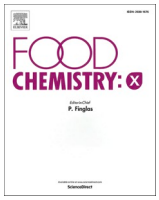
Food Chemistry: X 19 (2023) 100851

Contents lists available at ScienceDirect

Food Chemistry: X

journal homepage: www.sciencedirect.com/journal/food-chemistry-x



An early asymptomatic diagnosis method for cork spot disorder in 'Akizuki' pear *(Pyrus pyrifolia* Nakai) using micro near infrared spectroscopy

Li Liu [a,1], Hanhan Zhang [a,1], Lin Wu [a], Shangfeng Gu [a,b], Jing Xu [a], Bing Jia [a], Zhenfeng Ye [a], Wei Heng [a,*], Xiu Jin [b,*]

[a] School of Horticulture, Anhui Agricultural University, 130 Changjiang West Road, Hefei 230036, China
[b] School of Information and Computer Science, Anhui Agriculture University, 130 Changjiang West Road, Hefei 230036, China

ARTICLE INFO

Keywords:
Near infrared spectrum (NIR)
Cork spot disorder
'Akizuki' pear
Early identification
Mineral elements

ABSTRACT

The early symptoms of cork spot disorder in 'Akizuki' pear (*Pyrus pyrifolia* Nakai) are challenging to distinguish from those in healthy fruits, hindering early identification in production. In this study, samples of cork-browned 'Akizuki' pears, asymptomatic fruits and healthy fruits were examined to determine the content of relevant mineral elements. A micro near-infrared spectrometer collected spectral information, and various pretreatment methods were applied to the near-infrared spectral data. Support vector machine (SVM) modelling using the original data achieved the highest overall recognition accuracy of 84.65% and an F1 value of 84.06%. For identifying fruits without cork spot disease, Autokeras modelled data processed with the SG method, achieving the best accuracy of 90%. These findings establish a reliable basis for the early identification and diagnosis of cork spot disorder in 'Akizuki' pear, enhancing pear production management.

1. Introduction

'Akizuki' pear (*Pyrus pyrifolia* Nakai) is a new mid-late maturing sand pear variety bred in Japan. Cork spot disorder is a fruit physiological disease caused by the weak adaptability of new cultivars to climate stress (Cui, Jiao, Wang, & Ma, 2020). The cork spot disorder reduces the commodity value of the fruit, reduces the quality of the fruit, and causes significant losses to the production of pears (Cui et al., 2020). According to the data, the cork spot disorder mainly occurs in the expansion period of the fruit. Vascular bundle dysfunction destroys the balance of water and nutrient transport in the fruit, resulting in changes in the pore structure of the pulp and an imbalance of mineral elements (Zhang & Cui, 2023). The related mineral elements are mainly calcium and boron (Cui et al., 2020) and some elements that can lead to calcium imbalance, such as K and Mg (Schönherr & Bukovac, 1973). The fruit of 'Akizuki' pear with cork disease mainly showed two kinds of symptoms. One was that the peel was rounded and depressed, gradually extending downwards to cover the pulp, and spots were formed in the pulp, which was similar to the bitter core symptoms of apple. The other was that there was no disease on the surface of the peel, and brown and lignified round spots appeared in the pulp near the peel (Zhang & Cui, 2023). The judgment of the degree of illness is mainly based on subjective sensory experience, which makes it difficult to achieve qualitative evaluation. In the early stage, the asymptomatic fruit of 'Akizuki' pear cork disease can be effectively identified, and corresponding measures can be taken in time to prevent the development of the disease. For mature pears, identifying asymptomatic diseased fruits is helpful for fruit grading. However, due to the concealment of this defect, there is currently no effective detection method. Therefore, it is urgent to achieve accurate, rapid and non-destructive methods for asymptomatic fruits of 'Akizuki' pear with cork disease to improve the production level of the whole pear industry. Fig. 1.

The color of the biological material give us the information about the health and quality of the agricultural product. Also, it is a criteria for marketing value of the agricultural product for consumers. Classification can be completed by computer vision technology (Bhargava & Bansal, 2018), and plant diseases can be automatically classified and detected according to shape, texture and color (Zhang et al., 2021). In addition to disease identification, non-destructive methods can also be used to evaluate the quality and safety of agricultural products and foodstuffs (El-Mesery, Mao, & Abomohra, 2019) and to determine the oxidized areas and color changes in damaged apples using image analysis

* Corresponding authors.
E-mail addresses: hengwei@ahau.edu.cn (W. Heng), jinxiu123@ahau.edu.cn (X. Jin).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.fochx.2023.100851
Received 6 June 2023; Received in revised form 20 August 2023; Accepted 28 August 2023
Available online 5 September 2023

techniques.

In the non-destructive testing of internal defects in fruits, many studies have been carried out using various technologies, including X-ray imaging (Qian, Wenqian, Qingyan, Jingzhu, & Jiangbo, 2022), thermal imaging and magnetic resonance imaging (MRI), visible-near infrared (Vis/NIR) spectroscopy is a more commonly used technology in the non-destructive testing of agricultural products. It has many advantages, such as good detection efficiency, accuracy, multicomponent simultaneous analysis and low cost (Xia, Xu, Li, Zhang, & Fan, 2019). Based on the differences in the frequency combinations and octave absorption intensities of C—H, N—H, and O—H groups in the spectra, combined with chemometric methods, internal physiological diseases of fruits can be identified (Anderson & Walsh, 2022). Near-infrared spectroscopy was used for early detection of strawberry anthracnose crown rot (Lu et al., 2017) and bitter pit disease in asymptomatic 'Fuji' apples (Mogollón, Contreras, de Freitas, & Zoffoli, 2021). In the research of our team, Ba (Ba et al., 2022) used near-infrared spectroscopy to model and analyse the asymptomatic diagnosis of wheat kernel scab and accumulated experience for other team members to use near-infrared spectroscopy for modelling and analysis. Therefore, in this paper, an identification model of 'Akizuki' pear cork disease was established by near infrared spectroscopy, and the feasibility of near infrared spectroscopy for the diagnosis of early 'asymptomatic' pear cork disease was discussed.

Fast detection has high requirements for the execution efficiency of the model. To improve the performance of the model and achieve the purpose of practical application, it is necessary to remove redundancy, noise and collinear processing of the original spectral data. In this experiment, various preprocessing methods are used to correct the scattering, baseline variation, peak shift, noise and missing values (Mishra, Biancolillo, Roger, Marini, & Rutledge, 2020). Various preprocessing methods have different degrees of impact on the data in different ways. Preprocessing methods always have the risk of losing relevant chemical information or changes related to the attributes of interest (Oliveri, Malegori, Simonetti, & Casale, 2019). However, the combination of different pretreatment methods can weaken or eliminate the influence of a single pretreatment method on spectral information. Therefore, this paper selects six preprocessing methods and two common preprocessing combinations to process the near-infrared spectrum to find a preprocessing method that can not only attenuate the noise and scattering in the data but also retain the spectral information to the greatest extent.

For NIRS classification, some traditional classifiers are used, such as support vector machines (SVM) (Lu, Zheng, Hu, Lou, & Kong, 2011) and ensemble algorithms. In this paper, the random forest algorithm (RF) (Bin et al., 2016) and the polar gradient descent algorithm (Xia, Qin, Jumei, Yingying, & Bingyu, 2022) are selected as representatives of the ensemble algorithm. Traditional machine learning methods rely on human experience to design and specify parameters, which greatly affect the performance of the model. This paper applies the automatic reference model. The model is an automatic machine learning (AutoML) library Autokeras, which can automatically complete the model selection and hyperparameter adjustment process (Alaiad et al., 2023).

In general, the purpose of this study was to determine a classification model to detect asymptomatic pears with 'Akizuki' pear cork disease. The specific purpose is as follows: (1) Using near-infrared spectroscopy combined with different classification models to explore the detection ability of 'Akizuki' pear cork disease.(2) The random forest (RF), eXtreme gradient boosting (XGBoost), adaptive boost (AdaBoost), gradient boosting decision tree (GBDT), Autokeras and support vector machine (SVM) modelling methods were established, and the classification model of 'Akizuki' pear cork disease was established. (3) The classification performance of each model was compared from the aspects of tolerance, stability, accuracy and complexity, and the best model for distinguishing diseased pears from healthy pears was determined.

2. Materials and methods

2.1. Samples

In early September 2022, the experiment was carried out at the pear demonstration base of Youjia Agricultural Development Co., Ltd., Yeji District, Lu'an Anhui Province. After the reliability of the sample was determined by several experienced fruit growers, 85 diseased fruits and 72 healthy fruits (control) of 'Akizuki' pear were picked as test samples. The surface of the pear fruit was cleaned, wiped clean and then numbered for use.

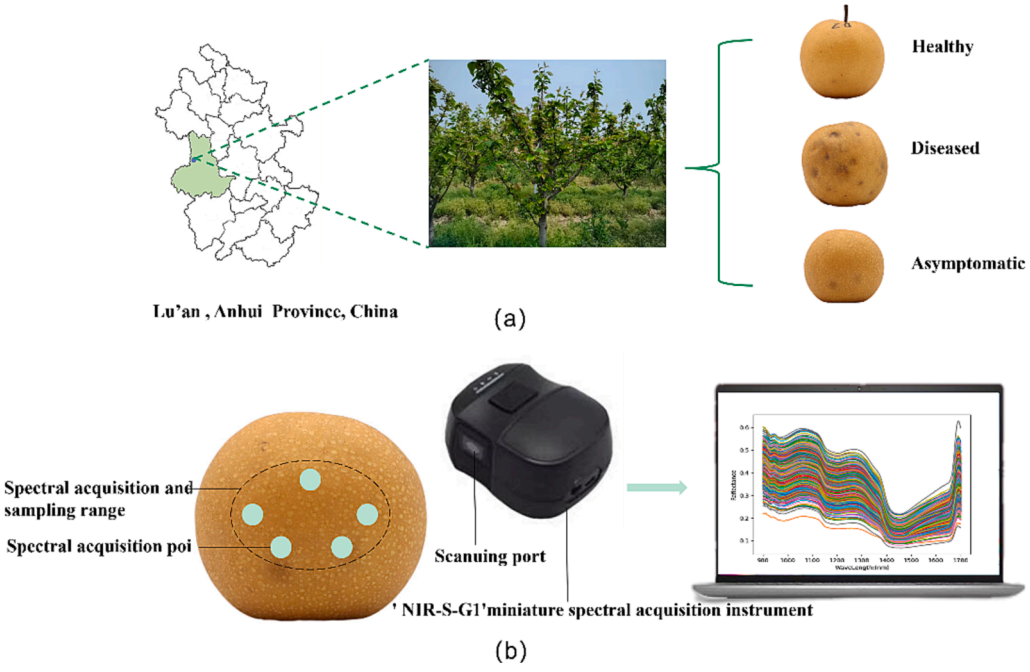


Fig. 1. Materials and Methods (a) The collection location of 'Akizuki' pear and appearance of three types of pear fruit; (b) The process of acquiring the spectra.

2.2. Near-infrared spectrum instrument

In this experiment, the instrument used to collect spectral data is a handheld portable miniature near-infrared spectrometer 'NIR-S-G1' produced by Shenzhen Green Union Company. The spectral wavelength detection range is 900 nm – 1700 nm, the spectral acquisition points are 228 bands, the spectral resolution is 3.89 nm, and the signal-to-noise ratio (SNR) is 5000:1. Before use, the app 'Instagram' on the mobile phone was connected to the instrument through Bluetooth, and then the spectral data were collected.

Before each measurement, the near-infrared spectrometer needs to be calibrated with a black and white board. The instrument is attached to the calibration whiteboard, and the light emitted is reflected back on the whiteboard. The reflected light is captured and recorded as the brightness value (W) of the whiteboard. We turn off the instrument's emission light source, which is recorded on the blackboard brightness value (B). After calibration, the instrument was used to collect the spectral data of the pear surface. The light source window of the instrument is close to the brightness value (R) of the reflected light obtained by the 'Akizuki' pear sample. The spectral reflectance of the sample is calculated by formula (1).

$$Refletance = \frac{I - B}{W - B} \times 100\% \quad (1)$$

2.3. Spectral data acquisition

In the process of spectral acquisition, to make the measurement results as accurate as possible, we selected multiple spectral sampling points on the surface of the pear for acquisition and averaging, as shown in Fig. 1. An ellipse with a short axis of 6 cm and a long axis of 7 cm was delineated near the equator on the surface of the pear fruit by pencil, and the sample points were marked in the ellipse by a five-point sampling method. For healthy fruits, we delimited an elliptical area every 120 degrees, and for diseased fruits, we delimited an ellipse on the surface of 1–3 lesions and the asymptomatic surface. The scanning window of the front end of the micro handheld spectrometer is close to the designated range. The number of scans in each sample area was 5. Each data file is named according to the sample number. After the spectral data acquisition is completed, all files are exported. The data in each document include wavelength, intensity spectrum, absorbance spectrum and reflectivity. The average value of the reflectivity spectral data scanned five times in each region is used as the original modelling spectral data.

2.4. Detection of mineral elements in pear fruit

After collecting spectral data from all pear samples, healthy fruits, diseased parts of diseased fruits and asymptomatic parts of diseased fruits were peeled and removed. The pulp was cut off manually and placed in a Kraft paper bag. The pulp was placed in kraft paper bags, dried at 70 degreesC until constant weight, mechanically pulverized, placed in self-sealing bags, labelled with a number, and stored for use. Three replicates were set up for each part.

First, 0.1 g of dry sample powder was weighed into an ablution jar, and 2 ml of concentrated sulfuric acid and 8 ml of distilled water were added. Microwave digestion was carried out using a microwave digestion apparatus (Milestone Ethos T) for 0.5 to 1 h.Then, inductively coupled plasma emission spectrometry (ICPES) was used to determine the contents of mineral elements in the fruits of autumn pear (Zhang, Nie, & Le, 2017). The detection basis was GB 5009.268-2016, and the main instrument used was an inductively coupled plasma emission spectrometer/B1126.

2.5. Spectral processing method

In the process of spectral data acquisition, there are inevitably some interference factors unrelated to spectral information acquisition, such as background noise of the experimental platform. This leads to some irrelevant information in the original spectral data. The research shows that data preprocessing helps to weaken the invalid information in the spectral curve, improve the utilization rate of the effective information of the spectrum, increase the signal-to-noise ratio of the near-infrared spectrum, and improve the accuracy and stability of the regression model established later. For this batch of data, this experiment uses the first derivative (FD), second derivative (SD), multiple scattering correction (MSC), standard normal variable (SNV), Savitzky Golay convolution smoothing (SG) and logarithmic transformation (LG) and the combination of two single preprocessing algorithms, SG + SNV and SG + MSC preprocessing methods.

The derivatives (FD and SD) can eliminate spectral offset and baseline effects (Yaqian, Xinyuan, Zhisheng, & Qiao. , 2016). Multivariate scattering correction (MSC) is based on the spectral array, which can eliminate the influence of the surface scattering of pears to a certain extent. The effect of the standard normal variate (SNV) is similar to that of the MSC, which helps to reduce the influence of factors such as pear fruit size and surface scattering on spectral diffuse reflection. The difference between the two is that MSC takes the average spectral reflectance of a single band of all samples and then performs subsequent processing, while SNV takes the average of the full-band spectral reflectance of a single sample and then performs subsequent processing. SG mainly eliminates the noise generated by the outside world or equipment. The logarithmic transformation (LG) helps to reduce the absolute value of the spectral data, which is beneficial to the later calculation (Yujie et al., 2022). Through the comprehensive evaluation of the performance of each model with different evaluation indexes, the best pretreatment method is selected for analysis. Then, we compare the modelling results under different preprocessing methods.

2.6. Modelling method

In this paper, the algorithms selected for modelling are random forest (RF), extreme gradient boosting (XGBoost), adaptive boosting (AdaBoost), support vector machine (SVM), gradient boosting tree (GBDT) and neural network structure search (Autokeras). SVM is a widely used supervised classification learning algorithm (Mammone, Turchi, & Cristianini, 2009). The classification data are predicted by finding the maximum boundary in the feature space. SVM predicts nonlinear data based on empirical risk minimization and prevents overfitting by setting penalty parameter C and kernel function parameter gamma (Li, Sun, & Li, 2020).

Autokeras is an efficient neural search with network morphism that uses Bayesian optimization to guide the search space through each optimal operation. Autokeras is an open source software library for automated machine learning that automatically searches for the architecture and hyperparameters of deep learning models. When using Autokeras to train models, there is almost no need to intervene. In this experiment, Autokeras is used for training data, including the following three aspects: search space, search strategy and evaluation index. The dataset of this experiment is smaller, and the network architecture is not complicated.

Ensemble learning is a machine learning technique that combines basic models according to different ideas to achieve better results. More than one single machine learning algorithm is trained and combined to produce a final output. (Alazzam, Alsmadi, & Akour, 2017) The ensemble model mainly includes the ensemble bagging model, the serial ensemble boosting model and the structured ensemble stacking model.

Random forest is an extended variant of bagging. Based on the bagging ensemble constructed by decision tree-based learners, RF further introduces random attribute selection into the training process of the decision tree.

AdaBoost belongs to the Boosting algorithm in integrated learning (Freund & Schapire, 1997). AdaBoost is a special machine learning

method for training a series of weak classifiers. In the AdaBoost training process, the weights of training samples are updated adaptively after each enhancement iteration. The weight of training samples misclassified by the current component classifier is increased, and the weight of training samples correctly classified is reduced. Finally, the weak classifiers are linearly combined to form a strong classifier.

To analyse the performance of different models, evaluation indicators are introduced, including accuracy, precision (P), recall rate (R) and F1 value. The most commonly used evaluation index in classification modelling is accuracy, but for models with uneven data distribution, the F1 value is better than accuracy and can be used to evaluate the goodness of fit of the model. Both accuracy and F1 value are considered when evaluating the model.

This section briefly introduces the methods used. According to the KS (Watson & Galliher, 2001) algorithm, 714 samples were divided into a training set and a test set at a ratio of 7:3. The test set is used to evaluate the accuracy of the model, that is, the generalization ability. The third chapter mainly introduces the results and discussions of mineral element detection, spectral characteristics, preprocessing and modelling.

All algorithms are implemented in PyCharm 2022.1.1 under Windows 10, and the programming language is Python 3.9.

3. Results and discussion

3.1. Comparison of the main mineral element contents between healthy and diseased 'Akizuki' pear fruits

The detection results of the main mineral elements in healthy pear fruit, diseased fruit and asymptomatic fruit are shown in Fig. 2. Through the analysis of the results, it can be found that there was a significant difference in the content of Ca in healthy fruits, diseased fruits and asymptomatic fruits. The content of K in asymptomatic fruit was significantly different from that in diseased fruit and healthy fruit, but there was no significant difference between diseased fruit and healthy fruit. The content of Mg in diseased fruit was significantly different from that in asymptomatic fruit and healthy fruit, but there was no significant difference between asymptomatic fruit and healthy fruit. The content of trace element B in diseased fruit was significantly different from that in asymptomatic fruit and healthy fruit, but there was no significant difference between asymptomatic fruit and healthy fruit. The K/Ca, Mg/Ca and K + Mg/Ca ratios of healthy fruit, diseased fruit and asymptomatic fruit were significantly different.

During the course of the disease, the contents of Ca and B decreased gradually. After the formation of the lesion, the content of Ca was significantly higher than that of the diseased site. Ca/B is often used to determine whether the fruit is deficient in B in production. The Ca/B difference between healthy fruit, diseased fruit and asymptomatic fruit is extremely significant.

The symptoms of pear cork disease are similar to those of apple bitter core, which has been confirmed to be related to calcium deficiency. The contents of K and Mg and their ratios to Ca are important indicators related to the occurrence of cork spots in other studies (Miqueloto, Amarante, Steffens, dos Santos, & Mitcham, 2014). Potassium (K) and magnesium (Mg) can antagonize the absorption or function of calcium on the cell membrane or compete for the binding sites of calcium on the cell membrane (Ho & White, 2005). The balance between calcium and

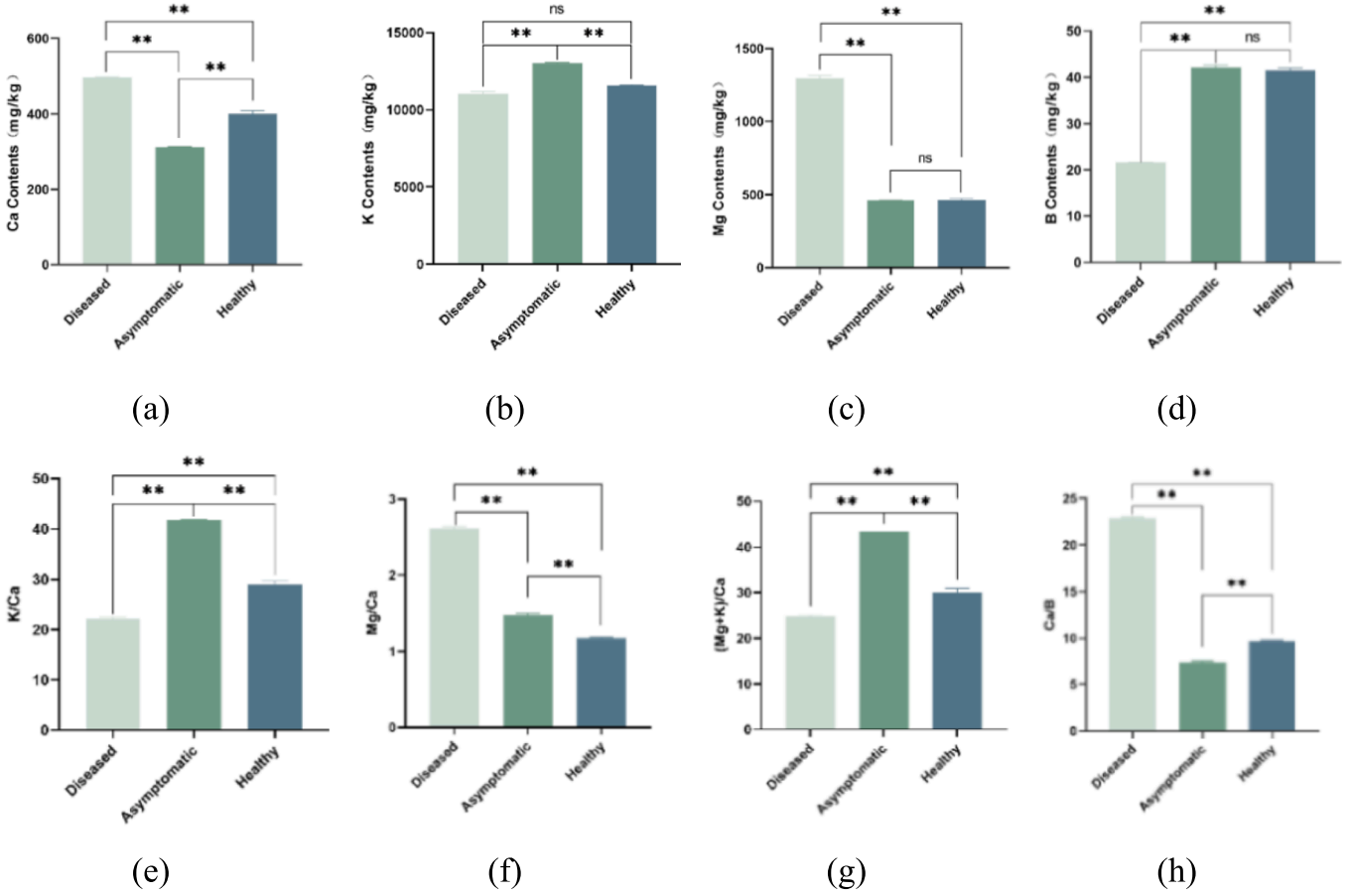


Fig. 2. Comparison of mineral element content in different degrees of disease. (a) Comparison of Ca content; (b) Comparison of K content. (c) Comparison of Mg content; (d) Comparison of B content. (e) Comparison of K/Ca content; (f) Comparison of Mg/Ca content. (g) Comparison of (Mg + K)/Ca content; (h) Comparison of Ca/B content.

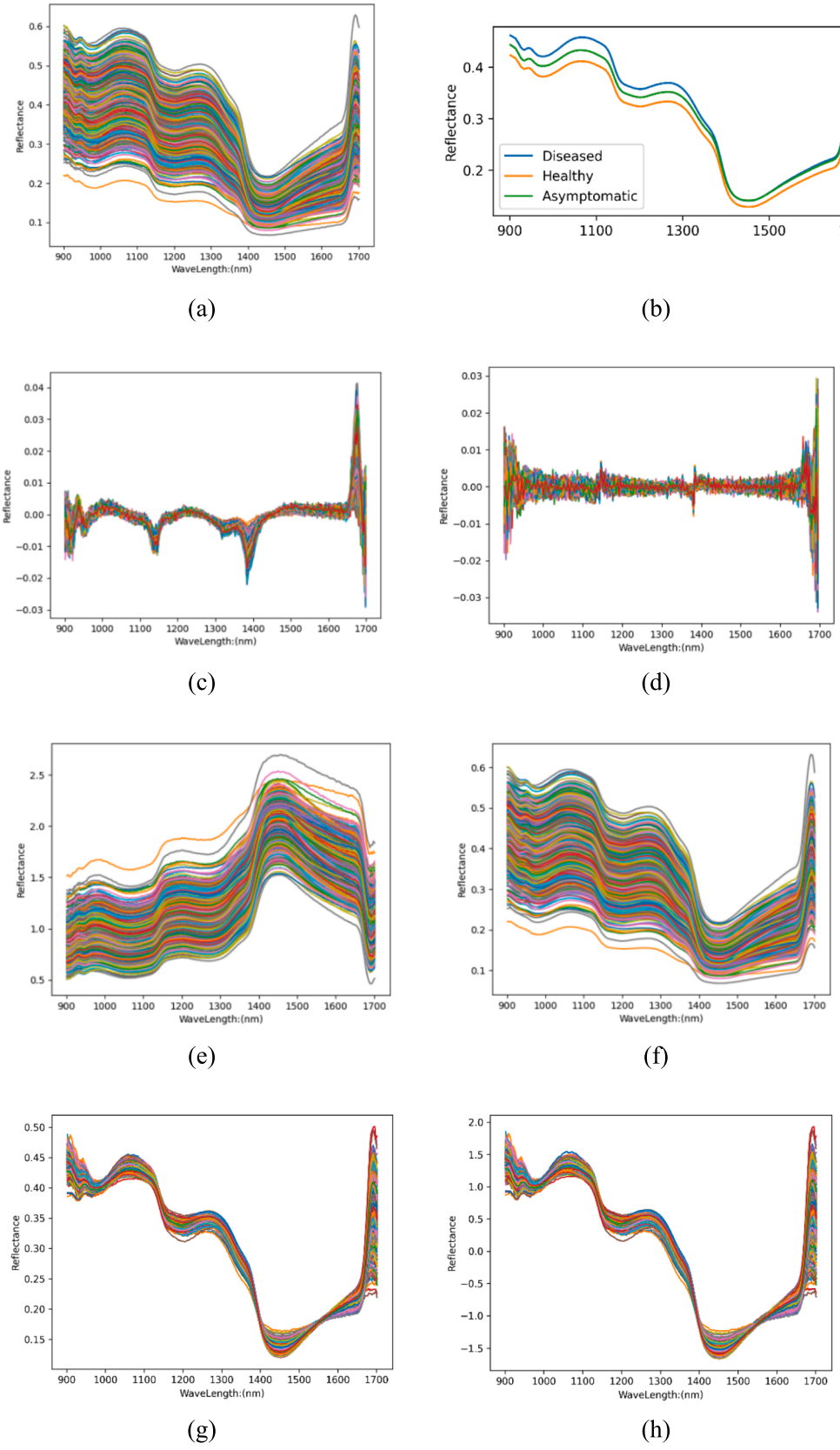


Fig. 3. Spectral reflectance of pear fruit (a) The original spectral reflectance of pear fruit; (b) The average spectral reflectance of pears with different degrees of disease. Preprocessed spectral curve:(c) spectrum with the FD method; (d) spectrum with the SD method; (e) spectrum with the LG method; (f) spectrum with the SG method; (g) spectrum with the MSC method; (h) spectrum with the SNV method; (i) spectrum with the SG + MSC method; (j) spectrum with the SG + SNV method.

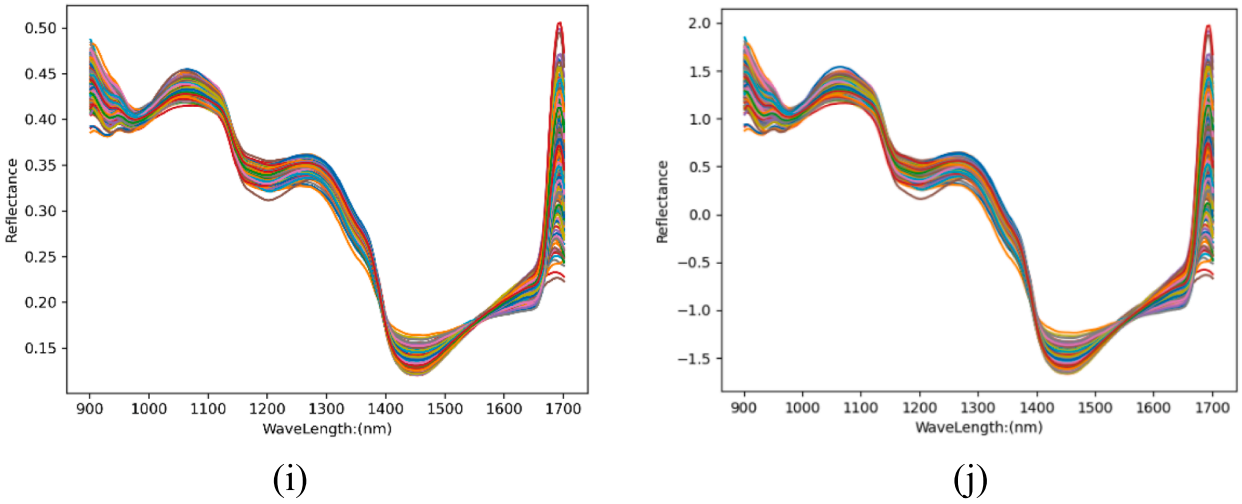


Fig. 3. (*continued*).

other mineral elements may also be the cause of the disease. Ca mainly exists in the form of ions in plants, and Ca^{2+} in plant cells is mainly stored in cell walls and vacuoles (Woodbridge, 1971). Fruits with Ca^{2+} deficiency symptoms even showed higher total Ca content than normal fruits due to increased Ca^{2+} in vacuoles being stored or bound to the cell wall (de Freitas, do Amarante, Labavitch, & Mitcham, 2010). Calcium is the component of calcium pectinate in the intercellular layer of the cell wall. K is an essential macroelement in plants that affects the function of enzymes in plants and mainly exists in the form of K^{+} in plant cells (Ciavarella, Batten, & Blakeney, 1998).

The differences in the content of mineral elements can be shown by spectral data. De Aldana reported that calcium pectate binds to plant cells and is sensitive in the NIR region (De Aldana, Criado, Ciudad, & Corona, 1995). Boron in crops mainly exists in the form of boric acid, which can be condensed with *cis*-diol to synthesize ester compounds. In plants, many compounds have adjacent *cis*-diol configurations, which can be condensed with boric acid to affect or mediate plant metabolism. As an inorganic acid, boric acid is difficult to directly reflect on the spectral curve. The content of B can be indirectly reflected by *cis*-diol. The wavelength at which K may have the strongest correlation with carbohydrates such as sucrose, starch and cellulose in plants can be detected by NIRS through carbohydrates and organic acids and can also be detected in cation-carbohydrate complexes (Cadet & Offmann, 2006). Mg was significantly correlated with chlorophyll bands, and chlorophyll bands were found by Clark in NIRS (Clark, Mayland, & Lamb, 1987).

3.2. Spectral characteristics and analysis

In this experiment, the reflection spectrum was selected as the modelling spectral data. The original near-infrared (900–1700 nm) spectra of diseased, asymptomatic and diseased fruits are shown in Fig. 3. Fig. 3(a) is the original reflection spectrum of pear fruit, and Fig. 3 (b) is the average reflection spectrum. The variation trend of the near-infrared spectra of healthy fruits, diseased fruits and asymptomatic fruits was consistent. There were significant differences in the range of 900–1300 nm. The spectral curves between diseased fruits and asymptomatic fruits in the range of 1400–1700 nm almost overlap, but the average spectral reflectance of the two is significantly different from that of healthy fruits. These average spectra reflect the effective chemical information of pears with different degrees of disease. This difference indicates that the chemical composition content of each disease degree is different. The spectral data meet the requirements of establishing an infrared species recognition model.

Because the stretching vibrations of the overtones of O—H, C—H and N—H are related to the concentration of some internal compounds with these bands, stronger absorption peaks were found in the spectral range of 900–1000 nm (Zahra, Bahareh, Mansour, & Yousef, 2018). It can be seen from the average spectral curve that there are differences in some common peaks, such as 990 nm, 1450 nm, and 1490 nm. Study on the attribution of different peaks. The results showed that the second harmonic absorption of O — H was observed at 990 nm, the first harmonic absorption of O — H was observed at 1450 nm, and the first harmonic absorption of O — H was observed at 1490 nm. The organic groups were starch, water and cellulose. The difference here may be related to the K content (Ciavarella et al., 1998).

The internal components of 'Akizuki' pear fruit may change after cork browning. This is mainly reflected in the starch, water and cellulose contents. The results showed that the pear belonged to the sugar accumulation type in the middle. In the early and middle stages of fruit development, starch mainly accumulated. In the late stage of fruit development, starch was degraded into soluble sugar. Some studies have measured that the starch content of the diseased fruit of 'Akizuki' pear is higher than that of the normal fruit, and there is a significant difference (Cui et al., 2020). This further illustrates that the reflectance of diseased fruit is higher than that of healthy fruit.

There are two different peaks and valleys in the spectral bands at 1300 and 1500 nm. Many studies have shown that the bands of C—C aromatic skeleton vibration at 1500 nm and the bands at 1300 nm represent C—C and C—O skeleton vibrations, which are highly correlated with lignin content (Barton & Himmelsbach, 1993). The peak difference at approximately 1400 nm may be due to the difference in lignin content between diseased and healthy fruits (Cui et al., 2020).

The wavelength is near 1490 nm, and the organic group reflected is cellulose. The cellulose content was significantly different between the diseased and normal fruits of 'Akizuki' pear, and the content of diseased fruits was higher than that of healthy fruits (Jamshidi, Mohajerani, & Jamshidi, 2016). There was no significant difference in content between different parts of the diseased fruit. In the spectral curve, the diseased fruit and the asymptomatic fruit overlap, and the curve difference with the healthy fruit is obvious. The differences in moisture, starch, cellulose and mineral elements between the lesion site and the healthy site and the asymptomatic site of the pear fruit were reflected in the near-infrared spectrum. This further proves the feasibility of using near-

infrared spectroscopy to classify 'Akizuki' pear fruits.

The original spectral data include the following three types: healthy fruit, asymptomatic fruit and diseased fruit, including 499 samples for the training set and 215 samples for the test set. In this experiment, 714 spectral data points collected from 167 freshly picked 'Akizuki' pear fruits were used as raw data. A total of 246 data points were labelled healthy fruits, 213 data points were labelled diseased fruits, and 255 data points were labelled asymptomatic fruits. The collected data are divided into a training set and a test set at a ratio of 7:3, which is used for model training and correction.

The micro near-infrared spectral data contain not only sample information but also noise and background information. To deeply analyse the spectral characteristics, the near-infrared spectral data are preprocessed. Spectral preprocessing eliminates the spectral changes unrelated to the chemical composition, which is considered to be a basic procedure before multivariate calibration. Selecting a separate

Table 1
Classification results based on the RF, XGBoost, AdaBoost, GBDT, Autokeras and SVM models.

Sodels	Pre processing method	Training set				Testing set			
		P	R	F1	Accuracy	P	R	F1	Accuracy
RF	R	95.60%	95.07%	95.28%	95.39%	64.21%	63.34%	63.19%	63.26%
	FD	99.45%	99.31%	99.38%	99.40%	77.42%	75.14%	75.27%	74.88%
	SD	99.28%	99.11%	99.19%	99.20%	74.91%	72.23%	70.00%	70.23%
	SG	95.07%	94.47%	94.70%	94.79%	64.24%	63.27%	63.10%	63.20%
	LG	92.66%	91.73%	92.05%	92.18%	63.44%	62.43%	62.24%	62.33%
	MSC	97.45%	97.26%	97.34%	97.39%	66.20%	73.37%	72.58%	72.56%
	SNV	94.82%	94.71%	94.74%	94.79%	72.998%	72.04%	71.15%	71.16%
	SG + MSC	92.62%	92.54%	92.54%	92.59%	72.73%	72.11%	71.17%	71.16%
	SG + SNV	96.79%	96.71%	96.74%	96.79%	73.48%	72.45%	71.71%	71.63%
XGBoost	R	96.565%	95.75%	96.03%	96.19%	60.88%	60.22%	60.08%	60.47%
	FD	97.30%	96.78%	96.99%	96.99%	67.18%	65.03%	65.66%	65.12%
	SD	96.91%	96.88%	96.87%	96.79%	68.03%	67.59%	67.48%	66.98%
	SG	97.77%	97.35%	97.52%	97.60%	58.72%	56.60%	56.37%	57.21%
	LG	93.50%	92.11%	92.57%	92.59%	60.30%	59.51%	59.25%	60.00%
	MSC	96.80%	96.92%	96.85%	96.79%	74.29%	74.15%	73.54%	73.49%
	SNV	95.64%	95.71%	95.64%	95.59%	71.74%	71.36%	70.81%	70.70%
	SG + MSC	97.28%	97.31%	97.25%	97.19%	75.71%	75.18%	74.49%	74.42%
	SG + SNV	97.67%	97.66%	97.63%	97.60%	74.04%	73.46%	73.20%	73.02%
AdaBoost	R	78.96%	78.91%	78.92%	78.96%	59.86%	59.76%	59.69%	60.47%
	FD	93.48%	93.43%	93.43%	93.19%	80.73%	80.51%	80.59%	79.07%
	SD	93.38%	93.24%	93.23%	92.99%	71.41%	70.74%	70.80%	68.37%
	SG	92.36%	92.47%	92.39%	92.38%	57.88%	57.48%	57.44%	58.17%
	LG	78.96%	78.91%	78.92%	78.96%	59.86%	59.76%	59.69%	60.47%
	MSC	90.83%	90.92%	90.87%	90.78%	67.45%	65.73%	66.14%	65.58%
	SNV	90.78%	90.78%	90.71%	90.58%	69.21%	67.65%	68.12%	67.44%
	SG + MSC	87.79%	88.04%	87.76%	87.78%	68.98%	68.51%	68.35%	67.91%
	SG + SNV	89.21%	89.46%	89.22%	89.19%	70.11%	67.02%	67.26%	67.44%
GBDT	R	97.35%	97.13%	97.22%	97.19%	60.566%	60.53%	60.40%	60.47%
	FD	96.27%	96.14%	96.19%	96.19%	68.25%	67.22%	67.87%	67.44%
	SD	97.38%	97.49%	97.42%	97.39%	63.57%	64.10%	63.80%	63.26%
	SG	96.20%	95.58%	95.99%	95.99%	53.54%	53.62%	53.52%	53.95%
	LG	94.50%	94.04%	94.19%	94.19%	58.37%	58.22%	58.25%	58.60%
	MSC	96.07%	96.02%	96.02%	95.99%	66.93%	67.28%	67.09%	66.98%
	SNV	92.25%	92.25%	92.24%	92.18%	68.42%	68.89%	68.40%	68.37%
	SG + MSC	96.80%	96.84%	96.81%	96.79%	67.68%	68.30%	67.42%	67.44%
	SG + SNV	93.45%	93.37%	93.39%	93.39%	68.44%	69.16%	68.32%	68.37%
Autokeras	R	83.32%	80.78%	80.94%	80.96%	74.75%	71.78%	71.51%	72.09%
	FD	100%	100%	100%	100%	67.00%	68.33%	67.40%	66.98%
	SD	100%	100%	100%	100%	59.88%	63.79%	60.13%	60.00%
	SG	82.38%	80.83%	81.09%	80.96%	77.38%	76.55%	76.31%	77.21%
	LG	92.15%	91.93%	92.00%	91.98%	67.78%	68.07%	67.89%	68.37%
	MSC	96.31%	96.14%	96.21%	96.19%	79.80%	79.36%	79.55%	79.07%
	SNV	97.15%	96.99%	97.03%	978.51%	78.51%	78.30%	78.39%	77.67%
	SG + MSC	97.21%	97.22%	97.22%	97.19%	77.47%	78.07%	77.71%	77.21%
	SG + SNV	91.38%	91.23%	91.25%	91.18%	77.71%	78.07%	77.88%	77.21%
SVM	R	91.13%	89.95%	90.28%	90.18%	85.16%	84.71%	84.67%	84.65%
	FD	84.83%	82.70%	83.29%	83.37%	75.46%	73.80%	73.77%	73.49%
	SD	88.60%	88.27%	88.17%	88.18%	68.35%	69.14%	68.29%	68.84%
	SG	88.38%	87.06%	87.47%	87.37%	83.55%	82.63%	82.51%	82.33%
	LG	93.22%	92.41%	92.69%	92.59%	80.59%	80.43%	80.40%	80.47%
	MSC	92.60%	91.91%	92.18%	91.98%	84.14%	83.60%	83.53%	83.26%
	SNV	90.38%	89.79%	90.01%	89.78%	84.49%	83.69%	83.51%	83.26%
	SG + MSC	88.25%	87.61%	87.83%	87.58%	83.40%	82.92%	82.53%	82.33%
	SG + SNV	88.10%	87.41%	87.63%	87.37%	83.40%	82.92%	82.53%	82.33%

preprocessing method and a combination can reduce the interference of irrelevant information (Shi & Yu, 2017). FD, SD, SG, MSC, SNV, LG, SG + MSC, and SG + SNV were performed on the spectral data. To prevent the difference in sample data observed after correction by different pretreatment methods, the results of the conversion spectra were compared. The results of the spectral data processed by the derivative method are shown in Fig. 3(c and d). Compared with the original spectrum, the spectral difference after derivative transformation is more obvious. This is because the derivative can randomly reduce the influence of noise, highlight the fine band shape and retain the relative band intensity information. It can be seen from the graph that the baseline shift phenomenon has been improved in the spectral data after derivative processing. The first derivative is used to eliminate the drift change of the spectral line, and the second derivative is used to remove the drift caused by different slopes. Fig. 3(e) shows the logarithmic conversion spectrum, which shows the opposite trend to the original spectrum due to the conversion of reflectivity to absorbance. The spectral curve processed by the SG convolution smoothing method is shown in Fig. 3(f). Some irrelevant noises are removed on the basis of the original spectral data. Compared with the original image, the spectral curve is smoother. The spectral curves of MSC and SNV with pretreatment are shown in Fig. 3(g) and F. 5(h), respectively. The two processing methods have similar effects and belong to scattering processing, which can reduce the scattering effect during spectral data acquisition, improve the signal-to-noise ratio of spectral data, and correct the spectral baseline offset without affecting the spectral absorption data corresponding to the sample. The spectra treated by MSC and SNV are more compact than the original spectra. The MSC algorithm is mainly used to reduce the influence of uneven solid distribution and sample particle size on spectral variables, while SNV is mainly used to eliminate the influence of optical path change and surface scattering. The data after SG + SNV and SG + MSC mixed processing methods are shown in Fig. 3(i and j). Both graphs contain the preprocessing results of SG and compare them with the original spectrum. SG removes some noise and makes the results more compact.

This section analyses the correlation of the data preprocessing results. To evaluate the relative robustness of various preprocessing methods, the next step is modelling. The modelling results are shown in Section 3.3.

3.3. Analysis of modelling results

To analyse the micronear infrared spectrum, six modelling methods, including support vector machine (SVM), random forest (RF), extreme gradient boosting (XGBoost), gradient boosting tree (GBDT) adaptive boosting (AdaBoost) and neural network structure search (Autokeras), were selected. Different preprocessing methods are trained by cross validation, and 54 prediction models are established. In this test, K-fold cross validation requires human setup. Since several trials chose 5-fold cross-validation and achieved better experimental results(Tu et al., 2021). Therefore 5-fold cross validation was chosen in this experiment. The results are shown in Table 1.

According to the accuracy value of each model in Table 1, SVM is superior to RF, XGBoost, GBDT, AdaBoost and Autokeras in the classification and modelling of 'Akizuki' pear cork disease. In addition, the best result obtained by SVM is based on the original data; the accuracy rate is 84.65%, and the F1 score is 84.69%. The model has high discriminant ability. Through GBDT modelling, the results of the SG processed data are the worst, with an accuracy rate of 53.95% and an F1 score of 53.52%.

(1) Support vector machines have many advantages in solving small-sample, nonlinear and high-dimensional pattern recognition problems and are widely used in classification and recognition. Considering the nonlinear correlation between the absorption and the real value of the sample, the RBF kernel function is selected. The grid search method combined with the tenfold cross validation method is used to optimize the parameters of the penalty factors C and gamma in the RBF kernel function. When C = 850 and gamma = 1.0, the classification effect of SVM is the best. The accuracy of the training set is 95.39%, and the accuracy of the test set is 84.65%.

(2) RF and XGBoost are traditional ensemble learning algorithms, and XGBoost is an algorithm based on GBDT. RF modelling of the spectral data processed by the first derivative is carried out to obtain the best recognition effect. The accuracy of the training set is 99.40%, and the accuracy of the test set is 74.88%. For the spectral data processed by SG + MSC, XGBoost modelling is performed to obtain the best recognition effect. At this time, the accuracy of the training set is 97.19%, and the accuracy of the test set is 74.42%. The results obtained by modelling the data after LG processing are the worst results, and the accuracies of the test set are 62.33% and 60.00%, respectively.

(3) AdaBoost is also an ensemble learning algorithm that can combine multiple weak learners to become a strong learner. The spectral data processed by the first derivative are modelled to obtain the best recognition effect. The accuracy of the training set is 93.19%, and the accuracy of the test set is 79.07%. However, the accuracy of the data modelling results processed by SG is the lowest, and the accuracy of the test set is only 58.17%.

The low performance of traditional ensemble learning algorithms may be due to the relatively small number of real samples as the approximation target when generating samples. Therefore, compared with the real samples, their diversity is still relatively moderate.

(4) Autokeras is an architecture for automated machine learning that provides an automatic search for deep learning models. In this experiment, the initial model number (mls) and the number of iterations (epochs) were set. When mls is 50 and epochs are 500, the data modelling after MSC processing achieves the best classification effect. The accuracy of the training set is 96.19%, and the accuracy of the test set is 79.07%.

The analysis results of each model under different pretreatment methods are shown in Fig. 4. The results show that the accuracy and FI of different pretreatment models are the same. Fig. 4 shows that the accuracy and F1 values of SVM models under other preprocessing results are higher than those of other models except FD and SD processing. After the spectrum was processed by SG transformation, the classification ability of each model for the 'Akizuki' pear data was poor except for SVM and Autokeras. However, in SVM, the original data obtained the highest accuracy and F1 score on the test set. The better classification results in other preprocessing methods include MSC, SNV, SG + MSC, SG + SNV. This may be due to the light scattering shape and surface difference characteristics of the pear fruit surface during the analysis. Both MSC and SNV can reduce the occurrence of light scattering and baseline offset. SNV has been shown to minimize the scattering effect (Alisaac, Behmann, Kuska, Dehne, & Mahlein, 2018).

Due to similar characteristics, pears in asymptomatic parts are easily misclassified as healthy or diseased fruits. This is because for asymptomatic fruit, the internal components of the plant tissue have changed but not in appearance. The accuracy results of asymptomatic fruits are shown in Table 2. Although the overall recognition accuracy of the Autokeras model is lower than that of the SVM modelling results, for the classification of asymptomatic fruits, the modelling results of the data processed by SG are the best, reaching 90%, of which 4 data points are misjudged as diseased fruits and 4 data points are misjudged as healthy fruits. The spectral data after SD processing had the worst modelling results, only 43.75%, of which 15 data points were misjudged as diseased fruits and 30 data points were misjudged as healthy fruits. The comprehensive recognition accuracy of the original data of the SVM model is the highest, but for the classification of asymptomatic fruits, the spectral data after SNV are modelled, and the accuracy rate reaches

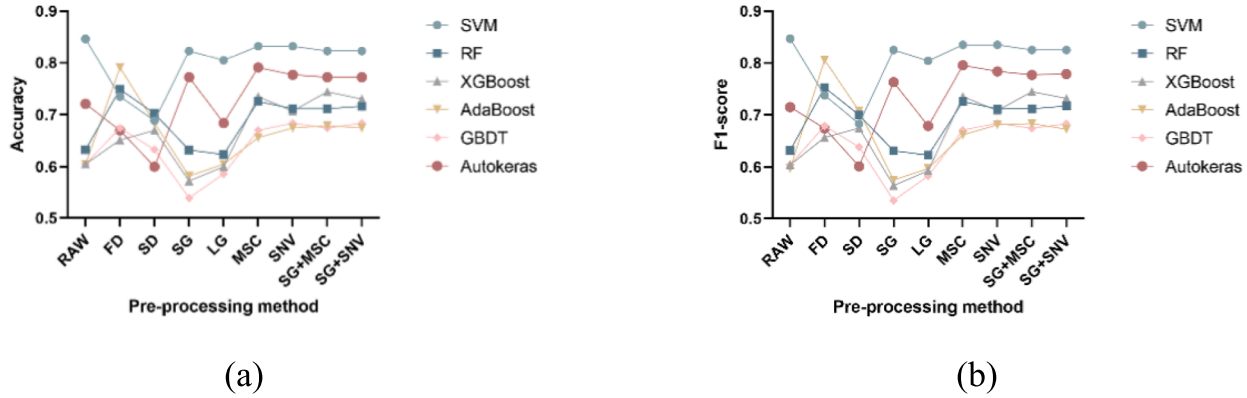


Fig. 4. The classification performance of different models under different preprocessing methods: (a) the accuracy of the test set; (b) F1 score of the test set.

Table 2
Classification accuracy of epidermal asymptomatic fruits based on Autokeras, SVM, RF, XGBoost, AdaBoost and GBDT models.

Models	Preprocessing method								
	RAW	FD	SD	SG	LG	MSC	SNV	SG + MSC	SG + SNV
Autokeras	86.25%	63.75%	43.75%	90%	70%	78.75%	76.25%	75%	73.75%
SVM	83.58%	80.60%	55.22%	86.57%	82.09%	83.58%	88.06%	85.29%	86.57%
RF	71.46%	83.58%	85.07%	71.64%	71.64%	86.57%	86.57%	85.07%	85.07%
XGBoost	65.67%	61.19%	62.87%	61.19%	64.18%	77.61%	77.61%	86.57%	80.60%
AdaBoost	68.29%	71.95%	52.43%	65.85%	68.29%	71.95%	71.95%	73.17%	82.93%
GBDT	59.76%	69.51%	60.98%	57.32%	60.98%	69.51%	71.95%	70.73%	70.73%

88.06%. Among them, 3 data points were misjudged as diseased fruits, and 5 data points were misjudged as healthy fruits. The spectral data after SD processing had the worst modelling results, only 55.22%, of which 15 data points were misjudged as diseased fruits and 15 data points were misjudged as healthy fruits. For RF, the accuracy of MSC and SNV was the highest, both 86.57%. The spectral data processed by SG + SNV are modelled, and the highest accuracies are obtained in AdaBoost and XGBoost, which are 82.93% and 86.57%, respectively. The data processed by SNV were modelled by GBDT, and the classification accuracy of the asymptomatic fruit was the highest, which was 71.95%.

The model can identify most healthy fruits, diseased fruits and asymptomatic fruits. However, because the appearance of asymptomatic fruits is too similar to the surface characteristics of healthy fruits, the recognition results are more or less biased. For the classification and recognition of asymptomatic fruits, the Autokeras model after SG treatment achieved the best results, reaching 90%, which could accurately differentiate asymptomatic fruits from diseased fruits and healthy fruits. The optimal result of the Autokeras model for the classification and recognition of healthy fruits, diseased fruits and asymptomatic fruits is 79.07%, and the classification result is also better.

In this experiment, asymptomatic fruits were classified as diseased or healthy fruits, resulting in poor accuracy of some models. The reason may be that asymptomatic fruits and healthy fruits are difficult to distinguish in appearance and are similar to diseased fruits in internal composition, resulting in easy misjudgment of fruits. This sample is a biological sample with a high degree of complexity, and the classification index of this model is acceptable under this study. The results of the model can be combined with the characteristics of fruit disease observed by the naked eye for comparative analysis to further improve the accuracy of asymptomatic fruit recognition and achieve the purpose of early identification of 'Akizuki' pear cork disease.

In summary, the modelling of Autokeras processed by SG achieves the best classification results in the identification of asymptomatic fruits, and the model has high robustness and accuracy. This study can provide more accurate results for the early identification of cork spot disorder in 'Akizuki' pear.

Based on near-infrared spectroscopy, the early recognition of 'Akizuki' pear cork disease was studied. The recognition results of asymptomatic fruit were poor, and it was easy to misclassify as healthy fruit and diseased fruit, resulting in low accuracy of individual classification. This may be because the test materials are field materials, and diseased fruits appeared in the field. Some fruits may have been diseased, but their appearance has not yet shown related symptoms. This paper mainly focuses on the early identification of 'Akizuki' pear cork disease. Considering the variability of asymptomatic samples in the later stage, the classification indexes obtained in this study are acceptable.

4. Conclusion

Studies have shown that it is feasible to use the Autokeras method to detect early asymptomatic 'Akizuki' pear cork disease under near-infrared spectroscopy. These studies have found that near-infrared spectroscopy can classify healthy, asymptomatic and diseased fruits based on differences in mineral elements and related components in diseased, asymptomatic and healthy fruits. Compared with the original data, SG eliminates most of the effects caused by noise and nonsmooth surfaces. However, in the field environment, there are many interference factors, and the appearance of asymptomatic fruit and healthy fruit is difficult to distinguish. The 'Akizuki' pear under natural conditions may be diseased without any symptoms, resulting in the incorrect classification of healthy pears. The internal components of asymptomatic fruits and diseased fruits are similar, and the misjudgment of the two types is within the normal range. This study provides a new perspective and idea for the study of early asymptomatic identification of 'Akizuki' pear cork disease, which can be used for early disease assessment and prediction in 'Akizuki' pear planting areas.

CRediT authorship contribution statement

Li Liu: Conceptualization, Methodology, Validation, Writing – review & editing, Resources, Funding acquisition. **Hanhan Zhang:** Software, Formal analysis, Data curation, Investigation, Writing – original

draft, Visualization. **Lin Wu:** Methodology, Investigation. **Shangfeng Gu:** Data curation, Investigation. **Jing Xu:** Investigation. **Bing Jia:** Investigation. **Zhenfeng Ye:** Investigation. **Wei Heng:** Resources. **Xiu Jin:** Investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This research was supported by the China Agriculture Research System of MOF and MARA, the Key Research Project of Natural Science in Colleges and Universities of the Anhui Provincial Department of Education in 2021 (KJ2021A0188), and Anhui Agricultural University Talent Project (rc322213).

References

Alaiad, A., Migdady, A., Al-Khatib, R.e. M., Alzoubi, O., Zitar, R. A., & Abualigah, L. (2023). Autokeras Approach: A Robust Automated Deep Learning Network for Diagnosis Disease Cases in Medical Images. *Journal of Imaging, 9*(3), 64. https://doi.org/10.3390/jimaging9030064

Alazzam, I., Alsmadi, I., & Akour, M. (2017). Software fault proneness prediction: A comparative study between bagging, boosting, and stacking ensemble and base learner methods. *International Journal of Data Analysis Techniques and Strategies, 9*(1), 1. https://doi.org/10.1504/IJDATS.2017.083058

Alisaac, E., Behmann, J., Kuska, M. T., Dehne, H. W., & Mahlein, A. K. (2018). Hyperspectral quantification of wheat resistance to Fusarium head blight: Comparison of two Fusarium species. *European Journal of Plant Pathology, 152*, 869–884. https://doi.org/10.1007/s10658-018-1505-9

Anderson, N. T., & Walsh, K. B. (2022). Review: The evolution of chemometrics coupled with near infrared spectroscopy for fruit quality evaluation. *Journal of Near Infrared Spectroscopy, 33*(3), 243–257. https://doi.org/10.1177/09670335211057235

Ba, W., Jin, X., Lu, J., Rao, Y., Zhang, T., Zhang, X., … Li, S. (2022). Research on predicting early Fusarium head blight with asymptomatic wheat grains by micro-near infrared spectrometer. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 287*, Article 122047. https://doi.org/10.1016/j.saa.2022.122047

Barton, F. E., & Himmelsbach, D. S. (1993). Two-Dimensional Vibrational Spectroscopy II: Correlation of the Absorptions of Lignins in the Mid- and Near-Infrared. *Applied Spectroscopy, 47*(11), 1920–1925. https://doi.org/10.1366/0003702934066091

Bhargava, A., & Bansal, A. (2018). Fruits and vegetables quality evaluation using computer vision: A review. *Journal of King Saud University - Computer and Information Sciences, 33*(3), 243–257. https://doi.org/10.1016/j.jksuci.2018.06.002

Bin, J., Ai, F.-F., Fan, W., Zhou, J.-H., Yun, Y.-H., & Liang, Y.-Z. (2016). A modified random forest approach to improve multi-class classification performance of tobacco leaf grades coupled with NIR spectroscopy. *RSC Advances, 6*(36), 30353–30361. https://doi.org/10.1039/c5ra25052h

Cadet, F., & Offmann, B. (2006). Evidence for Potassium-Sucrose Interaction in Biological Mid-Infrared Spectra by Multidimensional Analysis. *Spectroscopy Letters, 29*(7), 1353–1365. https://doi.org/10.1080/00387019608007128

Ciavarella, S., Batten, G. D., & Blakeney, A. B. (1998). Measuring potassium in plant tissues using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy, 6*(A), A63–A66. https://doi.org/10.1255/jnirs.167

Clark, D. H., Mayland, H. F., & Lamb, R. C. (1987). Mineral Analysis of Forages with near Infrared Reflectance Spectroscopy. *Agronomy Journal, 79*(3), 485–490.

Cui, Z., Jiao, Q., Wang, R., & Ma, C. (2020). Investigation and analysis of relationship between mineral elements alteration and cork spot physiological disorder of Chinese pear 'Chili' (Pyrus bretschneideri Rehd.). *Scientia Horticulturae, 260*(108883). https://doi.org/10.1016/j.scienta.2019.108883

De Aldana, B. R. V., Criado, B. G., Ciudad, A. G., & Corona, M. E. P. (1995). Estimation of mineral content in natural grasslands by near infrared reflectance spectroscopy. *Communications in Soil Science and Plant Analysis, 26*(9–10), 1383–1396. https://doi.org/10.1080/00103629509369379

de Freitas, S. T., do Amarante, C. V. T., Labavitch, J. M., & Mitcham, E. J. (2010). Cellular approach to understand bitter pit development in apple fruit. *Postharvest Biology and Technology, 57*(1), 6–13. https://doi.org/10.1016/j.postharvbio.2010.02.006

El-Mesery, H., Mao, H., & Abomohra, A. (2019). Applications of Non-destructive Technologies for Agricultural and Food Products Quality Inspection. *Sensors, 19*(4), 846. https://doi.org/10.3390/s19040846

Freund Y., Schapire R.E. (1997). *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting.* Paper presented at the Conference on Learning Theory.

Ho, L. C., & White, P. J. (2005). A Cellular Hypothesis for the Induction of Blossom-End Rot in Tomato Fruit. *Annals of Botany, 95*(4), 571–581. https://doi.org/10.1093/aob/mci065

Jamshidi, B., Mohajerani, E., & Jamshidi, J. (2016). Developing a Vis/NIR spectroscopic system for fast and non-destructive pesticide residue monitoring in agricultural product. *Measurement, 89*, 1–6. https://doi.org/10.1016/j.measurement.2016.03.069

Li, J., Sun, L., & Li, R. (2020). Nondestructive detection of frying times for soybean oil by NIR-spectroscopy technology with Adaboost-SVM (RBF). *Optik - International Journal for Light and Electron Optics, 206*(September (6)), Article 164248. https://doi.org/10.1016/j.ijleo.2020.164248

Lu, J., Ehsani, R., Shi, Y., Abdulridha, J., de Castro, A. I., & Xu, Y. (2017). Field detection of anthracnose crown rot in strawberry using spectroscopy technology. *Computers and Electronics in Agriculture, 135*, 289–299. https://doi.org/10.1016/j.compag.2017.01.017

Lu, H., Zheng, H., Hu, Y., Lou, H., & Kong, X. (2011). Bruise detection on red bayberry (Myrica rubra Sieb. & Zucc.) using fractal analysis and support vector machine. *Journal of Food Engineering, 104*(1), 149–153. https://doi.org/10.1016/j.jfoodeng.2010.12.007

Mammone, A., Turchi, M., & Cristianini, N. (2009). Support vector machines. *Wiley Interdisciplinary Reviews: Computational Statistics, 13*(4), 18–28. https://doi.org/10.1109/5254.708428

Miqueloto, A., Amarante, C. V. T.d., Steffens, C. A., dos Santos, A., & Mitcham, E. (2014). Relationship between xylem functionality, calcium content and the incidence of bitter pit in apple fruit. *Scientia Horticulturae, 165*, 319–323. https://doi.org/10.1016/j.scienta.2013.11.029

Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., & Rutledge, D. N. (2020). New data preprocessing trends based on ensemble of multiple preprocessing techniques. *Trends in Analytical Chemistry, 132*(1), Article 116045. https://doi.org/10.1016/j.trac.2020.116045

Mogollón, M. R., Contreras, C., de Freitas, S. T., & Zoffoli, J. P. (2021). NIR spectral models for early detection of bitter pit in asymptomatic 'Fuji' apples. *Scientia Horticulturae, 280*, Article 109945. https://doi.org/10.1016/j.scienta.2021.109945

Oliveri, P., Malegori, C., Simonetti, R., & Casale, M. (2019). The impact of signal pre-processing on the final interpretation of analytical outcomes – A tutorial. *Analytica Chimica Acta, 1058*, 9–17. https://doi.org/10.1016/j.aca.2018.10.055

Qian, Z., Wenqian, H., Qingyan, W., Jingzhu, W., & Jiangbo, L. (2022). Detection of pears with moldy core using online full-transmittance spectroscopy combined with supervised classifier comparison and variable optimization. *Computers and Electronics in Agriculture, 200*, Article 107231. https://doi.org/10.1016/j.compag.2022.107231

Schönherr, J., & Bukovac, M. J. (1973). Ion exchange properties of isolated tomato fruit cuticular membrane: Exchange capacity, nature of fixed charges and cation selectivity. *Planta, 109*(1), 73–93. https://doi.org/10.1007/bf00385454

Shi, H., & Yu, P. (2017). Comparison of grating-based near-infrared (NIR) and Fourier transform mid-infrared (ATR-FT/MIR) spectroscopy based on spectral preprocessing and wavelength selection for the determination of crude protein and moisture content in wheat. *Food Control, 82*, 57–65. https://doi.org/10.1016/j.foodcont.2017.06.015

Tu, Y., Chen, B., Lang, W., Chen, T., Li, M., Zhang, T., & Xu, B. (2021). Uncovering the Nature of Urban Land Use Composition Using Multi-Source Open Big Data with Ensemble Learning. *Remote Sensing, 13*(21), 4241. https://doi.org/10.3390/rs13214241

Watson, M. E., & Galliher, T. L. (2001). Comparison of Dumas and Kjeldahl methods with automatic analyzers on agricultural samples under routine rapid analysis conditions. *Communications in Soil Science and Plant Analysis, 32*(13–14), 2007–2019. https://doi.org/10.1081/css-120000265

Woodbridge, C. G. (1971). Calcium level of pear tissues affected with cork and black end. *HortScience: A publication of the American Society for. Horticultural Science, 6*(5), 451–453. https://doi.org/10.21273/HORTSCI.6.5.451

Xia, J., Qin, Z., Jumei, Y., Yingying, D., & Bingyu, L. (2022). Remote Sensing Monitoring of Winter Wheat Stripe Rust Based on mRMR-XGBoost Algorithm. *Remote Sensing, 14*(3), 756. https://doi.org/10.3390/rs14030756

Xia, Y., Xu, Y., Li, J., Zhang, C., & Fan, S. (2019). Recent advances in emerging techniques for non-destructive detection of seed viability: A review. *Artificial Intelligence in Agriculture, 1*, 35–47. https://doi.org/10.1016/j.aiia.2019.05.001

Yaqian, C., Xinyuan, S., Zhisheng, N.a., & Qiao. (2016). MDL and RMSEP assessment of spectral pretreatments by adding different noises in calibration/validation datasets. *Spectrochimica acta. Part A. Molecular and biomolecular spectroscopy, 163*, 20–27. https://doi.org/10.1016/j.saa.2016.03.017

Yujie, W., Qingqing, C., Shanshan, J., Chao, Z., Yonghua, L., Yilei, Y., … Zhengzhu, Z. (2022). Tea Analyzer: A low-cost and portable tool for quality quantification of postharvest fresh tea leaves. *LWT - Food Science and Technology, 159*, Article 113248. https://doi.org/10.1016/j.lwt.2022.113248

Zahra, B., Bahareh, J., Mansour, R., & Yousef, A. G. (2018). Detection of sunn pest-damaged wheat samples using visible/near-infrared spectroscopy based on pattern recognition. *Spectrochimica Acta Part A Molecular and Biomolecular Spectroscopy, 203*, 308–314. https://doi.org/10.1016/j.saa.2018.05.123

Zhang, X., & Cui, Z. (2023). Review of fruit cork spot disorder of Asian pear (Pyrus spp.). Frontiers. *Plant Science, 14*, Article 1211451. https://doi.org/10.3389/fpls.2023.1211451

Zhang, W., Liu, Y., Chen, K., Li, H., Duan, Y., Wu, W., … Guo, W. (2021). Lightweight Fruit-Detection Algorithm for Edge Computing Applications. *Frontiers in Plant Science, 12*, Article 740936. https://doi.org/10.3389/fpls.2021.740936

Zhang, L., Nie, X. Y., & Le, T. (2017). Rapid determination of 15 metallic elements in 2 kinds ofcoixseeds by inductively coupled plasma-mass spectrometry after closed-vessel microwave digestion. *Journal of Food Process Engineering*. https://doi.org/10.1111/jfpe.12528