RESEARCH

Non-destructive online detection of early moldy core apples based on Vis/NIR transmission spectroscopy

Xiaogang Jiang[1,2], Kang Ge[1,2], Zhi Liu[1,2], Nan Chen[1,2], Aiguo Ouyang[1,2], Yande Liu[1,2*], Yuyang Huang[3], Jinghu Li[4] and Mingmao Hu[5]

Abstract

Apple moldy core is a fungus-infested disease that is extremely insidious, usually occurring inside the fruit, making it very difficult to distinguish from the exterior with the naked eye. Using VIS/NIR transmission spectroscopy, this study successfully detected moldy core apples. By combining four wavelength selection algorithms (CARS, CARS-SPA, MC-UVE, and MC-UVE-SPA) with four classifiers (SVM, ELM, KNN, and LDA-KNN), discrimination models were established for two-class and three-class classifications. MC-UVE-SPA-LDA-KNN achieved an AUC of 0.99 and an accuracy of 98.82% for two-class classification, while MC-UVE-SPA achieved an AUC of 0.99 and an accuracy of 97.64% for three-class classification. This confirms MC-UVE-SPA as an effective tool for selecting wavelengths specific to moldy core apples, facilitating precise identification and differentiation of apple states. This study advances dynamic online detection of early-stage moldy core conditions in apples, reducing post-harvest disease occurrence and preserving fruit quality effectively.

Keywords Moldy core, Transmission spectroscopy, Online detection, Wavelength selection
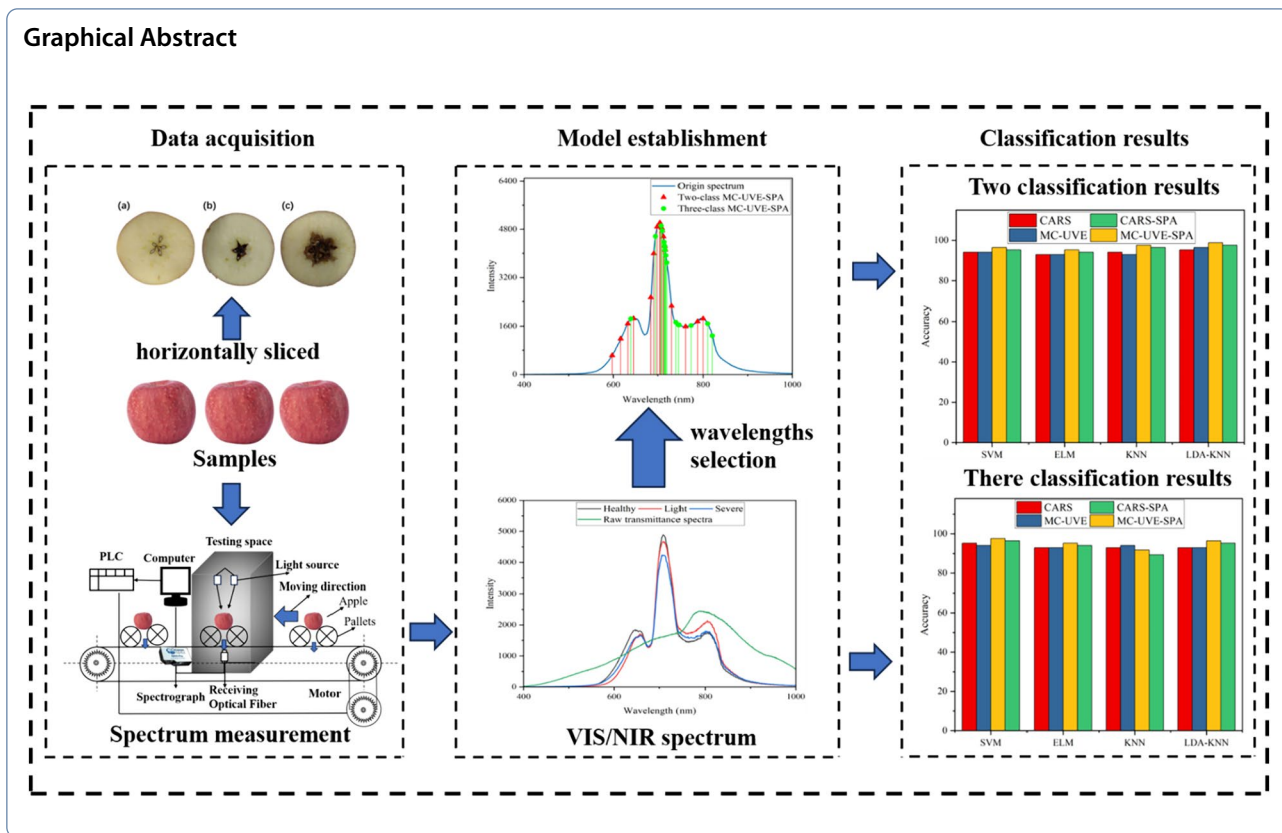
---

*Correspondence:
Yande Liu
jxliuyd@163.com
Full list of author information is available at the end of the article

Jiang *et al. Chem. Biol. Technol. Agric.* (2024) 11:63

Page 2 of 12

Graphical Abstract

Introduction

In recent years, apples have become a widely grown and consumed fruit around the world, making a significant contribution to the world's food supply. According to the statistics of the FAOSTAT (Food and Agriculture Organization), the total apple production reached 83 million tons in 2023. As one of the world's best-selling fruits, the quality of apples is closely related to consumer purchasing demand. Apple moldy core is a disease caused by fungal infection leading to internal decay [1]. In the initial stage of fungal infection, small dark brown spots are formed in the center of the apple core, which then becomes dark brown over time. The infected part grows grey or white hyphae, and the bacteria continue to expand outward, eventually leading to fruit deterioration from the inside out, thus seriously affecting apples' quality and shelf life [2]. This disease occurs mainly inside the fruit, and its diseased fruit is not visibly different from normal fruit in appearance, and can only be detected as diseased if the apple is cut open [3]. Once infected fruits are considered healthy and enter the market, it will not only cause food safety problems, but also affect the reputation of the fruit growers and the place of origin, and even lead to the bulk return of the export trade, which will inevitably affect the development of the apple industry. If apple moldy core

can be effectively identified at its early stage, the spread of the disease can be effectively prevented. Hu et al. developed a classification model of the moldy core using the acquired NIR transmittance spectra and back propagation network with a success rate of 95% [4]. However, due to the hidden nature of its early lesions and low detection rate. Therefore, there is an urgent need to achieve accurate, rapid, and nondestructive detection of early moldy core in apples, to safeguard the quality of apples.

Non-destructive detection methods for fruits that are commonly used include magnetic resonance imaging (MRI) [5], X-ray [6], acoustic vibration technology [7], electronic nose technology [8], and Vis/NIR [9] Because of its high detection rate, accuracy, efficiency, and cost-effectiveness, Vis/NIR spectroscopy technology has gained favor in the non-destructive testing of agricultural products [10]. According to the difference of frequency combination and frequency doubling absorption intensity of C–H, N–H, O–H, and other groups in spectra, combined with chemometrics methods, the internal physiological diseases of fruit can be identified [11]. To achieve fruit quality detection, Yao et al. [12] used a homemade portable near-infrared diffuse reflectometer for evaluating and monitoring the SSC of apples. Ma et al. [13] designed a multi-fiber-based Vis/NIR spatially resolved

Jiang *et al. Chem. Biol. Technol. Agric.* (2024) 11:63

Page 3 of 12

spectrometry system for the simultaneous evaluation of soluble solids content and hardness of apples. Baranowski [14] used a hyperspectral camera to Combine Vis/NIR and SWIR spectral properties for the prediction of days to bruise in apples. Cruz et al. [15] explored the use of VIS/NIR spectrometry for the prediction of SSC, hardness, and presence of internal browning in 'Rocha' pears. Burdon [16] used the flesh color of kiwifruit, SSC, hardness, and dry matter for nondestructive prediction of the presence or absence of frostbite during its storage period. The above study mainly utilized the diffuse reflectance mode to detect fruit quality. Since early moldy core usually occurs inside apples, it is more effective to use near-infrared transmission mode than diffuse reflection mode for its internal quality detection [17]. Several online systems utilizing NIR spectroscopy have been developed to address the need for quality testing and classification of a significant quantity of fruits. These systems are capable of detecting internal fruit quality, including freezing damage [18] and blackheart [19], among others. To enhance the performance of the model, it is possible to implement effective methods for selecting wavelengths. These methods not only improve the performance of the model but also decrease its complexity and runtime [20]. Some commonly employed methodologies encompass principal component analysis (PCA) [21], Monte Carlo-uninformative variable elimination (MC-UVE) [22], competitive adaptive reweighted sampling (CARS) [23], and successive projection algorithm (SPA) [24]. This study utilized the MC-UVE-SPA algorithm as the wavelength selection method to dynamically detect moldy core in apples. The rationale behind this choice was the potential synergistic effects that could be achieved by combining two complementary wavelength selection algorithms.

This study aims to develop a classification model suitable for dynamic online that can accurately detect early moldy core apples. The main aims of this study are as follows: (1) Study of Vis/NIR transmission spectroscopy combined with various classification models for the identification of early moldy core apples. (2) Four classification algorithms (SVM, ELM, KNN, and LDA-KNN) were utilized to classify two-class model including healthy and moldy core, and three-class model including healthy, light, and severe. (3) The determination of effective wavelengths for detecting apple moldy core was conducted using four variable selection methods: CARS, CARS-SPA, MC-UVE, and MC-UVE-SPA.

Materials and method

Apple samples

A total of 340 Red Fuji apples, all from the same growing area within the orchard and confirmed by the farmer to be infected with moldy core, were collected at the same stage of ripeness and presenting a similar coloration, thus ensuring consistency in growing conditions, ripeness and coloration. The apples were carefully selected to ensure that their appearance was free from any physical damage. Immediately upon arrival at the lab, the apples are washed and then air-dried. Each apple is then labeled and placed in an environment with a room temperature of 23 °C and a relative humidity of 30% to 40% for 24 h. After a resting period, spectral data collection began.

Transmittance spectra acquisition

The VIS/NIR spectroscopy collection system utilized in this study was a dynamic online transmittance detection device that was independently researched and developed [25], the specific structure is shown in Fig. 1. The QE65Pro spectrometer from Ocean Optics was selected for this study, which has higher sensitivity and lower stray light and excellent thermal stability. The light sources were 2 rows of 5 halogen lamps, each with a specific specification of 12 V and 100 W. The apples were carefully arranged on the fruit cup to guarantee that their stem axes were parallel and precisely aligned with the conveyor belt. They were then transported to a dark box for measurement. When the halogen lamps illuminated the apples, the transmitted light entering the apple was captured by optical fibers and transmitted to a computer through a spectrometer. The spectrometer had a wavelength range of 350 to 1150 nm and required a 30-min preheating before acquiring spectral data.

Moldy core degree assessment

The apple samples were horizontally sliced to conduct moldy core discrimination. Images of the apples were captured using a camera to assess the disease's degree. The CAD software was utilized to determine the proportion of the impacted region within the cross-sectional area. Based on the percentage, the severity of moldy core disease was classified into three groups: Healthy apples (Sd = 0), Light moldy core (Sd < 10%), and Severe moldy
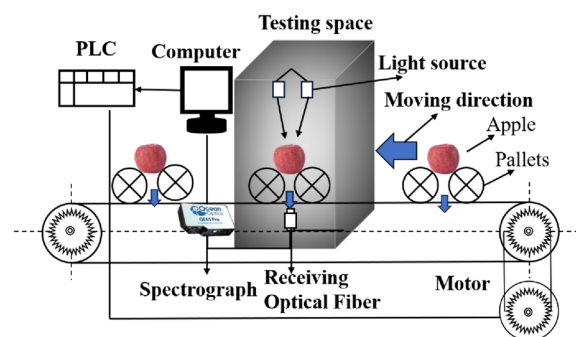

Fig. 1 Near-infrared spectra acquisition device

Jiang *et al. Chem. Biol. Technol. Agric.*   (2024) 11:63

Page 4 of 12

core (Sd > 10%). The level of internal disease severity in the apples is illustrated in Fig. 2.

Spectral pre-processing and effective wavelength selection algorithm

VIS/NIR spectra frequently display noise, baseline drift, and similar phenomena, which can be attributed to instrument variations, sample backgrounds, and other influencing factors. To address the potential impact of these detrimental factors on the model, this study utilizes three preprocessing techniques: Savitzky–Golay (S–G) smoothing, Multiplicative Scatter Correction (MSC), and Standard Normal Variate (SNV). Finally, a comparison and analysis are conducted between the raw spectra and the preprocessed spectra to ascertain the most suitable preprocessing method for the construction of the model. However, the presence of redundant or uninformative variables in spectral data may affect the performance and computational efficiency of classification models. To address this issue, four variable selection algorithms were chosen to extract the effective wavelengths, such as CARS, CARS-SPA, MC-UVE, and MC-UVE-SPA. These effective wavelengths are subsequently combined with the classifiers to improve the accuracy of two-class and three-class classifications.

Classification algorithms

Selecting a suitable classification algorithm is of utmost importance in enhancing the accuracy of classification. In this article, we have chosen four classification algorithms, namely K-Nearest Neighbors (KNN), Linear Discriminant Analysis-K Nearest Neighbors (LDA-KNN), Extreme Learning Machine (ELM), and Support Vector Machine (SVM), to conduct discriminant analysis on the moldy core of apples.

KNN is a classification algorithm that functions based on principle of proximity [26]. When classifying a new sample, the classification algorithm identifies the K nearest training samples and assigns the new sample to the category with the highest number of votes, which is determined based on the categories of these K neighbors.

LDA-KNN is a hybrid approach that combines LDA and KNN [27]. Initially, LDA is utilized to reduce the dimensionality of the original data. This process entails the computation of the between-class scatter and within-class scatter to determine the most suitable discriminant vectors and subsequently project the data onto a discriminant subspace of lower dimensionality. Afterward, the KNN is employed to classify the dimension-reduced data. It achieves this by calculating the distances between the sample to be classified and the training samples that have been dimension reduced. The primary benefit of the LDA-KNN algorithm lies in its efficient utilization of LDA for reducing dimensionality and extracting features, in conjunction with KNN for classification. By implementing classification in a lower-dimensional space, this approach effectively enhances classification accuracy and generalization, while simultaneously reducing the dimensionality of the data.

The Extreme Learning Machine (ELM) is an algorithm for single-hidden-layer feedforward neural networks. It has been shown to exhibit superior generalization performance and higher learning efficiency during training when compared to traditional feedforward network learning algorithms. This characteristic renders it an appealing technique in the field of machine learning [28]. ELM demonstrates notable efficacy in various tasks, including classification, regression, and density estimation. By employing suitable network architectures and parameter settings, Extreme Learning Machines (ELM) have demonstrated the ability to effectively capture intricate patterns within data, leading to favorable learning results and extensive practical implementations.

Support Vector Machine (SVM), which was proposed by Vapnik [29] as a statistical learning theory, is a highly influential and extensively employed machine learning
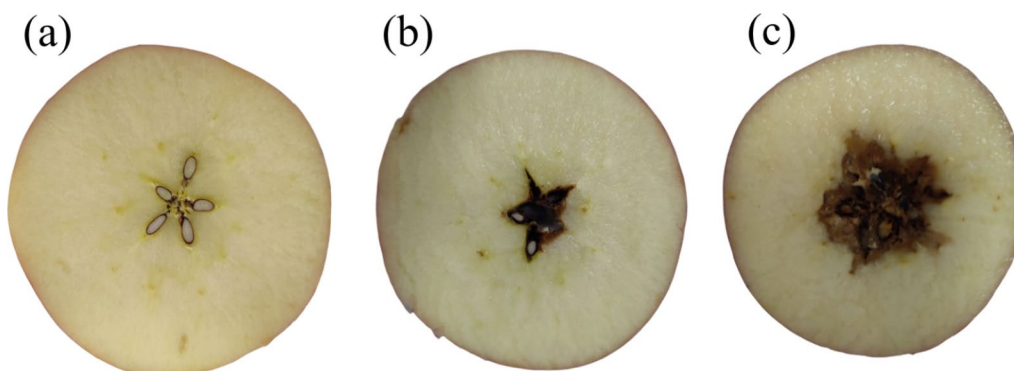

Fig. 2 Healthy (**a**), light (**b**), severe (**c**)

Jiang *et al. Chem. Biol. Technol. Agric.*   (2024) 11:63

Page 5 of 12

algorithm utilized for both classification and regression tasks. The fundamental concept revolves around the identification of an optimal hyperplane that maximizes the margin between various classes, thereby effectively segregating the data into distinct categories. SVM demonstrates exceptional performance across a wide range of applications, particularly in effectively managing datasets with high dimensions and tackling intricate classification tasks.

Establishment of classification models

This study developed two-class classification models, distinguishing between healthy and moldy core samples, as well as three-class classification models, differentiating between healthy, light, and severe. The main aim of developing the three-class model was to evaluate its effectiveness in identifying apples with a light moldy core. Accurately identifying apples exhibiting a mild moldy core is of utmost importance to prevent their further deterioration into a severe state and to safeguard the

well-being of other unaffected fruits. Additionally, the identification of moldy core can aid farmers in implementing targeted orchard management strategies aimed at mitigating the prevalence of diseased fruits. For two classes, a dataset consisting of 100 healthy apples and 240 apples with moldy core was utilized. In the three classes, there were a total of 340 apples categorized as follows: 100 apples were classified as healthy, 160 apples had a mild moldy core, and 80 apples had a severe moldy core. As shown in Table 1, the samples were divided into calibration and prediction sets using the Kennard–Stone algorithm in a 3:1 ratio [30].

Results and discussion
Spectral features

Figure 3a displays the spectra of healthy and moldy core apples. From the figure, it is evident that there exist three distinct absorption peaks at approximately 650 nm, 715 nm, and 815 nm. The absorption peak observed at approximately 650 nm is attributed to chlorophyll [31], whereas the peak at 715 nm is predominantly linked to the second harmonic stretching vibration of C-H and O–H bonds. The peak observed at 815 nm can be primarily attributed to the second harmonic absorption of C–H and N–H bonds [32]. From the figure, it is evident that the spectral curves of healthy and moldy core apples exhibit a high degree of similarity. Additionally, there is a certain degree of spectral overlap between healthy and diseased apples, which poses a challenge for direct

Table 1 Moldy core degree classification and sample set division

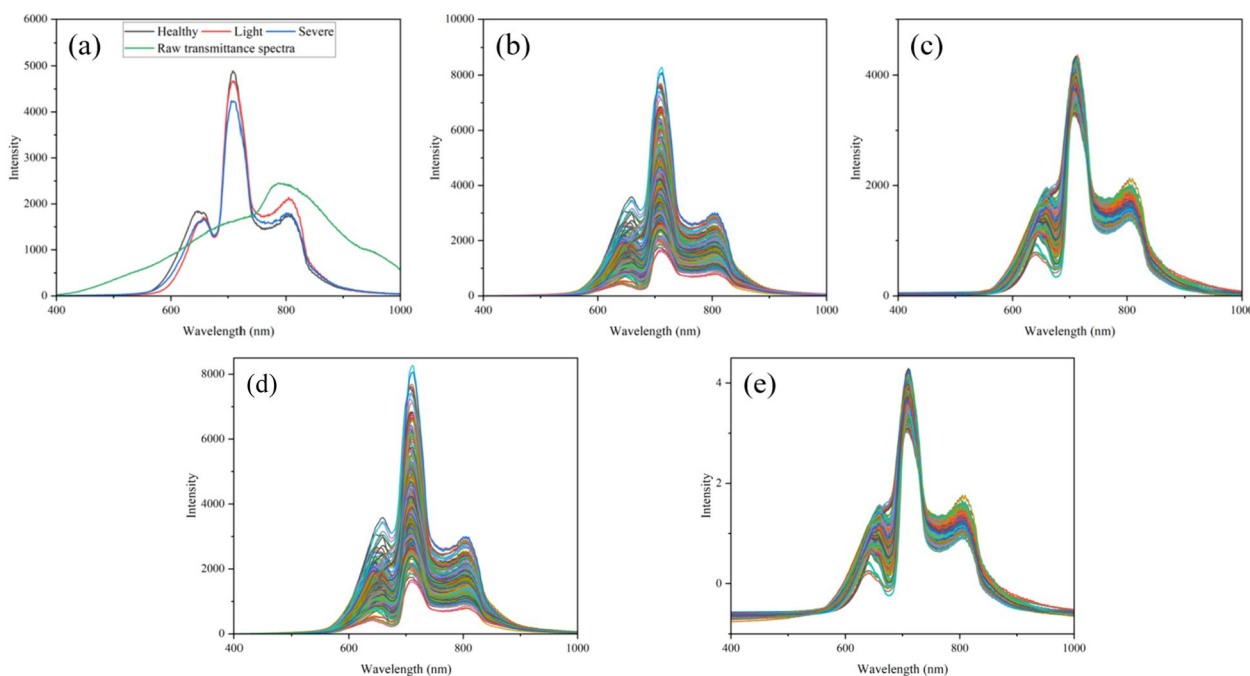| Moldy core degree	Area ratio	Total	Calibration set	Prediction set
Healthy	0	100	75	25
light	< 10%	160	120	40
Severe	≥ 10%	80	60	20

Fig. 3 Three types of different states and original transmission spectra (**a**), origin transmittance spectra (**b**), MSC (**c**), S–G smoothing (**d**), and SNV (**e**)

Jiang *et al. Chem. Biol. Technol. Agric.*   (2024) 11:63

Page 6 of 12

differentiation. Despite the observed differences in spectral intensities between healthy and moldy core apples, and the fact that the characteristic peaks were not significantly different, directly categorizing the three apples by observing only the spectral information is still a great challenge.

The development of classification models that effectively integrate the information present in spectra is crucial for achieving precise identification of diverse samples. In addition, to improve the accuracy of the spectra and to build a reliable classification model, three preprocessing techniques, namely SNV, Savitzky–Golay smoothing, and MSC, were used to eliminate unwanted factors such as noise, baseline drift, and light scattering from the original transmission spectra data, where Fig. 3b is the raw transmission spectrum, Fig. 3c is MSC, Fig. 3d is S–G smoothing and Fig. 3e is SNV.

Two-class classification by full wavelengths

Table 2 provides a comprehensive overview of the results obtained from two-class experiments using SVM, ELM, KNN, and LDA-KNN models. Compared to the unprocessed raw spectra, the utilization of the MSC, Savitzky–Golay smoothing, and SNV preprocessing techniques led to improved prediction accuracy of the model. Among the four classification models that were investigated, SVM and LDA-KNN exhibited exceptional performance. Specifically, the utilization of the SNV preprocessing technique resulted in an enhancement of prediction accuracy from 89.41 to 92.94%. The classification

performance of the ELM and KNN models was found to be satisfactory. The ELM model, when utilizing the MSC preprocessing technique, demonstrated an increase in predictive accuracy from 90.59 to 92.94%. The application of the KNN model, in conjunction with SNV preprocessing, resulted in a significant improvement in prediction accuracy, the accuracy increased from 87.06 to 90.58%. The analysis of the obtained results demonstrates that the implementation of preprocessing techniques led to diverse enhancements for different classifiers. Among the various methods that were evaluated, the SNV preprocessing technique exhibited the most superior performance, resulting in a significant enhancement in accuracy.

Effective wavelengths selection based on two-class classification

Four variable selection algorithms, namely CARS, CARS-SPA, MC-UVE, and MC-UVE-SPA, were utilized to identify optimal wavelengths from an initial pool of 1044 spectrums to enhance and streamline the model. Taking the two-class classification as an example, Fig. 4a shows the variation of RMSECV with the number of repetitions in the 50 rounds of repetition, from which it can be seen that the RMSECV value is the smallest at the 18th round, and the number of selected variables is 119. Fig. 4b demonstrates the results of MC-UVE in determining the optimal wavelength. The two red dashed lines indicate the stability boundaries and the variables located on the boundaries

Table 2 The two-class classification results based on SVM, ELM, KNN, and LDA-KNN models with full spectral

| Classification algorithms	Preprocessing methods	Accuracy (%)	Calibration set		Accuracy (%)	Prediction set	
			Healthy 75	Diseased 180		Healthy 25	Diseased 60
SVM	None	91.37	63	170	89.41	16	60
	MSC	93.33	65	173	91.76	18	60
	S–G smoothing	92.94	65	172	90.59	17	60
	SNV	95.29	68	175	92.94	19	60
ELM	None	91.37	61	172	90.59	18	59
	MSC	93.33	63	175	92.94	20	59
	S–G smoothing	91.37	61	172	90.59	18	59
	SNV	92.94	63	174	92.94	20	59
KNN	None	89.80	60	169	87.06	15	59
	MSC	92.16	61	174	89.41	17	59
	S–G smoothing	91.37	63	170	88.24	16	59
	SNV	92.94	65	172	90.59	18	59
LDA-KNN	None	92.94	66	171	89.41	16	60
	MSC	94.11	70	170	92.94	19	60
	S–G smoothing	94.90	69	173	91.76	19	59
	SNV	96.08	70	175	92.94	19	60
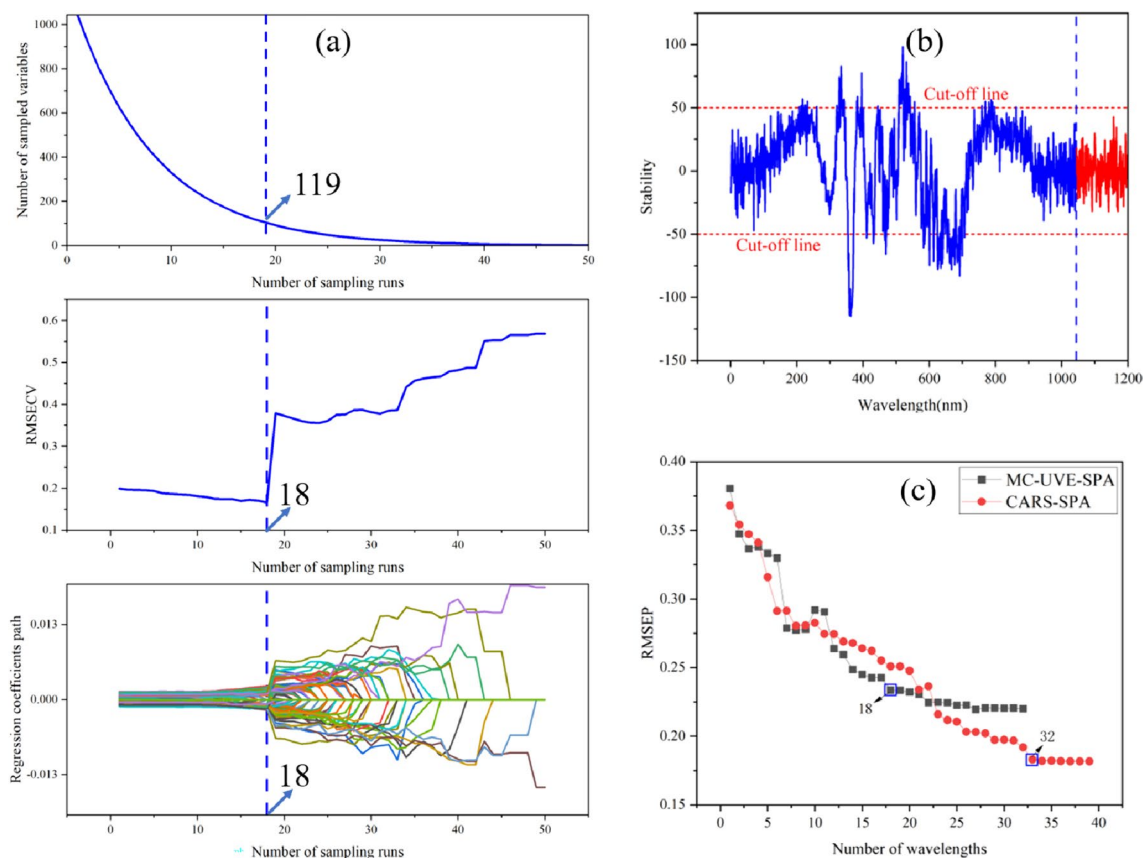
Fig. 4 Two-class wavelength selection by CARS (**a**), wavelength selection by MC-UVE (**b**), and wavelength selection by SPA (**c**)

are regarded as non-informative variables. When constructing the model, the wavelength corresponding to the stable value within the boundary is chosen. The red curve indicates noise, which should be ignored. In the end, the MC-UVE method selected a total of 295 wavelengths from the entire spectrum, while the wavelength selection algorithm using CARS identified 119 wavelengths. 295 and 119 wavelengths were retained by MC-UVE and CARS, respectively, which were then used as inputs to the SPA. The SPA eliminates redundant variables, thus reducing the complexity of the model. Fig. 4c depicts the wavelength selection trend of the models using CARS-SPA and MC-UVE-SPA as the number of wavelengths increases. The black square in the figure corresponds to MC-UVE-SPA, which has an optimal RMSEP value of 0.2335. The red circle in the figure denotes CARS-SPA, which records an optimal RMSEP value of 0.1918. In the end, 32 and 18 optimal wavelengths were identified by the CARS-SPA and MC-UVE-SPA methods, respectively.

Two-class classification based on the effective wavelengths
Figure 5 shows the TPR, TNR, and Accuracy histogram of the SVM, ELM, KNN, and LDA-KNN classification models established for healthy and moldy core apples using the effective wavelengths determined by CARS, CARS-SPA, MC-UVE, and MC-UVE-SPA. By comparing TPR in Fig. 5a, TNR in Fig. 5b, and Accuracy in Fig. 5c it can be found that after selecting the effective wavelengths, the model classification performances of all the four classifiers are improved compared with the full spectrum model, which indicates that the selected effective wavelengths are highly representative. On the prediction set, we can see that the TPR of the model prediction results are mixed, indicating that the model has a high misclassification rate of healthy apples, which is not conducive to the guidance of the actual production. The MC-UVE-SPA-LDA-KNN model phenotype is better than the others, probably because the dimensionality reduction of the data by LDA facilitates the subsequent KNN for classification. It is worth mentioning that the TNR and Accuracy of all model classification results performed well, probably because the models were more
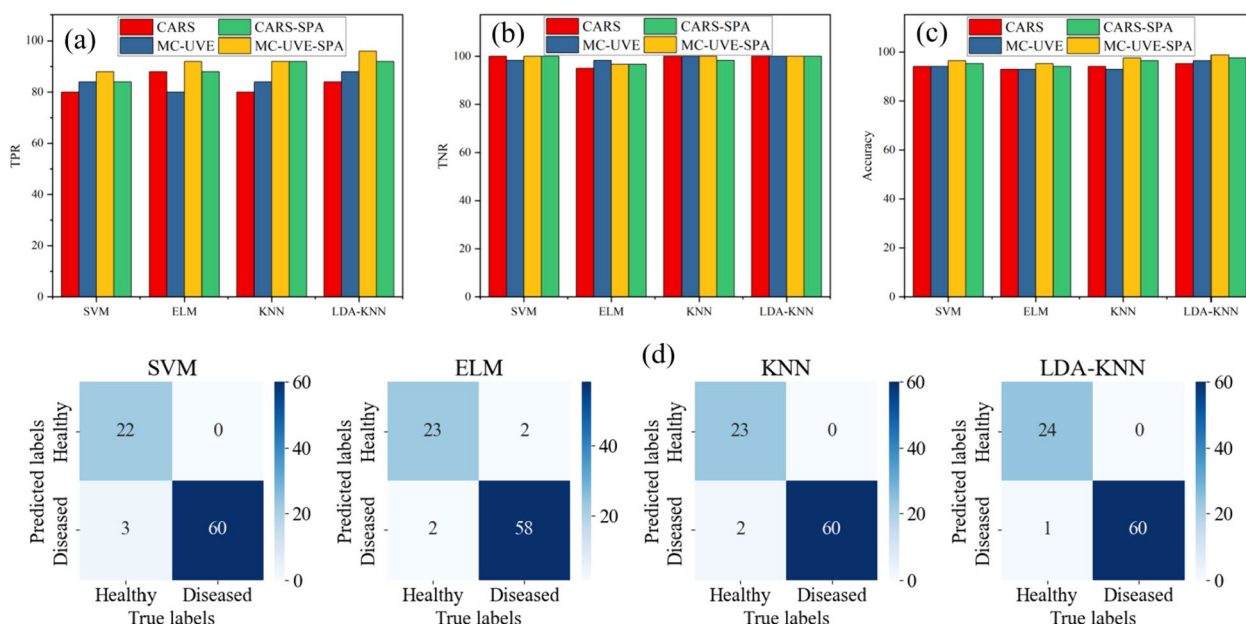
Jiang *et al. Chem. Biol. Technol. Agric.*   (2024) 11:63

Page 8 of 12

Fig. 5 Two-class classification results: TPR (**a**), TNR (**b**), accuracy (**c**), confusion matrix of prediction sets based on MC-UVE-SPA model (**d**)

sensitive to feature recognition of healthy and moldy core apples. Figure 5d shows the confusion matrix of the prediction model based on MC-UVE-SPA, from which it can be seen that LDA-KNN performs the best with only one misclassification followed by KNN and SVM, whereas ELM performs the worst with four misclassifications in which it misclassifies the two moldy core apples as healthy apples, which is extremely bad for both consumers and fruit farmers. The model based on MC-UE-SPA uses 18 of the 295 wavelengths extracted by MC-UVE, while the model based on the CARS-SPA algorithm uses 32 of the 119 wavelengths extracted by CARS. Without

affecting the accuracy of the model, the combined variable can be selected to select the combined algorithm to reduce irrelevant variables, improve the calculation speed, and reduce the complexity of the model. Therefore, the model based on MC-UVE-SPA is the most suitable choice, it has fewer variables and better performance than the model based on CARS-SPA.

Figure 6 shows the ROC curve of the model based on CARS-SPA and MC-UE-SPA. From the figure, LDA-KNN performs well, and the AUC value is close to 1, which is the highest among all classifiers. The performance of KNN and ELM is relatively stable, and both
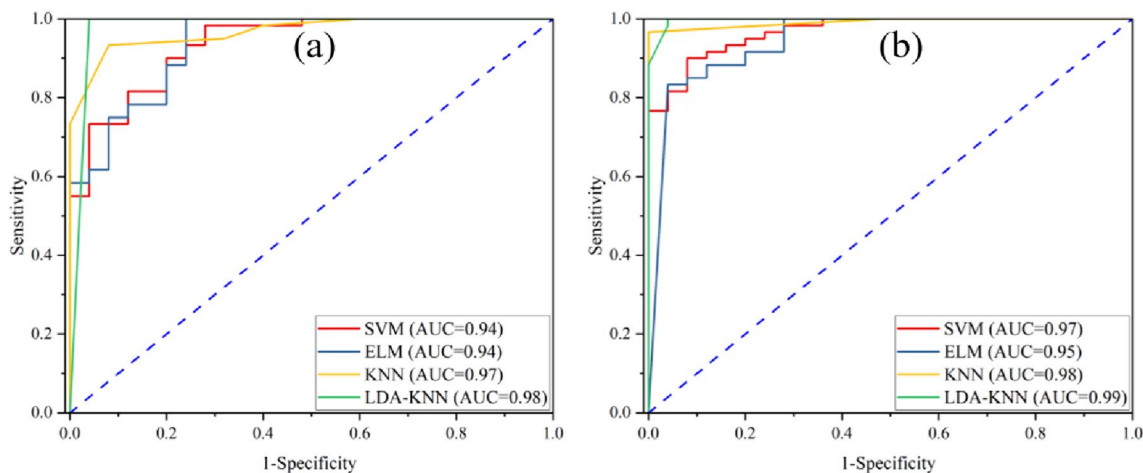

Fig. 6 Two-class classifications ROC curve of model prediction set: based on CARS-SPA (**a**), based on MC-UVE-SPA (**b**)

Jiang *et al. Chem. Biol. Technol. Agric.*   (2024) 11:63

Page 9 of 12

have achieved good results. The performance of SVM based on MC-UE-SPA is better than that of CARS-SPA, which is related to the variables they screen out. The results show that MC-UVE is a reliable variable selection algorithm, and the combination of MC-UVE and SPA can extract representative wavelengths for the classification of moldy core apples. Therefore, considering the TPR, TNR, Accuracy, and ROC curves of the model, the MC-UE-SPA-KNN and MC-UE-SPA-LDA-KNN models have excellent performance for the classification of healthy apples and moldy core apples. Thanks to its good data dimensionality reduction ability, the MC-UE-SPA-LDA-KNN model should be the best choice.

Three-class classification based on the effective wavelengths

Based on the four variable selection algorithms of CARS, CARS-SPA, MC-UVE, and MC-UVE-SPA, 154, 29, 214, and 14 effective wavelengths were screened out, respectively, and the number of effective wavelengths used in the three-class classification was found to be less than two-class classification through comparison. Figure 7 is the TPR, TNR, and Accuracy histogram of the three-level classification models of SVM, ELM, KNN, and LDA-KNN established for healthy apples, light, and severe apples using the determined effective wavelengths. From Fig. 7a, it can be seen that the TPR of all models performed well, and only MC-UE-SPA-KNN performed the worst. Combined with the confusion matrix in Fig. 7d, it can be found that there are a total of 7 misclassifications.

Among them, two light moldy core apples are misclassified as healthy apples, which is unacceptable to consumers. Figure 7b, c shows that the TNR and Accuracy performance of the model is relatively close and the overall performance is good. Figure 7d shows the confusion matrix of the prediction model based on MC-UE-SPA. From the figure, it can be seen that compared with the other three models, SVM performs best, with only two wrong classifications, followed by LDA-KNN, and ELM and KNN performance is just passable. It is worth noting that MC-UVE-SPA used only 14 of the 214 wavelengths extracted by SPA, and CARS-SPA used only 29 of the 154 wavelengths extracted by SPA. Thus, the MC-UVE-SPA and CARS-SPA wavelength selection algorithm significantly simplifies the model without affecting the classification accuracy. By comparing Figs. 5 and 7, it can be found that the overall effect of the three-class classification model is worse than the two-class classification model. Although the three-class classification model uses a smaller number of effective wavelengths, these wavelengths may not be sufficient to capture the complex relationships between their categories.

Figure 8 shows the ROC curve of the model based on CARS-SPA and MC-UE-SPA, from which it can be seen that the performance of MC-UVE-SPA-SVM is the best, and its AUC is close to 1, indicating that a good balance has been achieved between the TPR and the FPR, showing strong classification performance, next LDA-KNN. ELM's AUC is slightly lower but still relatively high, showing good performance. KNN's AUC is
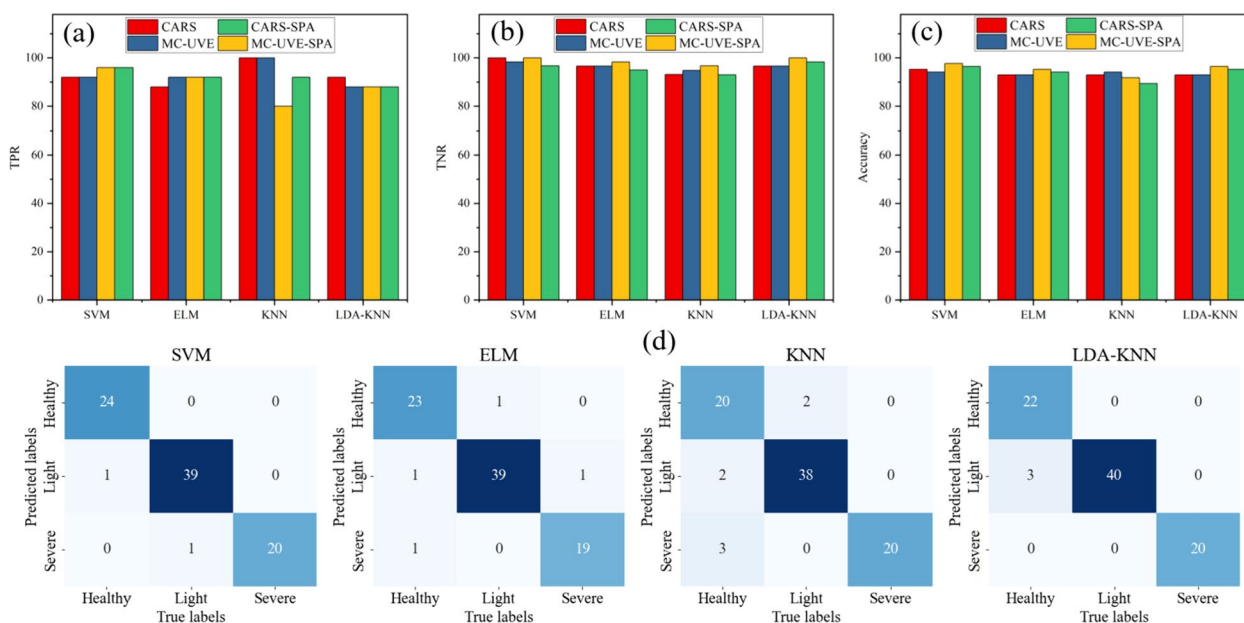

Fig. 7 Three-class classification results: TPR (**a**), TNR (**b**), accuracy (**c**), confusion matrix of prediction sets based on MC-UVE-SPA model (**d**)
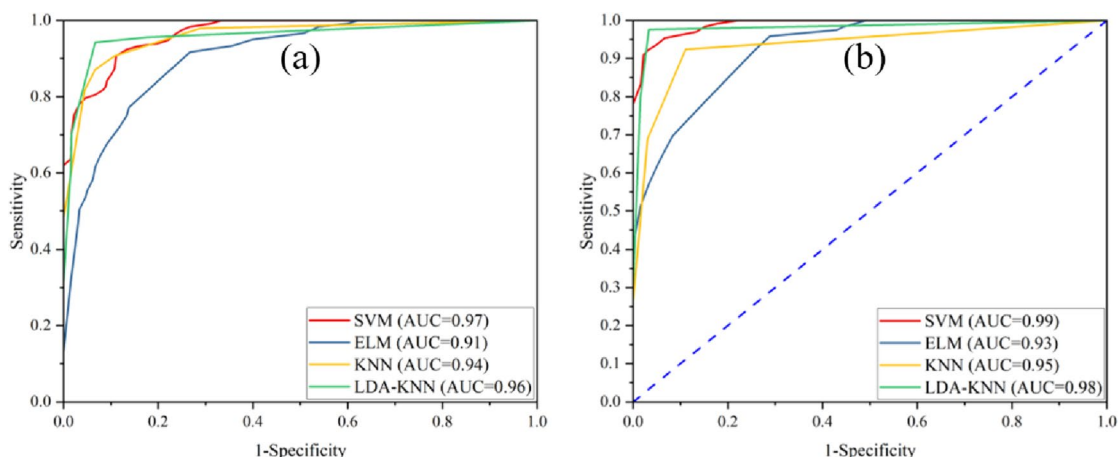
Fig. 8 Three-class classifications ROC curve of model prediction set: based on CARS-SPA (**a**), based on MC-UVE-SPA (**b**)

medium, indicating that it has a certain performance on this task, but it is slightly insufficient compared with SVM. Combining the performance of each model in Figs. 7 and 8, MC-UE-SPA-SVM is more suitable for the three-class classification problems of health, light, and severe.

By analyzing Figs. 7 and 8, we can find that the models based on CARS-SPA and MC-UVE-SPA perform well in all three classification tasks. In practical applications, the model's accuracy in detecting apples with moldy core should be higher than that for detecting healthy apples because misclassifying diseased apples as healthy can incur higher costs. In the detection of moldy core in apples, the most stringent requirement is to have zero tolerance for misclassifying severely moldy core apples as healthy. This is followed by the requirement to avoid misclassifying mildly moldy core apples as healthy. There is usually more tolerance for misclassifying healthy apples as diseased. In the study of early detection of apple moldy core, by comparing the performance of four different models, the MC-UVE-SPA-SVM model stood out with its excellent TPR, TNR, Accuracy, and ROC curves, proving its high accuracy and balance in identifying healthy apples and apples with different degrees of moldy core. In particular, the model successfully avoided misclassifying severe apples as healthy apples, demonstrating its reliability in food safety. In contrast, other models misclassified when distinguishing light from healthy apples, which is unacceptable from a food safety perspective. Therefore, MC-UVE-SPA-SVM not only proved to be the optimal model for early moldy core detection in apples but also highlighted the effectiveness of the MC-UVE-SPA variable selection algorithm, which made a significant contribution to improving the yield of dynamic online sorting and ensuring food safety.

Wavelength analysis used in the optimal models

In this study, the MC-UVE-SPA wavelength selection algorithm was utilized to select different sets of wavelengths for two-class and three-class classification tasks as shown in Fig. 9. For two-class classification, 18 wavelengths were chosen: 597, 616, 632, 645, 684, 690, 698, 704, 705, 708, 709, 712, 715, 716, 730, 760, 787, and 799 nm. Most of these wavelengths are within the visible spectrum, primarily because of the color difference between the healthy apple tissue and the moldy core tissue [33]. The moldy core tissue, being brown, is mainly composed of red, orange, and yellow, with 597 nm falling in the yellow band, 616 nm in the orange band, and the rest in the red band, which can be reflected in the transmission spectrum. For three-class classification, 14 wavelengths were selected: 639, 694, 707, 709, 714, 715, 716, 717, 720, 740, 746, 773, 810, and 820 nm. These
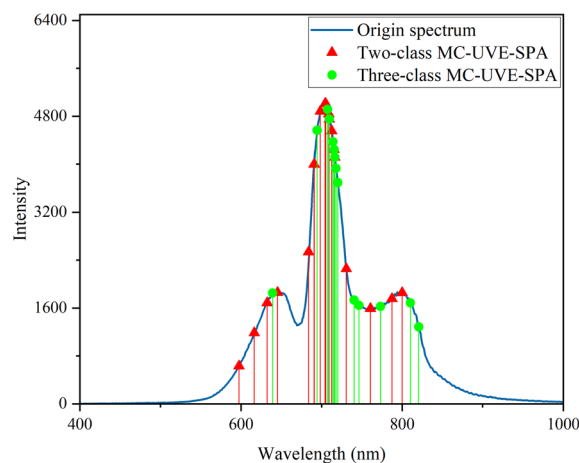

Fig. 9 Distribution of variables selected in two-class and three-class classifications based on MC-UVE-SPA

Jiang *et al. Chem. Biol. Technol. Agric.*   (2024) 11:63

Page 11 of 12

wavelengths overlap with some of the wavelengths used in binary classification, further highlighting their importance for distinguishing between healthy and moldy core tissues. Additionally, there is a small portion in the near-infrared spectrum, which contributes to differentiating various degrees of moldy core. This can be confirmed by Fig. 3a, where various degrees of the moldy core can be observed along with the decrease in absorption peak intensity as the condition of moldy core worsens.

Conclusions

This study successfully detected moldy core apples using VIS/NIR transmission spectroscopy. Discrimination models for two-class and three-class classifications were established by combining four wavelength selection algorithms (CARS, CARS-SPA, MC-UVE, and MC-UVE-SPA) with four classifiers (SVM, ELM, KNN, and LDA-KNN). The results show that MC-UVE-SPA-LDA-KNN is suitable for two-class classification between healthy and moldy core apples, with an AUC and accuracy of 0.99 and 98.82%, respectively. MC-UVE-SPA is applicable for three-class classifications of healthy, mild, and severe conditions, with an AUC and accuracy of 0.99 and 97.64%, respectively. This confirms that MC-UVE-SPA is an effective tool for selecting wavelengths specific to moldy core apples, aiding in the precise identification and differentiation of apple states. This study advances the dynamic online detection of early-stage moldy core conditions in apples, contributing to a reduction in postharvest disease occurrence and ensuring the effective preservation of fruit quality.

Author contributions
Xiao Gang Jiang: research, writing. Kang Ge and Zhi Liu: preparation, experiment. Nan Chen and Ai Guo Ou Yang: theoretical guidance. Yan de Liu and Yu yang Huang: grammar check. Jing hu Li and Ming mao Hu: format detection. All authors have read and agreed to the published version of the manuscript.

Availability of data and materials
The data that support the funding of this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]School of Mechatronics & Vehicle Engineering East, China Jiaotong University, Nanchang 330013, China. [2]Institute of Intelligent Mechanical and Electrical Equipment Innovation, Nanchang 330013, China. [3]Suzhou Suna Optoelectronics Company Limited, Suzhou 215000, China. [4]Xiamen Yixinyuan Semiconductor Technology Company Limited, Xiamen 361000, China. [5]School of Mechanical Engineering, Hubei University of Automotive Technology, Shiyan 442002, China.

References
1. Ali S, Abbasi P, Rehman S, Ellouze W. First report of moldy core of sweet tango apples from New Zealand caused by *Alternaria arborescens*. Plant Dis. 2021. https://doi.org/10.1094/PDIS-01-21-0025-PDN.
2. Li L, Peng Y, Li Y, Yang C, Chao K. Rapid and low-cost detection of moldy apple core based on an optical sensor system. Postharvest Biol Technol. 2020;168: 111276. https://doi.org/10.1016/j.postharvbio.2020.111276.
3. Shenderey C, Shmulevich I, Alchanatis V, Egozi H, Hoffman A, Ostrovsky V, Lurie S, Ben Arie R, Schmilovitch ZE. NIRS detection of moldy core in apples. Food Bioprocess Technol. 2009;3:79–86. https://doi.org/10.1007/s11947-009-0256-1.
4. Hu Q, Tian J, Fang Y. Detection of moldy cores in apples with near-infrared transmission spectroscopy based on wavelet and BP network. Int J Pattern Recognit Artif Intell. 2019;33:1950020. https://doi.org/10.1142/s0218001419500204.
5. Suchanek M, Kordulska M, Olejniczak Z, Figiel H, Turek K. Application of low-field MRI for quality assessment of 'conference' pears stored under controlled atmosphere conditions. Postharvest Biol Technol. 2017;124:100–6. https://doi.org/10.1016/j.postharvbio.2016.10.010.
6. Herremans E, Verboven P, Bongaers E, Estrade P, Verlinden B, Wevers MHE, Maarten LATM, Nicolai BM. Characterisation of 'Braeburn' browning disorder by means of X-ray micro-CT. Postharvest Biol Technol. 2013;75:114–24. https://doi.org/10.1016/j.postharvbio.2012.08.008.
7. Fathizadeh Z, Aboonajmi M, Beygi SRH. Nondestructive firmness prediction of apple fruit using acoustic vibration response. Sci Hortic. 2020;262: 109073. https://doi.org/10.1016/j.scienta.2019.109073.
8. Brezmes J, Llobet E, Vilanova X, Saiz G, Correig X. Fruit ripeness monitoring using an electronic nose. Sens Actuators B Chem. 2000;69:223–9. https://doi.org/10.1016/S0925-4005(00)00494-9.
9. Li D, et al. Detection and classification of citrus fruit infestation by *Bactrocera dorsalis* (Hendel) using a multi-path Vis/NIR spectroscopy system. Agriculture. 2023;13:1642. https://doi.org/10.3390/agriculture13081642.
10. Xia Y, Xu Y, Li J, Zhang C, Fan S. Recent advances in emerging techniques for non-destructive detection of seed viability: a review. Artif Intell Agric. 2019;1:35–47. https://doi.org/10.1016/j.aiia.2019.05.001.
11. Anderson NT, Walsh KB. Review: The evolution of chemometrics coupled with near infrared spectroscopy for fruit quality evaluation. J Near Infrared Spectrosc. 2022;30:3–17. https://doi.org/10.1177/09670335211057235.
12. Yao Y, Ma K, Zhu J, Huang F, Kuang L, Wang X, Li S. Non-destructive determination of soluble solids content in intact apples using a self-made portable NIR diffuse reflectance instrument. Infrared Phys Technol. 2023;132: 104714. https://doi.org/10.1016/j.infrared.2023.104714.
13. Ma T, Xia Y, Inagaki T, Tsuchikawa S. Rapid and nondestructive evaluation of soluble solids content (SSC) and firmness in apple using Vis–NIR spatially resolved spectroscopy. Postharvest Biol Technol. 2021;173: 111417. https://doi.org/10.1016/j.postharvbio.2020.111417.
14. Baranowski P, Mazurek W, Pastuszka WJ. Supervised classification of bruised apples with respect to the time after bruising on the basis of hyperspectral imaging data. Postharvest Biol Technol. 2013;86:249–58. https://doi.org/10.1016/j.postharvbio.2013.07.005.
15. Cruz S, Guerra R, Brazio A, Cavaco AM, Antunes D, Passos D. Nondestructive simultaneous prediction of internal browning disorder and quality attributes in 'Rocha' pear (*Pyrus communis* L.) using VIS–NIR spectroscopy. Postharvest Biol Technol. 2021;179: 111562. https://doi.org/10.1016/j.postharvbio.2021.111562.
16. Burdon J, Wohlers M, Pidakala P, Laurie T, Punter M, Billing D. The potential for commonly measured at-harvest fruit characteristics to predict chilling

susceptibility of 'Hort16A' kiwifruit. Postharvest Biol Technol. 2014;94:41–8. https://doi.org/10.1016/j.postharvbio.2014.03.005.

17. Fu X, Ying Y, Lu H, Xu H. Comparison of diffuse reflectance and transmission mode of visible–near infrared spectroscopy for detecting brown heart of pear. J Food Eng. 2007;83:317–23. https://doi.org/10.1016/j.jfoodeng.2007.02.041.

18. Tian S, Wang S, Xu H. Early detection of freezing damage in oranges by online Vis/NIR transmission coupled with diameter correction method and deep 1D-CNN. Comput Electron Agric. 2022;193: 106638. https://doi.org/10.1016/j.compag.2021.106638.

19. Sun X, Liu Y, Li Y, Wu M, Zhu D. Simultaneous measurement of brown core and soluble solids content in pear by on-line visible and near infrared spectroscopy. Postharvest Biol Technol. 2016;116:80–7. https://doi.org/10.1016/j.postharvbio.2016.01.009.

20. Mishra P, Woltering E, Brouwer B, Hogeveen-van EE. Improving moisture and soluble solids content prediction in pear fruit using near-infrared spectroscopy with variable selection and model updating approach. Postharvest Biol Technol. 2021;171: 111348. https://doi.org/10.1016/j.postharvbio.2020.111348.

21. Bro R, Smilde AK. Principal component analysis. Anal Methods. 2014;6:2812–31. https://doi.org/10.1039/c3ay41907j.

22. Li J, Huang W, Chen L, Fan S, Zhang B, Guo Z, Zhao C. Variable selection in visible and near-infrared spectral analysis for noninvasive determination of soluble solids content of 'Ya' pear. Food Anal Methods. 2014;7:1891–902. https://doi.org/10.1007/s12161-014-9832-8.

23. Li H, Liang Y, Xu Q, Cao D. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. Anal Chim Acta. 2009;648:77–84. https://doi.org/10.1016/j.aca.2009.06.046.

24. Mário CUA, Teresa CBS, Roberto KHG. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemom Intell Lab Syst. 2001;57:65–73. https://doi.org/10.1016/S0169-7439(01)00119-8.

25. Jiang X, Zhu M, Yao J, Zhang Y, Liu Y. Calibration of near infrared spectroscopy of apples with different fruit sizes to improve soluble solids content model performance. Foods. 2022;11:1923. https://doi.org/10.3390/foods11131923.

26. Kadir S. Different apple varieties classification using kNN and MLP algorithms. Int J Intell Syst Appl Eng. 2016;4:166–9. https://doi.org/10.18201/ijisae.270416.

27. Georgouli K, Martinez DRJ, Koidis A. Continuous statistical modelling for rapid detection of adulteration of extra virgin olive oil using mid infrared and Raman spectroscopic data. Food Chem. 2017;217:735–42. https://doi.org/10.1016/j.foodchem.2016.09.011.

28. Huang G, QinYu Z, CheeKheong S. Extreme learning machine a new learning scheme of feedforward neural networks. IEEE Xplore. 2004;2:985–90. https://doi.org/10.1109/IJCNN.2004.1380068.

29. Vapnik VN. The nature of statistical learning theory. New York: Springer; 2000.

30. Wei X, He J, Zheng S, Ye D. Modeling for SSC and firmness detection of persimmon based on NIR hyperspectral imaging by sample partitioning and variables selection. Infrared Phys Technol. 2020;105: 103099. https://doi.org/10.1016/j.infrared.2019.103099.

31. Merzlyak MN, Solovchenko AE, Gitelson AA. Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit. Postharvest Biol Technol. 2003;27:197–211. https://doi.org/10.1016/s0925-5214(02)00066-2.

32. Yuan L, Cai J, Sun L, Han E, Ernest T. Nondestructive measurement of soluble solids content in apples by a portable fruit analyzer. Food Anal Methods. 2015;9:785–94. https://doi.org/10.1007/s12161-015-0251-2.

33. Clark CJ, McGlone VA, Jordan RB. Detection of brownheart in 'Braeburn' apple by transmission NIR spectroscopy. Postharvest Biol Technol. 2003;28:87–96. https://doi.org/10.1016/s0925-5214(02)00122-9.

Publisher's Note