Article

Research on Defect Detection in Kubo Peach Based on Hyperspectral Imaging Technology Combined with CARS-MIV-GA-SVM Method

Lixiu Zhang [1,2], Pengcheng Nie [1,*], Shujuan Zhang [2,*], Liying Zhang [3] and Tianyuan Sun [2]

[1] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China; z18503483861@163.com

[2] College of Agricultural Engineering, Shanxi Agricultural University, Jinzhong 030800, China; nicosty@163.com

[3] State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Beijing Normal University, Beijing 100875, China; zlyzhangliying@mail.bnu.edu.cn

* Correspondence: pcn@zju.edu.cn (P.N.); zsujuan1@163.com (S.Z.);Tel.: +86-0571-8898-2456 (P.N.); +86-13935491091 (S.Z.)

Abstract: Due to the dark red surface of ripe fresh peaches, their internal injury defects cannot be detected using the naked eye and conventional images. The rapid and accurate detection of fresh peach defects can improve the efficiency of fresh peach classification. The goal of this paper was to develop a nondestructive approach to simultaneously detecting internal injury defects and external injuries in fresh peaches. First, we collected spectral data from 347 Kubo peach samples using hyperspectral imaging technology (900–1700 nm) and carried out pretreatment. Four methods (the competitive adaptive reweighting algorithm (CARS), the combination of CARS and the average influence value algorithm (CARS-MIV), the combination of CARS and the successive projections algorithm (CARS-SPA), and the combination of CARS and uninformative variable elimination (CARS-UVE)) were used to extract the characteristic wavelength. Based on the characteristic wavelength extracted using the above methods, a genetic algorithm optimization support vector machine (GA-SVM) model and a least-squares support vector machine (LS-SVM) model were used to establish classification models. The results show that the combination of CARS and other feature wavelength extraction methods can effectively improve the prediction accuracy of the model when the number of wavelengths is small. Among them, the discriminant accuracy of the CARS-MIV-GA-SVM model reaches 93.15%. In summary, hyperspectral imaging technology can accomplish the accurate detection of Kubo peaches defects, and provides feasible ideas for the automatic classification of Kubo peaches.

Keywords: Kubo peach; defect; CARS-MIV; GA-SVM; nondestructive testing

1. Introduction

"Early maturity Okubo Peach", referred to as the "Kubo peach", is a kind of premature variety of peach that originated in Japan, in Okubo; this fruit type has a fruit weight of about 230–280 g, dense meat, less fiber, and more juice; contains 16.48% soluble solid (16.48%); and is high-quality, rich in nutrition, and popular with consumers [1]. In the growing and harvesting process of Kubo peaches, due to climate and unavoidable collision in the harvesting process, the Kubo peach is prone to defects such as dark wounds, lesions, rot, scars, and marks [2,3]. The existence of these defects on the fruit surface will reduce the quality and market value of the fruit. When purchasing fruits, consumers tend to choose fruits with a good appearance. However, for obvious external obvious defects such as scab, rot, and others, peaches can be detected using conventional visual technology. While mature peaches are mostly red, the naked eye cannot identify damaged peaches, and ordinary machine vision cannot identify them, which leads to the classification and sorting efficiency of peaches not being high, affecting the commodity value of the fruit, and then,

affecting the export and sales of Kubo peaches. At present, the peach fruit on the market mainly relies on manual grading. Manual classification has the problems of low efficiency and low accuracy [4]. Therefore, it is of practical value to study a fast and efficient method for the detection of dark wounds and obvious external defects in Kubo peaches.

Hyperspectral imaging is a non-contact image acquisition technology used to obtain continuous spectral information for objects in different wavelength ranges. It combines optical and digital image processing technology to provide rich spectral data and spatial information, making the detection and analysis of the surface composition, chemical composition, and specific characteristics of the object more accurate and comprehensive. Hyperspectral imaging combined with chemometric approaches is proven to be a powerful tool for the quality evaluation and control of fruits [5,6], as it enables the assessment of internal properties that cannot be inspected using computer vision, including soluble solid content [7], acidity [8], and texture [9]. Therefore, to detect fruit quality defects, such as immaturity [10], bruising [11], etc., it is feasible to use hyperspectral imaging technology to detect the dark wound defects and external defects of peaches.

In recent years, hyperspectral imaging technology has been widely applied to the external quality detection of fruits and vegetables, and the research subjects mainly include dates, cucumbers, cherries, citrus fruits, apples, peaches [12–22], etc. Most scholars have combined hyperspectral imaging technology with the related stoichiometry, and have obtained relatively scientific research results. In terms of internal quality detection, Li et al. [13] combined hyperspectral imaging technology with a multiple linear regression model to predict the soluble solid content of Hami jujubes, and the correlation coefficient of the prediction set reached 0.857. Li. et al. [14] combined hyperspectral technology with the CARS method to detect cucumber hardness and water loss, two indexes representing cucumber freshness. The final correlation coefficient of the PLSR model for hardness was 0.942, and that of the PLSR model for moisture was 0.822. Pullanagari et al. [15] combined hyperspectral imaging technology with a partial least-squares regression (PLSR) model and a Gaussian process regression (GPR) model to detect the hardness and total soluble solids of sweet cherries. Finally, the correlation coefficient of total soluble solids predicted using the GPR model was 0.88. The correlation coefficient of hardness prediction was 0.60. He et al. [16] combined hyperspectral imaging technology with multiplicative scattering correction, Savitzky–Golay, and first-order derivative pretreatment methods to establish a PLSR model to predict the water content of dried purple potatoes, and the correlation coefficient of the final model reached 0.975. In terms of defect detection, Tang et al. [17] fitted the hyperspectral imaging data of apples with different damage levels using piecewise nonlinear curves, studied the spectral data of damaged apples within the band range of 386–1016 nm, and concluded that the final score detection accuracy of this method reached 97.33%. Xu et al. [18] combined hyperspectral imaging technology with a partial least-squares regression model to detect the relationship between the damage degree and internal attribute quality of mangos, and graded mangos according to the damage degree. The final classification accuracy was 77.8%. WANG et al. [19] selected different pretreatment methods to establish an LS-SVM model based on the spectral data of the upper and lower surfaces of citrus leaves with yellow dragon disease. The results showed that the recognition rates of the upper and lower surfaces of citrus leaves were 100% and 92.5%, respectively, when the second derivative was selected as the pretreatment method. Zhang et al. [20] used mean-PC image, an improved watershed segmentation algorithm, and hyperspectral technology to distinguish normal oranges from defective oranges, and the overall classification accuracy reached 97.73%. Based on mean-PC5 and the simple global threshold method, rotten oranges and intact oranges were identified, and the recognition rate reached 100%. Zhang H. et al. [21] used hyperspectral imaging technology to detect scab, black spot, root rot, and brown disease in citrus, with a 94% final discrimination rate. Chen Si et al. [22] used hyperspectral imaging technology and the band selection method to propose defect region segmentation and recognition algorithms for peach brown rot and scab, with final recognition efficiencies of 96.9% and 88.4%, respectively.

These studies used hyperspectral imaging technology to model and analyze the internal qualities and external defects of fruits; they achieved good results with research methods primarily based on one or more single-feature wavelength extraction methods, which have wavelength redundancy problems or wavelengths that are too small. The wavelength combination method can maximize useful information from samples and improve detection accuracy. There are few reports regarding using the CARS-MIV combination method to extract characteristic wavelengths and establish GA-SVM and LS-SVM classification models to detect peach dark wound defects.

Based on the above problems, this study aimed to use hyperspectral imaging technology combined with stoichiometric methods to achieve rapid, efficient, nondestructive batch identification of the Kubo peach. This study's specific process included the following:

(1) Collecting Kubo peach fruit with different defect types and obtaining hyperspectral image data between 900 nm and 1700 nm;

(2) Selecting the appropriate pretreatment method by establishing a partial least squares (PLS) model;

(3) Introducing the CARS-MIV group method to extract the feature wavelength and compare the results with other classical feature wavelength extraction methods;

(4) Introducing the GA-SVM discriminant model and comparing it with the classical LS-SVM model to select the best prediction model.

In this study, HSI imaging technology provided a new research method for the rapid, efficient batch identification of Kubo peach external defects and the accurate classification of Kubo peach fruit.

2. Materials and Methods

2.1. Sample Collection

The sample consisted of Kubo peaches harvested on 12 July 2022, from Xi shan di Village, Jin zhong City, Shanxi Province, China. To ensure accuracy, the principles of uniform size (single fruit weight 170–180 g) and complete defect types (intact peaches, scab peaches, and rotten peaches) were selected during harvest. Damaged Kubo peach samples were placed in a paper box to simulate collision during harvest and transportation. After selection and analysis, 347 Kubo peaches were chosen for spectral data analysis. Selected Kubo peaches were left at room temperature for 24 h, and the three sample types were randomly divided into 259 correction sets and 88 prediction sets according to the Kennard–Stone algorithm ratio of 3:1. Figure 1 shows sample images of four Kubo peach types.
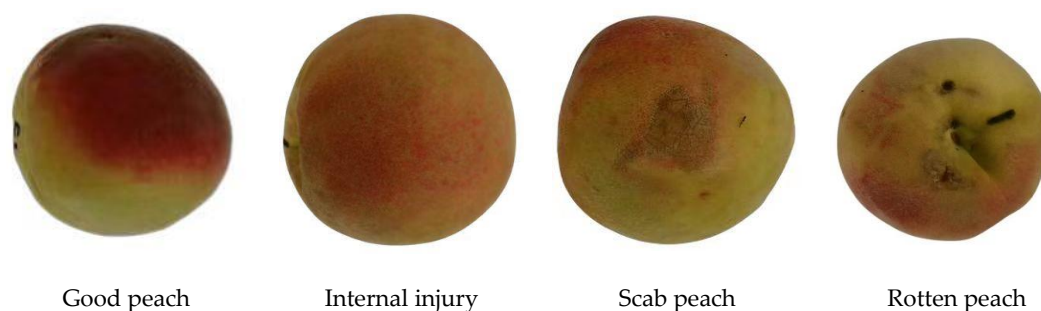

Good peach   Internal injury   Scab peach   Rotten peach

Figure 1. Sample images of an intact peach and defective peaches.

2.2. Hyperspectral Imaging System

Hyperspectral imaging technology combines spectral information with spatial information [23]. It has hyperspectral and spatial resolution, and higher resolution and accuracy for the surface components, component distribution, and feature extraction of objects [24]. The hyperspectral sorting instrument adopted in this study is shown in Figure 2, and the specific parameters of the hyperspectral sorting instrument are shown in Table 1.

Figure 2. Hyperspectral sorter.

Table 1. Hyperspectral sorter's parameters.

Name	Parameter
Model number	ZOLIX Gaia Sorter-type "Gaia" hyperspectral sorter
Main instrument parts	Image-lambda-n17e spectral camera, camera obscura, electronic control platform, bromo-tungsten lamp (4), external computer
Sample exposure time	20 ms
Lens distance from sample height	22 cm
Forward speed of the electronically controlled mobile platform	2 cm/s

As the spectrum was affected by changes in light intensity and the dark current in the lens [25] during the acquisition process, the acquired images were not clear, so it was necessary to perform black-and-white correction before spectrum acquisition [26].

The following calculation formula was used for black-and-white correction.

$$R = \frac{IR - ID}{IW - ID} \tag{1}$$

Note: R—corrected image; IR—original image; ID—blackboard correction image; IW—whiteboard correction image.

2.3. Main Data Processing Methods

2.3.1. CARS-MIV Feature Variable Extraction Method

The competitive adaptive reweighting algorithm (CARS) is an unstable algorithm; the regression coefficient R of the variable set changes as the sample numbers change [27]. Therefore, the absolute value of the regression coefficient R cannot effectively reflect a variable's importance, and the feature variable extracted by a single CARS algorithm will occupy approximately 1/3 of the spectral data's full wavelength. Modeling data are not concise enough. Therefore, the average influence value algorithm (MIV) [28] was introduced in this study to conduct secondary screening of characteristic wavelengths screened using the CARS algorithm, further simplifying the data and improving the model's accuracy. The average impact value algorithm (MIV) was used for feature selection. It evaluated the features' importance by calculating their average influence value on the predicted results, and then, carried out feature selection. The MIV algorithm process was as follows (Figure 3):
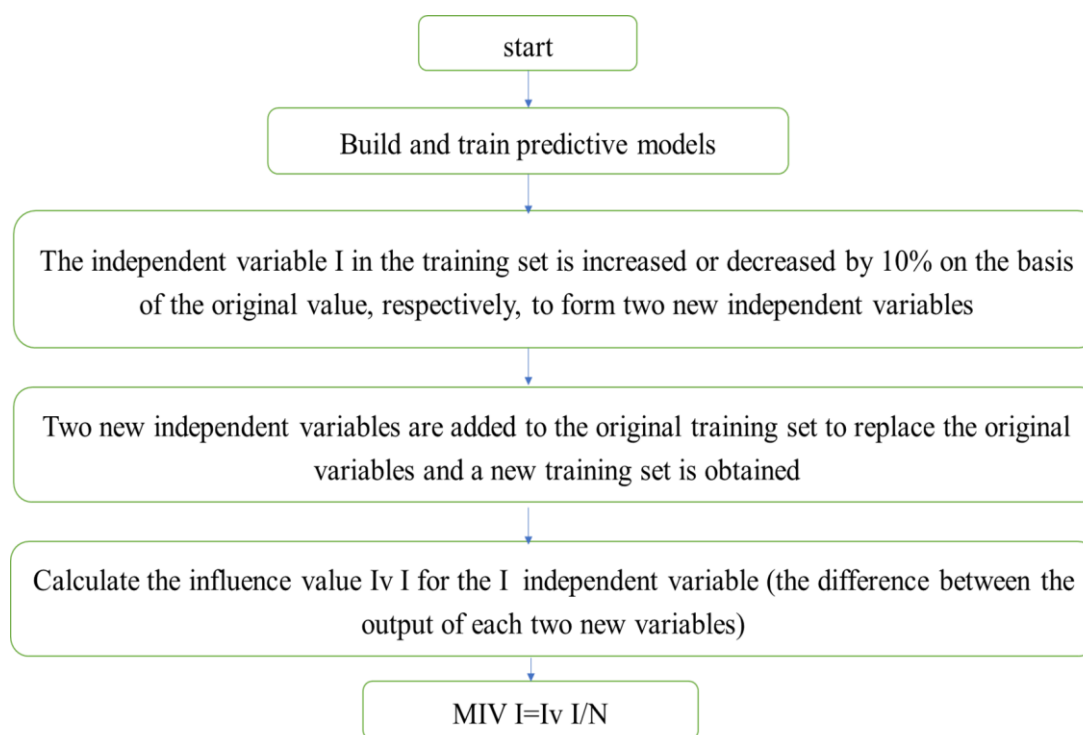
```
                          ┌─────────────┐
                          │    start    │
                          └─────────────┘
                                 │
                    ┌────────────────────────┐
                    │ Build and train        │
                    │ predictive models      │
                    └────────────────────────┘
                                 │
        ┌──────────────────────────────────────────────────┐
        │ The independent variable I in the training set is │
        │ increased or decreased by 10% on the basis        │
        │ of the original value, respectively, to form two  │
        │ new independent variables                         │
        └──────────────────────────────────────────────────┘
                                 │
        ┌──────────────────────────────────────────────────┐
        │ Two new independent variables are added to the    │
        │ original training set to replace the original     │
        │ variables and a new training set is obtained      │
        └──────────────────────────────────────────────────┘
                                 │
        ┌──────────────────────────────────────────────────┐
        │ Calculate the influence value Iv I for the I      │
        │ independent variable (the difference between the  │
        │ output of each two new variables)                 │
        └──────────────────────────────────────────────────┘
                                 │
                    ┌────────────────────────┐
                    │     MIV I=Iv I/N        │
                    └────────────────────────┘
```

Figure 3. MIV algorithm flowchart.

Step 1: Use the original sample set to establish the Shenjing network training model Svmtrain.

Step 2: Increase and decrease the original P1 and P2 by 10%, respectively, to form new training set samples, P1 and P2.

Step 3: Use the new training set samples, P1 and P2, as the neural network's input model, and obtain two output results, y1 and y2.

Step 4: The difference between the two output results (y1–y2) is the impact value (IV) on the output variable after changing the original input variable.

Step 5: Average the IV value according to the number of observed cases to obtain the MIV of the input to the output variable. A loop structure is used to successively calculate each input variable's MIV, and the characteristic wavelength variables whose contribution rates are greater than 95% are screened out according to the descending order of the absolute value.

2.3.2. GA-SVM Modeling Method

A support vector machine (SVM) is a powerful classification and regression model; however, it also has some disadvantages. An SVM has relatively low processing efficiency of large-scale data sets and high-dimensional data, presents difficulty when selecting the appropriate kernel function for nonlinear problems, requires selection of the appropriate penalty parameter C, etc. As a classical optimization algorithm, a genetic algorithm has good applicability, a fast search speed, and high efficiency [29]. In this study, a genetic algorithm (GA) was used to optimize the C and r parameters of the support vector machine (SVM). The steps based on the GA-SVM [30] algorithm are as follows; the process is shown in Figure 4.

Step 1: Set the parameters and reasonably select the genetic algorithm's operating parameters and the SVM's parameter optimization range.

Step 2: Form the initial population, convert the kernel function parameter r and SVM penalty factor C into genotype data with genetic characteristics via binary coding; randomly generate the initial population.

Step 3: Calculate each individual's fitness to evaluate their importance.

Step 4: Use selection, crossover, and mutation operations in the genetic algorithm to optimize parameters c and g to determine whether each individual's fitness meets the termination criteria. If yes, perform Step 6; otherwise, perform Step 5.

Step 5: Select individuals with high fitness for crossover and mutation operations to form a new population. Return to Step 3 until the stop criteria are met, ending the loop.

Step 6: Parse the code and output the parameter value.

Step 7: Apply optimal parameters to the SVM model for solution analysis.
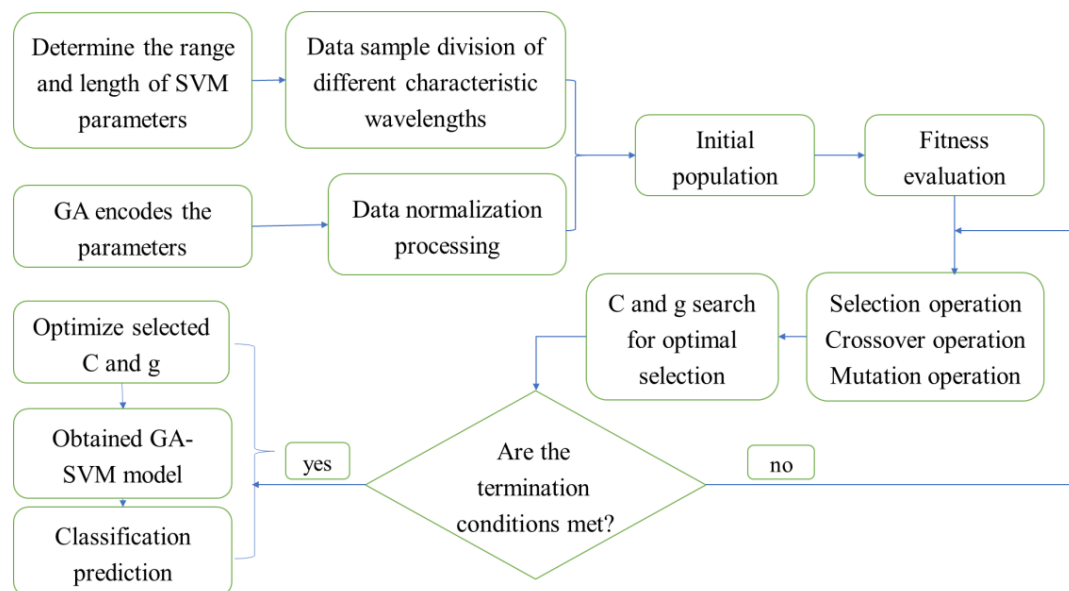
Figure 4. GA-SVM algorithm flowchart.

2.4. Model Evaluation Index

In this study, five indexes (namely, the coefficient of determination of the training set (RC2), the root-mean-square error (RMSEC), the coefficient of determination of the test set (RP2), the root-mean-square error of the training set (RMSEP) and the discrimination accuracy of the prediction set) were used to comprehensively evaluate the model. The closer the correlation coefficient between the correction set and the prediction set is to 1, the closer the root-mean-square error is to 0, and the higher the discrimination accuracy is, the better the model is considered to be.

3. Results and Discussion

3.1. Average Spectral Curves of Four Types of Kubo Peach Samples

The ROI region selection should best represent image content features; this can help reduce the calculation complexity, improve the algorithm's performance, and provide increased accuracy [31]. In this study, the ROI region in the center of the equator of intact peach samples and ROI regions in the defects of damaged peaches, scab peaches, and scarred peaches were selected. After sorting the spectral data from the four sample types, average spectral curves for each of the four sample types were drawn using Origin8.5 software, as shown in Figure 5.

As can be seen in Figure 5, there were significant differences in the average spectral curves of the four Kubo peach sample types. This was due to differences in the skin thickness and spectral reflectance of intact and defective peaches. Wave peaks at 1100 nm and 1290 nm of the four sample types were related to the Kubo peach's excessive reflectivity, and troughs at 1190 nm and 1450 nm were related to the Kubo peach's internal water and sugar absorption, which are O-H third-order and second-order frequency-doubling characteristic absorption peaks [32]. The four sample types' overall reflectance showed a decreasing trend.
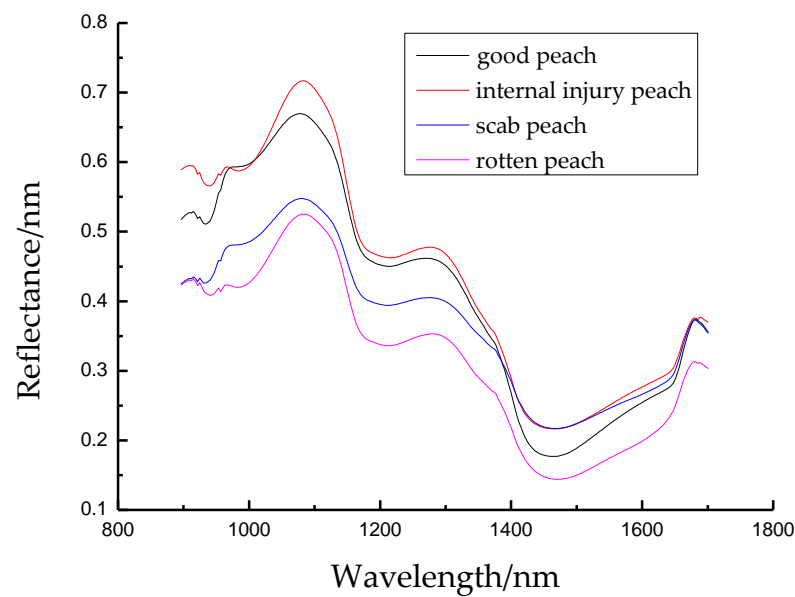
Figure 5. Average spectral curves of four sample types.

3.2. Spectrum Pretreatment

Spectral pre-processing refers to a series of processing operations on acquired spectral data to extract useful information or improve data quality [33]. In this study, derivative-gap-segment (D-GS), standard normalization (SNV), baseline, and a smoothing-median filter (SMF) were used. Derivative Savitzky–Golay (D-SG) and five pretreatment methods were used to process the original spectral data, and PLS models were established. The prediction results are shown in Table 2.

Table 2. PLS models established using different pretreatment methods.

Pretreatment Method	Correction Set		Prediction Set	
	RC2	RMSEC	RP2	RMSEP
Primary spectrum	0.72	0.60	0.86	0.45
DGS	0.74	0.57	0.83	0.47
DSG	0.75	0.56	0.85	0.45
Baseline	0.71	0.61	0.82	0.51
SMF	0.72	0.60	0.85	0.45
SNV	0.65	0.67	0.72	0.62

As shown in Table 1, the model pre-treated using DSG had relatively high accuracy and a relatively low standard deviation. Its correction set determination coefficient was 0.75 with a standard deviation was 0.56, and its prediction set determination coefficient was 0.85 with a standard deviation was 0.45. Therefore, data pre-processed using DSG were selected for the follow-up study.

3.3. Selection of Characteristic Wavelength Variables

3.3.1. CARS Competitive Adaptive Weighting Algorithm

The CARS competitive adaptive reweighting algorithm is a sampling algorithm based on adaptive weights, and is often used to solve data imbalance problems. In data imbalance problems, the training set has a large difference in the number of samples from different classes, which may cause the classifier to tend to predict classes that occur more frequently, while the prediction effect is poor for rare classes. The CARS algorithm aims to solve this problem. By adaptively adjusting sample weights and iteratively updating sample weights and model parameters, the CARS algorithm can achieve a state of dynamic balance and

improve the prediction accuracy of rare categories [34]. The CARS feature wavelength extraction process is illustrated in Figure 6.
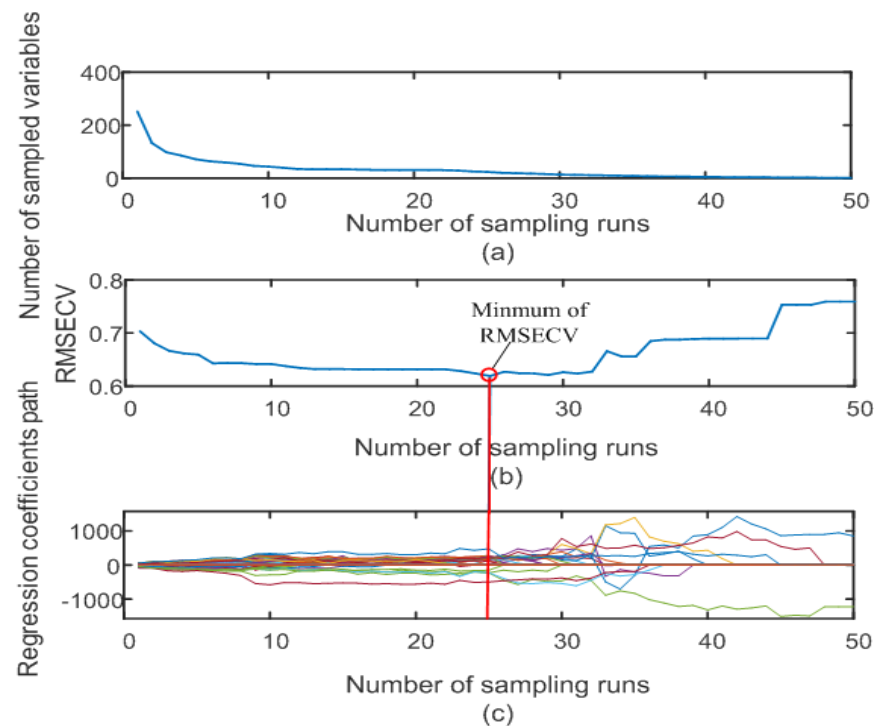

Figure 6. CARS feature wavelength extraction process. Variation of (**a**) the number of variables in the attenuation function, (**b**) the root-mean-square error value of cross-validation (**c**) the path value of the regression coefficient with the increase of the sampling number.

Pre-processed spectral characteristic variables were extracted based on the CARS algorithm, and the number of Monte Carlo samplings was set to 50. Figure 6a shows the attenuation function selection process; the number of variables showed a trend of rapid decline at first, and then, a steady decline with an increase in sampling times. Figure 6b shows that the root-mean-square error of cross-validation first slowly decreased, and then, rapidly increased with an increase in sampling times. Figure 6b shows that the root-mean-square error of cross-validation (RMSE) was smallest at the 25th sampling time, and its minimum value was 0.6190. Figure 6c shows the change process of the regression coefficient's path value with an increase in sampling times. A total of 24 characteristic wavelengths were screened out by running the CARS algorithm 25 times.

3.3.2. Feature Wavelength Extraction Using the CARS-MIV Method

The average influence value (MIV) algorithm can screen out variables with high correlation for modeling analysis. In this study, 24 feature variables were screened using the CARS algorithm, and feature variables with contribution rates greater than 95% were screened using the MIV algorithm to achieve a secondary screening, so that the feature wavelength variables were more representative. Selected wavelengths were sorted in descending order, and 11 characteristic wavelength variables were finally selected for subsequent modeling analysis. The MIV screening process is shown in Figure 7, the selected characteristic wavelength is shown in Figure 8, and the cumulative contribution rate of the characteristic wavelength is shown in Figure 9.
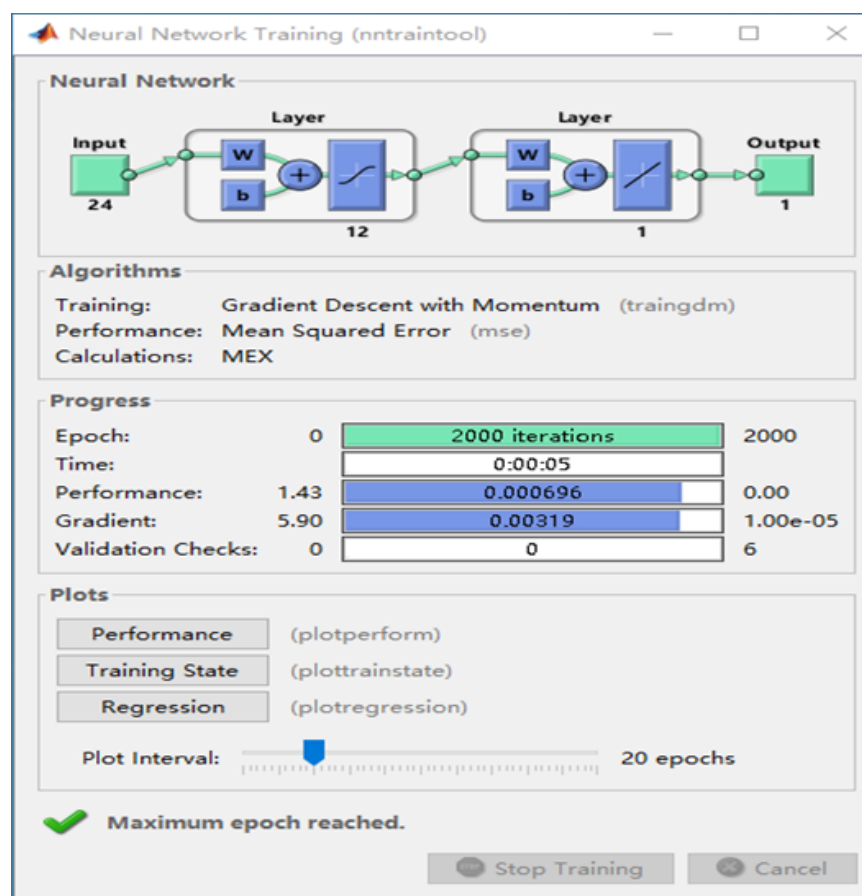
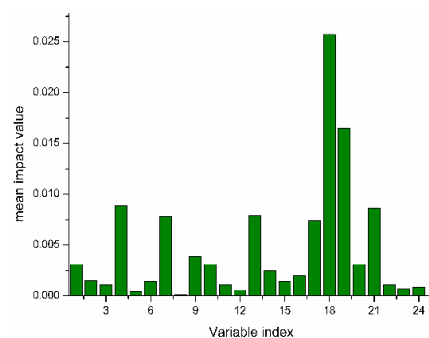**Figure 7.** Wavelength extraction process of MIV secondary screening.

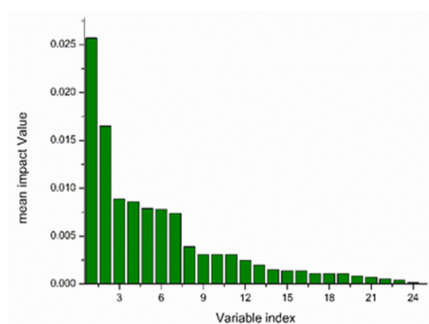**Figure 8.** CARS-MIV feature variables' values.

Figure 9. Importance ranking of CARS-MIV feature variables.

3.3.3. Feature Wavelength Extraction Using the CARS-SPA Method

There was a data imbalance problem among the 24 characteristic-wavelength-adjacent variables formed using the CARS algorithm. Therefore, the SPA method of screening characteristic wavelengths was used to reduce the dimension twice. Data imbalances in CARS extraction variables could be eliminated. The variation range of the number of characteristic spectrum variables was set to 1–25. The SPA algorithm reduced the dimensions of 24 characteristic spectra, and seven characteristic variables were extracted when the RMSEC was 0.62033. The characteristic wavelength variables are shown in Figure 10.
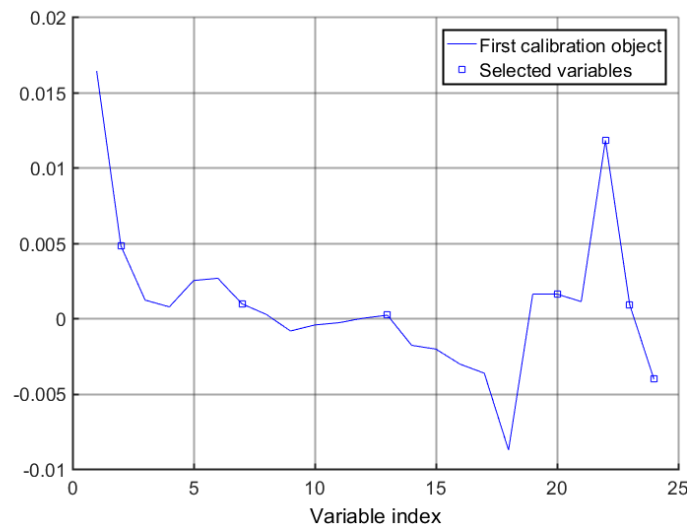

Figure 10. Distribution of CARS-SPA characteristic variables.

3.3.4. Feature Wavelength Extraction Using the CARS-UVE Method

The uninformative variable elimination (UVE) algorithm is a band selection algorithm based on the regression coefficient of the PLS model. The value of the regression coefficient is used to measure whether or not a variable is valid. A new matrix can be obtained by randomly setting a noise matrix and placing it after the spectral matrix. The PLS regression process is performed on the new matrix using the leave-one-interaction-out verification method. The reliability of each variable can be obtained by comparing the PLS results with the standard deviation and average value of the regression coefficients. Then, according to the reliability value, feature wavelengths with large reliability values are screened out [35]. The UVE algorithm can delete uninformative variables in the spectral information, but when the number of bands is large, the number of feature wavelengths selected by the algorithm remains large, which is not conducive to the development of spectral equipment. Therefore, the UVE algorithm is often combined with other algorithms to further screen feature variables [36]. In this study, the CARS algorithm was combined with the UVE algorithm. On the basis of 24 feature variables screened using the CARS algorithm, the UVE algorithm was used to screen 15 feature wavelengths twice. The filtering process is shown in Figure 11.

The dotted blue horizontal lines in Figure 11 are the upper and lower limits of the thresholds, and the variables between these lines are information unrelated to the Kubo peach classification prediction, which require deletion. The 15 variables outside the dotted line are the selected feature variables.

3.4. Modeling Results and Analysis

3.4.1. Least-Squares Support Vector Machine (LS-SVM) Model

The principle of the LS-SVM modeling method is to take RBF as the model's kernel calculation function, use a grid-search method on the basis of cross-validation, and adopt a global optimization method for the main SVM target parameters γ and δ^2. The princi-

ple [37] that minimizes the root-mean-square error RMSEC was used to optimize the target parameters' design.
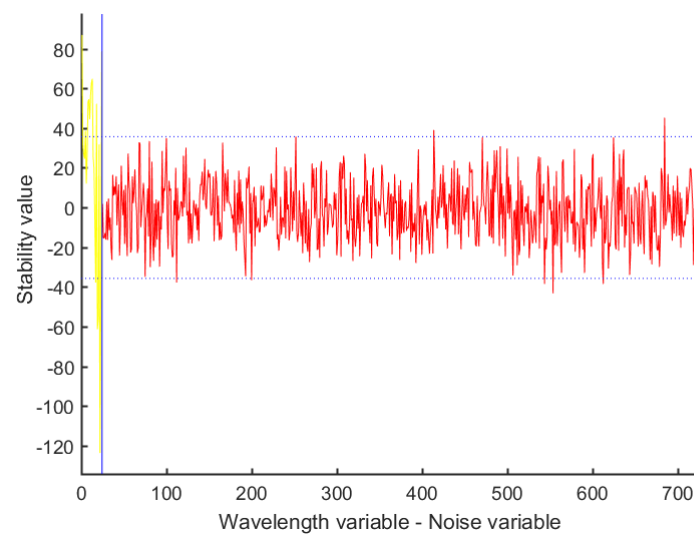

Figure 11. Feature wavelength extracted using the CARS-UVE method.

3.4.2. Support Vector Machine (GA-SVM) Parameter Optimization Model Based on the Genetic Algorithm

The genetic algorithm is an optimization algorithm based on the survival of the fittest and the natural selection mechanism in biological evolution. It has the characteristics of strong fitness, high optimization efficiency, and fast searching ability. In the GA-SVM model used in this study, the maximum number of genetic iterations was set to 100, the population size was set to 20, and the method of five-fold cross-validation was adopted to obtain the optimal penalty factor (cost) as C and core parameter (gamma) as γ. In this study, the feature wavelength sample data extracted using the four methods were normalized. After the initial training was complete, training sample data were input into the GA-SVM model for model training, and then, the prediction set sample was input into the trained model to obtain the training results. By comparing the predicted training results with the real predicted results, confusion matrices of the discriminant samples were obtained.

The prediction data results of the two models established using the four feature wavelength extraction methods are shown in Tables 3–6.

Table 3. Confusion matrices of CARS-GA-SVM and CARS-LS-SVM models.

GA-SVM				Model	LS-SVM			
Good Peach	Internal Injury Peach	Scab Peach	Rotten Peach	Classification	Good Peach	Internal Injury Peach	Scab Peach	Rotten Peach
49	0	0	0	Good peach	49	0	0	0
0	7	1	1	Internal injury peach	0	4	5	0
0	0	16	1	Scab peach	0	6	11	0
0	0	2	11	Rotten peach	0	0	3	10
49	9	17	13	Total	49	9	17	13
100	77.7	94.12	84.62	Discriminant Rate	100	44.44	64.71	76.92
	89.11			Total Discriminant rate		71.52		

Table 4. Confusion matrices of CARS-MIV-GA-SVM and CARS-MIV-LS-SVM models.

GA-SVM				Model	LS-SVM			
Good Peach	Internal Injury Peach	Scab Peach	Rotten Peach	Classification	Good Peach	Internal Injury Peach	Scab Peach	Rotten Peach
49	0	0	0	Good peach	49	0	0	0
0	8	1	0	Internal injury peach	0	4	5	0
0	1	16	0	Scab peach	0	5	12	0
0	0	1	12	Rotten peach	0	0	3	10
49	9	17	13	Total	49	9	17	13
100	88.9	91.4	92.3	Discriminant Rate	100	44.44	70.59	76.92
	93.15			Total Discriminant rate		71.52		

Table 5. Confusion matrices of CARS-SPA-GA-SVM and CARS-MIV-LS-SVM models.

GA-SVM				Model	LS-SVM			
Good Peach	Internal Injury Peach	Scab Peach	Rotten Peach	Classification	Good Peach	Internal Injury Peach	Scab Peach	Rotten Peach
49	0	0	0	Good peach	49	0	0	0
2	6	0	1	Internal injury peach	0	4	5	0
1	1	15	0	Scab peach	0	6	11	0
0	0	0	13	Rotten peach	0	0	2	11
49	9	17	13	Total	49	9	17	13
100	66.67	88.24	100	Discriminant Rate	100	44.44	64.71	84.62
	88.73			Total Discriminant rate		73.44		

Table 6. Confusion matrices of CARS-UVE-GA-SVM and CARS-MIV-LS-SVM models.

GA-SVM				Model	LS-SVM			
Good Peach	Internal Injury Peach	Scab Peach	Rotten Peach	Classification	Good Peach	Internal Injury Peach	Scab Peach	Rotten Peach
49	0	0	1	Good peach	49	0	0	0
0	6	2	0	Internal injury peach	0	7	2	0
0	0	17	12	Scab peach	0	4	13	0
0	0	1	13	Rotten peach	0	0	4	9
49	9	17	13	Total	49	9	17	13
100	66.67	100	92.31	Discriminant rate	100	77.77	76.47	69.23
	89.75			Total Discriminant rate		80.87		

Note: This is an example to explain the methods listed in the table: CARS-MIV means that the average influence value MIV algorithm was used to reduce the dimensions of 24 spectral variables extracted using the CARS method twice.

According to the discrimination results presented in Tables 3–6, the discrimination rate accuracy of the GA-SVM model established using the four methods of extracting characteristic wavelengths was much higher than that of the LS-SVM model. Therefore, the GA-SVM model was chosen as the optimal discrimination model. The comprehensive discrimination rates of GA-SVM models established using the CARS, CARS-SPA, and

CARS-UVE methods were 89.11%, 88.73%, and 89.75%, respectively. The comprehensive discriminant rate of the GA-SVM model established using the CARS-MIV method was 93.15%, and it had the best discriminant effect. Therefore, the CARS-MIV-GA-SVM model was chosen as the optimal discrimination model. The highest discriminant rate of an intact peach was 100%. One scab peach was misjudged as a dark wound because some scab peaches have a certain degree of dark wound inside, which is not visible to the human eye, so miscalculations can occur. One black peach was misjudged as a scab peach; due to the depth of the black wound, the sample's color characteristics showed obvious changes, resulting in machine identification as a scab peach. One rotten peach was misjudged as a scab peach; its rotten area was too small, resulting in machine identification as a scab peach. The GA-SVM model's optimization process and prediction results established using four feature wavelength extraction methods are shown in Figures 12–15, respectively.
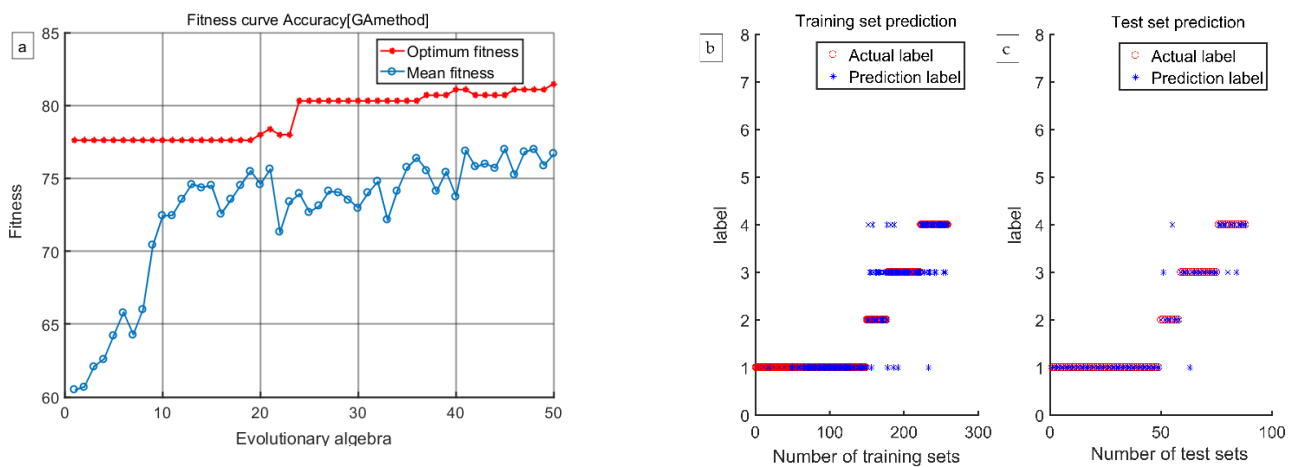
Figure 12. The CARS-GA-SVM model's optimization process and prediction results: (**a**) the optimization process, (**b**) training results, and (**c**) test results.
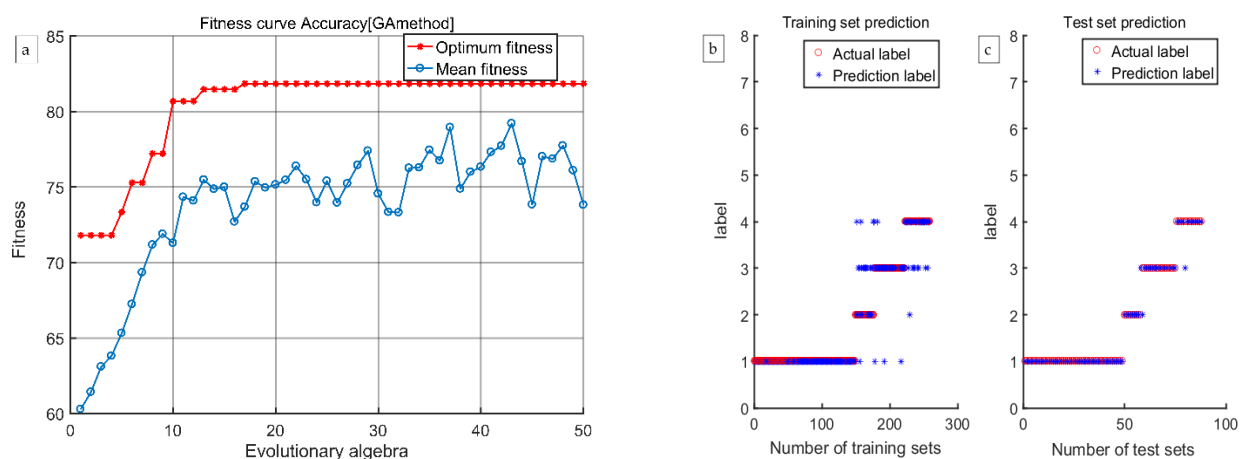

Figure 13. The CARS-MIV-GA-SVM model's optimization process and prediction results: (**a**) the optimization process, (**b**) training results, and (**c**) test results.

3.5. Discussion

In this experiment, we applied the CARS-MIV combination method to extract feature wavelengths and establish a detection model capable of simultaneously evaluating the internal and external qualities of Kubo peaches. Spectral data pre-processed using DSG were screened using four characteristic wavelength extraction methods: CARS, CARS-MIV, CARS-SPA, and CARS-UVE. Compared with other feature variable screening methods, the two models based on the CARS-MIV method had higher discriminant rates. Based

on the principle of the feature variables' importance, the CARS-MIV method was used as a secondary screening method after feature variables were screened based on CARS. First, a neural network training model was established according to selected characteristic variables, and a new independent variable was formed by adding and reducing 10%, respectively, based on the original independent variable; the difference value could be calculated by inputting the original independent variable into the trained neural network model to obtain the variable's influence change value. According to this principle, the importance of the selected variable was screened twice. Important wavelength variables can be screened more accurately. Xu L. [38] used hyperspectral imaging technology to nondestructively detect kiwi fruit's sugar content, a single feature wavelength extraction method, and a combination of multiple feature wavelengths to extract the feature wavelength, and then, carried out a comparative analysis. The research results showed that the model established using (CARS + IRIV)-SPA to extract the feature wavelength had the best prediction effect. This shows that it is feasible to use hyperspectral technology and a composition method to extract characteristic wavelengths and establish models to predict sugar content. Compared with the above studies, this experiment's combination algorithm added a step to sort according to wavelength importance, so that feature variables with higher importance could be screened more accurately.
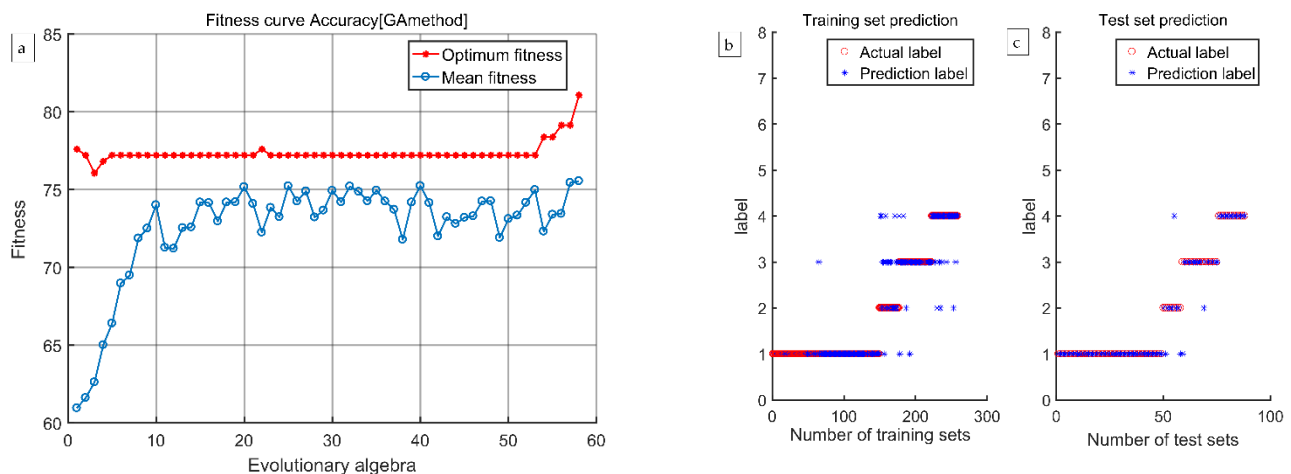

Figure 14. The CARS-SPA-GA-SVM model's optimization process and prediction results: (**a**) the optimization process, (**b**) training results, and (**c**) test results.
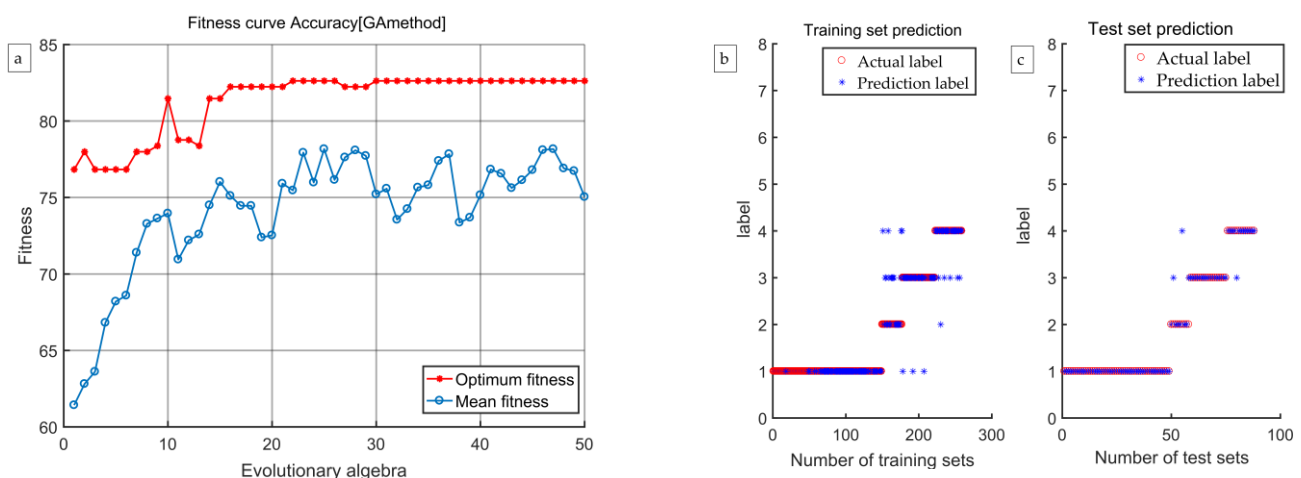

Figure 15. The CARS-UVE-GA-SVM model's optimization process and prediction results: (**a**) the optimization process, (**b**) training results, and (**c**) test results.

GA-SVM is also an optimization model that uses the classical genetic algorithm (GA) to optimize the C and γ parameters of SVM. Su J. [39] used the GA-SVM model formed using the GA optimization of SVM parameters to classify dried jujube varieties. Their research results showed that, compared with the traditional SVM model, the classification accuracy of the GA-SVM model was improved by more than 20%. Dan S. [40] used near-infrared spectroscopy and the GA-SVM model to study citrus-producing areas in 16 regions; their research results showed that the GA-SVM model's recognition rate for citrus-producing areas reached 94.58%. Compared with the above studies, in this study, the GA-SVM model was used for the first time to distinguish the external quality classification of the Kubo peach. The results presented in Tables 2–5 show that the discrimination rate accuracy of the GA-SVM model established using the four methods of extracting characteristic wavelengths was much higher than that of the LS-SVM model. Therefore, the GA-SVM model was chosen as the optimal discrimination model. The comprehensive discrimination rates of the GA-SVM model established using CARS, CARS-SPA, and CARS-UVE were 89.11%, 88.73%, 89.75%, and 88.59%, respectively. The comprehensive discriminant rate of the GA-SVM model established using CARS-MIV was 93.15%, and its discriminant effect was the best. Therefore, the CARS-MIV and GA-SVM models were combined to determine the Kubo peach's external quality.

In this study, we applied this model to identify the external quality of only one peach variety; it is not universal in the detection of the external quality of other peach varieties. Therefore, in a future study, we will use more fruit varieties to realize the model's potential detection capabilities, and to improve its universality.

4. Conclusions

To quickly detect Kubo peach internal injury defects and realize their accurate classification, Kubo peach internal and external defects were studied using hyperspectral imaging technology. A variety of pretreatment methods were used to process original spectral data, and the optimal pretreatment method was selected. The MIV algorithm was used to double-screen feature variables extracted using CARS according to their importance, GA was used to optimize the C and γ parameters of SVM to obtain the GA-SVM optimization model, and the optimized algorithm was compared with a single algorithm. Finally, it was concluded that combining CARS-MIV feature wavelength extraction with the GA-SVM model had the highest discrimination accuracy, optimized the external quality detection model for the Kubo peach, and provided a new research idea for the realization of Kubo peach quality detection. The conclusions drawn are as follows:

(1) The DSG spectral pretreatment method could better optimize spectral data, and spectral data after pretreatment could establish the PLS model with relatively high accuracy and low standard deviation. The correction set determination coefficient R_c^2 was 0.7521, and the root-mean-square error RMSEC was 0.5693. The prediction set determination coefficient R_p^2 was 0.8463, and the root-mean-square error RMSEP was 0.4563. Therefore, the data pre-processed using DSG were selected for the follow-up study.

(2) The data pre-treated using DSG were modeled and analyzed using four feature wavelength screening methods, namely, CARS, CARS-MIV, CARS-SPA, and CARS-UVE. The established GA-SVM model had a better discriminative effect than the LS-SVM model. Compared with the feature wavelength screening and model building methods, the GA-SVM model based on CARS-MIV-extracted feature wavelength data had the best prediction effect, and its discrimination accuracy reached 93.15%. Therefore, the CARS-MIV-GA-SVM model was selected as the optimal model to detect the Kubo peach's external quality.

In summary, we proposed a detection method based on hyperspectral imaging technology to simultaneously detect the internal injury and external defects of fresh Kubo peaches. We introduced the idea of extracting feature wavelengths by combining CARS-MIV, CARS-SPA, CARS-UVE, and other methods. We combined the feature wavelength algorithm and the optimization model (GA-SVM) to enhance the classification model's

accuracy. This approach addressed the limitations of existing peach grading methods that solely rely on appearance and overlook internal damage in peaches.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study can be made available by the corresponding author upon request.

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

References

1. Wang, X.D. New varieties of mid–ripening cold–resistant peach: Zhongnong Hanxiang Honey. *China Fruit Ind. Inf.* **2023**, *40*, 68. (In Chinese)
2. Tian, Y.; Sun, J.; Zhou, X.; Yao, K.; Tang, N. Detection of Soluble Solid Content in Apples Based on Hyperspectral Technology Combined with Deep Learning Algorithm. *J. Food. Preserv.* **2022**, *46*, 16414. [CrossRef]
3. Teerachaichayut, S.; Ho, H.T. Non–destructive Prediction of Total Soluble Solids, Titratable Acidity and Maturity Index of Limes by near Infrared Hyperspectral Imaging. *Postharvest Biol. Technol.* **2017**, *133*, 20–25. [CrossRef]
4. Zhong, Y.; Li, H.H.; Yuan, H.Y.; Xu, R.; Zhu, Y.Q.; Dai, X.J.; Zhou, Y.; Li, K.; Li, H.J. Effect of different grading methods on mechanical injury of Hongyang Kiwifruit. *Sichuan Agric. Sci. Technol.* **2018**, *10*, 41–43. (In Chinese)
5. Hu, M.; Dong, Q.; Liu, B.; Opara, U.; Chen, L. Estimating Blueberry Mechanical Properties Based on Random Frog Selected Hyperspectral Data. *Postharvest Biol. Technol.* **2015**, *106*, 1–10. [CrossRef]
6. Shao, Y.; Wang, Y.; Xuan, G.; Gao, Z.; Hu, Z.; Gao, C.; Wang, K. Assessment of Strawberry Ripeness Using Hyperspectral Imaging. *Anal. Lett.* **2020**, *54*, 1547–1560. [CrossRef]
7. Baranowski, P.; Mazurek, W.; Wozniak, J.; Majewska, U. Detection of Early Bruises in Apples Using Hyperspectral Data and Thermal Imaging. *J. Food Eng.* **2012**, *110*, 345–355. [CrossRef]
8. Mehl, P.; Chen, Y.; Kim, M.; Chan, D. Development of Hyperspectral Imaging Technique for the Detection of Apple Surface Defects and Contaminations. *J. Food Eng.* **2004**, *61*, 67–81. [CrossRef]
9. Babellahi, F.; Paliwal, J.; Erkinbaev, C.; Amodio, M.L.; Chaudhr, M.M.A.; Colelli, G. Early Detection of Chilling Injury in Green Bell Peppers by Hyperspectral Imaging and Chemometrics. *Postharvest Biol. Technol.* **2022**, *162*, 111100. [CrossRef]
10. Li, J.; Huang, W.; Tian, X.; Wang, C.; Fan, S.; Zhao, C. Fast Detection and Visualization of Early Decay in Citrus Using Vis–NIR Hyperspectral Imaging. *Comput. Electron. Agric.* **2016**, *127*, 582–592. [CrossRef]
11. Zheng, Z.; An, Z.; Yang, Y.; Chen, J.; Liu, X.; Lu, L. Metabolic profiling of blueberries (*Vaccinium* Spp.) to quantitatively and qualitatively assess bruise damage and fruit deterioration. *Postharvest Biol. Technol.* **2023**, *195*, 112135. [CrossRef]
12. Tang, Y.; Gao, S.; Zhuang, J.; Hou, C.; He, Y.; Chu, X.; Miao, A.; Luo, S. Apple Bruise Grading Using Piecewise Nonlinear Curve Fitting for Hyperspectral Imaging Data. *IEEE Access* **2020**, *8*, 147494–147506. [CrossRef]
13. Wang, Z.; Hu, D.; Sun, Z.; Ying, Y. Application status and prospect of spatial frequency domain imaging in agricultural product quality detection. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 275–288. (In Chinese)
14. Li, Y.; Ma, B.; Li, C.; Yu, G. Accurate prediction of soluble solid content in dried Hami jujube using SWIR hyperspectral imaging with comparative analysis of model. *Comput. Electron. Agric.* **2022**, *193*, 106655. [CrossRef]
15. Xu, D.; Wang, H.; Ji, H.; Zhang, X.; Wang, Y.; Zhang, Z.; Zheng, H. Hyperspectral Imaging for Evaluating Impact Damage to Mango According to Changes in Quality Attributes. *Sensors* **2018**, *18*, 3920. [CrossRef] [PubMed]
16. Li, Y.; Yin, Y.; Yu, H.; Yuan, Y. Fast detection of water loss and hardness for cucumber using hyperspectral imaging technology. *J. Food Meas. Charact.* **2022**, *16*, 76–84. [CrossRef]
17. Pullanagari, R.R.; Li, M. Uncertainty assessment for firmness and total soluble solids of sweet cherries using hyperspectral imaging and multivariate statistics. *J. Food. Eng.* **2021**, *289*, 110177. [CrossRef]
18. Wang, K.; Guo, D.; Zhang, G.; Deng, L.; Xin, R.; Lv, Q.; Yi, S.; Zheng, Y.; Ma, Y.; He, S. Detection of Huanglongbing (citrus greening) based on hyperspectral image analysis and PCR. *Front. Agric. Sci. Eng.* **2019**, *6*, 172–180. [CrossRef]
19. Zhang, H.; Chen, Y.; Liu, X.; Huang, Y.; Zhan, B.; Luo, W. Identification of Common Skin Defects and Classification of Early Decayed Citrus Using Hyperspectral Imaging Technique. *Food Anal. Methods* **2021**, *14*, 1176–1193. [CrossRef]

20. Heo, S.; Choi, J.; Kim, J.; Moon, K. Prediction of moisture content in steamed and dried purple sweet potato using hyperspectral imaging analysis. *Food Sci. Biotechnol.* **2021**, *30*, 783–791. [CrossRef]

21. Zhang, H.; Gao, J.; He, Y. Nondestructive detection of citrus defect based on hyperspectral imaging. *Trans. Chin. Soc. Agric. Mach.* **2013**, *44*, 177–181. (In Chinese)

22. Chen, S. Research on Surface Defect Detection Method of Peach Based on Hyperspectral Image Technology. Doctoral Thesis, Zhejiang University, Hangzhou, China, 2013. (In Chinese).

23. Wang, D.; Wang, K.; Wu, J.; Han, P. Research progress of non–destructive rapid measurement of seed quality based on spectral and imaging techniques. *Spectrosc. Spect. Anal.* **2021**, *41*, 52–59.

24. Song, S.; Yang, Y.; Wang, H.; Wang, X.; Rong, S.; Zhou, H. Anomaly detection method of hyperspectral image sparse constraint combined with autoencoder feature extraction. *J. Surv. Mapp.* **2019**, *52*, 932–943. (In Chinese)

25. Hu, H.; Wei, Y.; Xu, H.; Zhang, L.; Mao, X.; Zhao, Y. Identification of the age of rodenticle based on hyperspectral imaging and principal component analysis. *Spectrosc. Spect. Anal.* **2023**, *43*, 1953–1960.

26. Zhu, C.; Liu, Y.; Li, X.; Gong, W.; Guo, W. Hyperspectral detection of the freshness of *Litopenaeus vannamei*. *Spectrosc. Spect. Anal.* **2023**, *43*, 107–110.

27. Song, H.; Yan, Y.; Song, Z.; Sun, J.; Li, Y.; Li, F. Nondestructive testing model for maize grain moisture content established by screening dielectric parameters and variables. *Trans. Chin. Soc. Agric. Eng.* **2019**, *35*, 262–272. (In Chinese)

28. Chen, D.; Yan, M.; Li, Q.; Yu, L.; Jin, Y.; Xu, K. Development of a novel nonlinear variable selection strategy for spectral analysis. *Nanotechnol. Precis. Eng.* **2015**, *13*, 226–231.

29. Hou, J.; Li, Z.; Cao, X.; Song, X. Integrating genetic algorithm and support vector machine for polymer flooding production performance prediction. *J. Pet. Sci. Eng.* **2009**, *68*, 29–39. [CrossRef]

30. Harper, R.; Taylor, A. Genetic Algorithm Based Optimization of One–class Support Vector Machine for Abnormality Detection in Oilfield Centrifugal Pump Injection Stations. *Sci. Technol. Eng.* **2023**, *23*, 283–289. (In Chinese)

31. Lu, W.; Cai, M.; Zhang, Q.; Li, S. Rapid classification method of black Wolfberry based on hyperspectral and ensemble learning. Spectrosc. *Spect. Anal.* **2019**, *41*, 2196–2204.

32. Yang, Q.Z.; Feng, W. Effect of polyvinylpyrrolidone on the quality and storage of 'Okubo' peach fruits. *Food Sci.* **2016**, *37*, 264–269. (In Chinese)

33. Wang, W.; Feng, W.; Chang, N.; Liu, Q.; Li, Z.; Chen, Y.; Li, C.; Chen, X.; Zhang, Y. Prediction of chlorophyll content in flue–cured tobacco based on spectral pretreatment and machine learning algorithm. *Chin. J. Soil Fertil.* **2023**, *60*, 194–201. (In Chinese)

34. Yuan, L.; Shi, B.; Yu, J.; Tang, T.; Yuan, Y.; Tang, Y. Application of different smooth integrated CARS algorithms in the spectral identification of black tea grades. *Spectrosc. Spect. Anal.* **2020**, *40*, 3254–3259.

35. Xia, Y.; Feng, Q.; Yang, S.; Guo, F. Soil moisture estimation of Apple orchard based on hyperspectral remote sensing. *For. Mach. Woodwork. Equip.* **2019**, *51*, 24–32. (In Chinese)

36. Gao, S.; Xu, J.H. Prediction and visualization of total acid and hardness of red extract by hyperspectral imaging. *Food Sci.* **2023**, *44*, 327–336. (In Chinese)

37. Tan, H.Y. Research on classification and recognition of Chinese tea based on THz–TDS and PSO–LS–SVM. *Res. China Sci. Technol. J. Database* **2022**, *14*, 289–293. (In Chinese)

38. Xu, L.; Chen, M.; Wang, Y.; Chen, X.; Lei, X. Nondestructive detection of sugar content of kiwifruit using hyperspectral imaging. *Spectrosc. Spect. Anal.* **2019**, *41*, 2188–2195.

39. Su, J.; Rao, Y.; Zhang, J.; Li, S. Dry jujube variety classification based on GA optimization SVM. *J. Luoyang Inst. Technol.* **2018**, *28*, 65–69+93. (In Chinese)

40. Dan, S.J. Study on NIR spectral identification model of citrus origin based on SVM and GA–SVM. *J. Hunan Polytech. Ind.* **2021**, *22*, 30–35+46. (In Chinese)