

Zagajewski, B., Kluczek, M., Raczko, E., Njegovec, A., Dabija, A., & Kycko, M. (2021). Comparison of random forest, support vector machines, and neural networks for post-disaster forest species mapping of the krkonoše/karkonosze transboundary biosphere reserve. *Remote Sensing*, 13(13), 2581. http://dx.doi.org/10.3390/rs13132581.

Zhang, D., Liang, P., Chen, W., Tang, Z., Li, C., Xiao, K., Jin, S., Ni, D., & Yu, Z. (2021a). Rapid field trace detection of pesticide residue in food based on surface-enhanced Raman spectroscopy. *Mikrochimica Acta*, 188(11), 370. http://dx.doi.org/10.1007/s00604-021-05025-3. PMid:34622367.

Zhang, X., Lin, T., Xu, J., Luo, X., & Ying, Y. (2019). DeepSpectra: an end-to-end deep learning approach for quantitative spectral analysis. *Analytica Chimica Acta*, 1058, 48-57. http://dx.doi.org/10.1016/j.aca.2019.01.002. PMid:30851853.

Zhang, X., Sun, J., Li, P., Zeng, F., & Wang, H. (2021b). Hyperspectral detection of salted sea cucumber adulteration using different spectral preprocessing techniques and SVM method. *Lebensmittel-Wissenschaft + Technologie*, 152, 112295. http://dx.doi.org/10.1016/j.lwt.2021.112295.

Zhu, J., Sharma, A. S., Xu, J., Xu, Y., Jiao, T., Ouyang, Q., Li, H., & Chen, Q. (2021). Rapid on-site identification of pesticide residues in tea by one-dimensional convolutional neural network coupled with surface-enhanced raman scattering. *Spectrochimica Acta. Part A: Molecular Spectroscopy*, 246, 118994. http://dx.doi.org/10.1016/j.saa.2020.118994. PMid:33038862.