

Article

Research on the Maturity Detection Method of Korla Pears Based on Hyperspectral Technology

Jiale Liu ^{1,2} and Hongbing Meng ^{1,2,*}¹ College of Information Engineering, Tarim University, Alar 843300, China; 10757223144@stumail.taru.edu.cn² Key Laboratory of Tarim Oasis Agriculture, Tarim University, Ministry of Education, Alar 843300, China

* Correspondence: mhb@taru.edu.cn; Tel.: +86-186-9977-6234

Abstract: In this study, hyperspectral imaging technology with a wavelength range of 450 to 1000 nanometers was used to collect spectral data from 160 Korla pear samples at various maturity stages (immature, semimature, mature, and overripe). To ensure high-quality data, multiple preprocessing techniques such as multiplicative scatter correction (MSC), standard normal variate (SNV), and normalization were employed. Based on these preprocessed data, a custom convolutional neural network model (CNN-S) was constructed and trained to achieve precise classification and identification of the maturity stages of Korla pears. Additionally, a BP neural network model was used to determine the characteristic wavelengths for maturity assessment based on the sugar content feature wavelengths. The results demonstrated that the BP model, based on sugar content feature wavelengths, effectively discriminated the maturity stages of the pears. Specifically, the comprehensive recognition rates for the training, testing, and validation sets were 98.5%, 93.5%, and 90.5%, respectively. Furthermore, the combination of hyperspectral imaging technology and the custom CNN-S model significantly enhanced the detection performance of pear maturity. Compared to traditional CNN models, the CNN-S model improved the accuracy of the test set by nearly 10%. Moreover, the CNN-S model outperformed existing techniques based on partial least squares discriminant analysis (PLS-DA) and support vector machine (SVM) in capturing hyperspectral data features, showing superior generalization capability and detection efficiency. The superior performance of this method in practical applications further supports its potential in smart agriculture technology, providing a more efficient and accurate solution for agricultural product quality detection. Additionally, it plays a crucial role in the development of smart agricultural technology.

Keywords: hyperspectral; Korla pear; maturity detection; nondestructive testing; CNN-S model



Citation: Liu, J.; Meng, H. Research on the Maturity Detection Method of Korla Pears Based on Hyperspectral Technology. *Agriculture* **2024**, *14*, 1257. <https://doi.org/10.3390/agriculture14081257>

Academic Editor: Juan Ignacio Arribas

Received: 1 July 2024

Revised: 24 July 2024

Accepted: 27 July 2024

Published: 30 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Korla pears are renowned for their thin skin and juicy texture, enjoying a high reputation both domestically and internationally [1]. They are significant drivers of economic development in the Xinjiang region and are recognized as a geographical indication product under the China–EU agreement [2]. The cultivation of Korla pears has a history of over 1400 years in their native region of Korla, Xinjiang [3]. In addition to Korla, these pears are also grown in nearby areas such as Luntai County, Aral City, Aksu City, and Awat County [4]. Known for their rich flavor, thin skin without residue, crispness, juiciness, and high nutritional value, Korla pears are considered a premium variety of pears [5].

With the continuous improvement in living standards, consumers' demand for quality has been increasing. They require the pears not only to look good but also to have excellent internal nutritional qualities [6]. Currently, the determination of fruit maturity largely relies on farmers' experience, which can significantly affect the fruit's market value. Immature or overripe pears often fail to meet optimal production, processing, and storage standards, adversely affecting both farmers' interests and the nutritional value of the fruit.

Maturity in Korla pears refers to the physiological and quality characteristics that influence their flavor, storage, transportation, and processing. Higher maturity levels enhance qualities like color, aroma, sweetness, acidity, hardness, and juiciness, making them more appealing to consumers [7,8]. The maturity level also impacts the storage and transportation of the pears. Immature pears are harder and less prone to damage but lack flavor, making them hard to sell. Overripe pears, on the other hand, are softer and more prone to bruising and damage, which hinders long-distance transportation and storage, making them susceptible to pests and rot [9]. Maturity also affects the processing and utilization of the pears. Ideally matured pears can be processed into products like canned pears, pear juice, dried pears, pear wine, and pear vinegar, increasing their added value and economic benefits. Therefore, mastering the optimal harvest time and maturity evaluation methods is key to enhancing the industry level and competitiveness of Korla pears.

Traditional detection methods are mainly manual, consuming significant human, material, and financial resources. The study of detecting the maturity of Korla pears using hyperspectral imaging technology can significantly improve detection efficiency and accuracy, reduce reliance on human resources, and save considerable time and costs [10]. This technology enables rapid online monitoring of pear maturity, improving processing and sales efficiency, ensuring fruit quality, and increasing consumer satisfaction. Hence, this method holds substantial significance for the development of smart agriculture [11].

Several studies have demonstrated the effectiveness of hyperspectral imaging in non-destructive quality assessment of agricultural products [12]. Rajkumar et al. [13] used hyperspectral imaging to establish a multiple linear regression (MLR) model to predict the quality (moisture, hardness, and soluble solids) of bananas at different maturity stages, showing successful predictions. Zhang et al. [14] employed principal component analysis (PCA) to select optimal wavelengths (441.1 to 1013.97 nm) and combined texture features to develop a support vector machine (SVM) model for strawberry maturity identification, achieving an accuracy of 95.0%. Ribera-Fonsec et al. [15] used PCA to analyze the absorbance differences at 560 nm and 640 nm wavelengths for grape maturity, measuring quality indices (SSC, TA, DI), and showed good prediction capabilities. Munera S et al. [16] used NIR hyperspectral imaging to establish partial least squares regression (PLSR) models for the maturity of two nectarine varieties, achieving successful discrimination. Chu Zhang et al. [14] used hyperspectral imaging to establish an SVM classification model for different maturity stages of strawberries, demonstrating that the combination of spectral and texture features provided the best maturity classification results. Shao et al. [17] used hyperspectral imaging to classify mature, mid-ripe, and unripe strawberries, employing PLS-DA and LS-SVM for maturity evaluation, showing that the LS-SVM model with effective wavelengths performed best in field evaluations, with an accuracy of 96.7%. Sun et al. [18] established SVM prediction models for soluble solids content and hardness of Hami melons based on hyperspectral technology, demonstrating that hyperspectral technology can achieve simultaneous quantitative prediction of quality (SSC, hardness) and qualitative maturity classification. Shao et al. [19] studied the quality and maturity of Feicheng peaches using hyperspectral imaging, determining SSC and hardness values, showing that the optimal models for predicting SSC and hardness were CARS-MLR, and the ANN could effectively classify peach maturity. Jiang Hao et al. [20] used hyperspectral imaging technology to establish Fisher linear discriminant models for maturity classification of strawberries based on existing and newly constructed parameters, showing that hyperspectral imaging can achieve automatic strawberry maturity identification.

In summary, although hyperspectral technology is recognized as an effective method for nondestructive quality assessment of agricultural products, research on the quality and maturity detection of Korla pears is relatively limited, with most studies focusing on hardness and SSC determination. This study uses hyperspectral imaging technology to successfully collect spectral data of Korla pears and develop new machine learning models for maturity assessment, significantly improving detection efficiency and accuracy. Compared to traditional methods, our technique improved test set accuracy by nearly

10%, providing strong technical support and theoretical basis for the rapid, nondestructive detection of the quality and maturity of Korla pears. These improvements not only optimize data processing but also enhance model generalization, representing significant progress in pear quality monitoring and smart agriculture development.

2. Materials and Methods

2.1. Collection of Korla Pear Samples

Aral City, located in southern Xinjiang, is one of the main cultivation areas for Korla pears. The region experiences a typical temperate continental arid desert climate, with an average annual temperature of 10.7 °C to 11.2 °C, a frost-free period of 170 to 227 days, and annual sunshine hours ranging from 2762.1 to 3186.3 h. The abundant light and heat resources, along with the significant diurnal temperature variation, contribute to the superior quality of the pears.

To ensure the representativeness and scientific validity of the experimental data, this study harvested Korla balsam pear from Batuan pear orchard (81°41'39.86" E, 40°36'51.79" N) in Alar City. The orchard is a high-yield pear orchard (≥ 1580 kg/667 m²) located in the temperate desert climate zone; the soil is mainly sandy loam, and the spacing between rows of fruit trees is 4 m × 5 m. Fruit trees with an age of 10 to 15 years were selected in the four directions of each orchard, namely, east, west, south, and north, to harvest the fruit. In order to ensure the accuracy of experimental data, each stage of the collection of the appearance of the size and color of the balsam pear is basically the same. The location of the experimental field is shown in Figure 1.

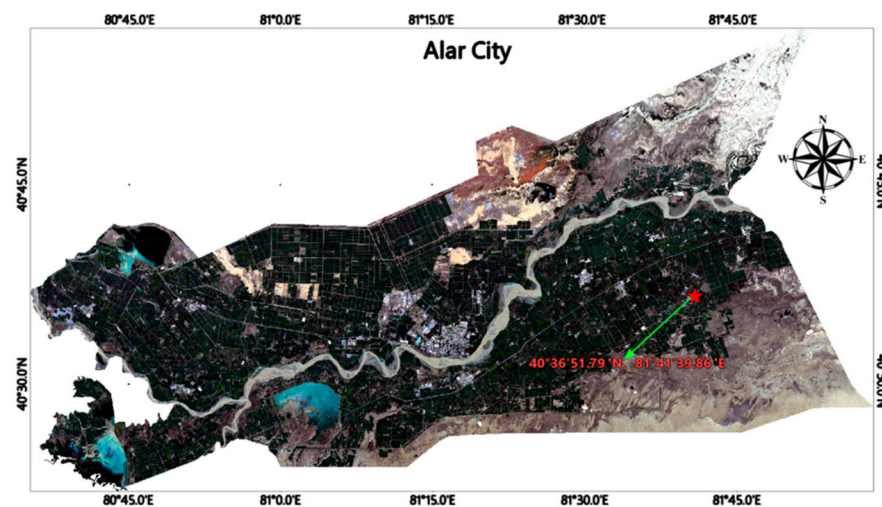


Figure 1. Location of the experimental field.

A total of 600 samples were collected, with 150 samples from each of four different stages. The early harvest stage (first stage) was in early August, specifically on 2 August 2023. The late August harvest (second stage) occurred on 29 August 2023. The mid-September harvest (third stage) was on 25 September 2023, and the late October harvest (fourth stage) was on 25 October 2023. Each stage is characterized as shown in Table 1.

Stratified sampling of the Korla balsam pear samples was conducted according to different time points (2 August, 29 August, 25 September, and 25 October), and the 600 samples were divided into prediction and validation sets at a ratio of 2:1 to ensure the representativeness and scientificity of the data. Hyperspectral images of balsam pear were collected using a hyperspectral imaging system (wavelength range 380 to 1000 nm). The sweetness value of the balsam pear was then measured using an ATAGO portable digital fruit brix meter. The acquired hyperspectral images were corrected with black and white plates, and preprocessing methods such as multiple scattering correction (MSC), standard normal variable transformation (SNV), and normalization were used to remove noise and interference

in the spectral data; then, these methods were compared and analyzed to select the best preprocessing method. The algorithm for extracting feature wavelengths was selected, and the three methods of principal component analysis (PCA), partial least squares regression (PLS), and successive projection method (SPA) were used for comparative analysis, to finalize the best feature wavelength extraction method and to screen out eight key wavelengths (460 nm, 570 nm, 594 nm, 610 nm, 670 nm, 730 nm, 810 nm, and 890 nm). Based on the full-spectrum information and the characteristic wavelength information, the ripeness classification and discrimination model of pear was established based on partial least squares (PLS), BP neural network, and support vector machine (SVM), respectively. The accuracy and robustness of the models were ensured through cross-validation and parameter optimization. The specific steps and methods of the research process are shown in Figure 2, and systematically evaluated the quality characteristics of the samples at different harvesting stages and ensured the scientific validity and reliability of the research results.

Table 1. Quality characteristics at different stages.

Different Phases	Color	Character	Flavor
Phase 1	It is starting to turn yellow, but there may be some green left.	Begins to soften, but may still be hard in the center.	Some sweetness begins to appear, but the sourness remains more pronounced.
Phase 2	Green or slightly yellowish rind with a more uniform overall color.	Flesh is firmer, slightly hard but not too hard, with some elasticity to the touch.	Light, lightly sweet flavor and less pronounced aroma. Suitable for storage, transportation, and processing.
Phase 3	Sharp yellow or yellow–green color, possibly with some reddish tinge.	The texture is soft and juicy.	Sweetness is noticeable and aroma is strong, making it the most suitable stage for consumption.
Phase 4	The color may become darker and some parts may begin to turn brown.	Too soft and may begin to show signs of rotting.	Overly sweet or fermented flavor and texture may begin to become muddy.

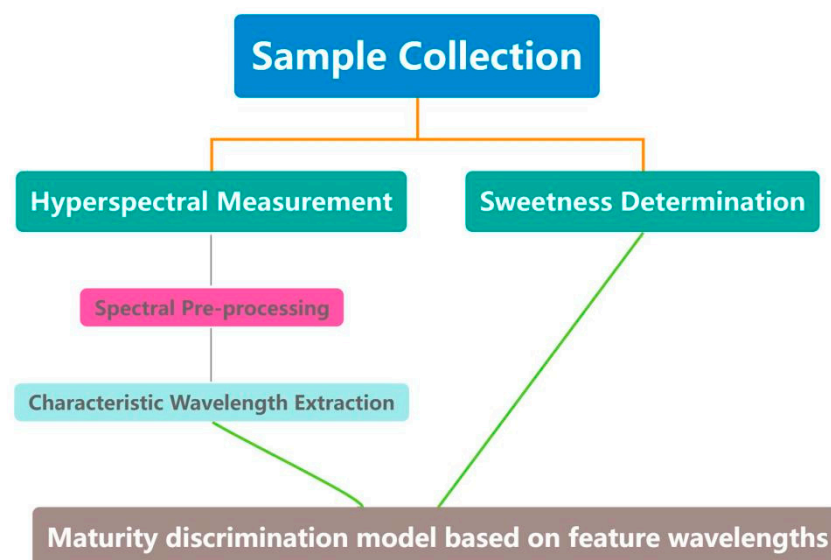


Figure 2. The procedure of spectral processing and sweetness determination in this study.

2.2. Hyperspectral Image Acquisition and Sugar Content Measurement of Korla Pears

Before each experiment, dust was carefully removed from the surface of the Kuril balsam pear using soft paper, and each balsam pear was numbered for hyperspectral imaging and brix determination. The hyperspectral imaging system used in this study

consisted mainly of an ImSpector V10E (Specim, Finland) imaging spectrometer. The system has a spectral range of 380 to 1000 nm and is equipped with a CCD array detector, two fiber-optic halogen lamps with adjustable light intensity of 150 W, a light source with emission wavelengths ranging from 400 to 1000 nm and high emissivity in this wavelength range, an OLE23 C-mount imaging lens, and OBF570 filters, as shown in Figure 3.

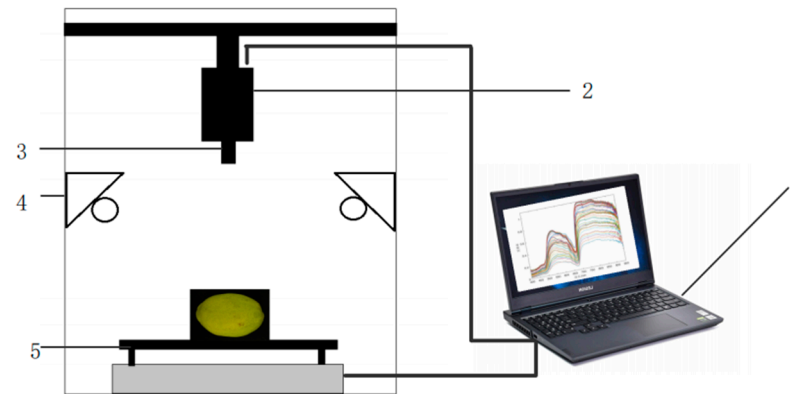


Figure 3. Block diagram of a hyperspectral imaging system. 1: Computer 2: Imaging spectrometer 3: Lens 4: Fiber optic halogen lamp 5: Motorized panning stage.

During the measurement process, a portable digital refractometer (ATAGO PAL-1) was used to measure the sugar content of the pears. The detection range of the refractometer was 0–53%, with a measurement accuracy of $\pm 0.2\%$. Before each experiment, the optimal data acquisition parameters for the hyperspectral instrument were determined: the distance between the object and the sensor was set to 300 mm, the exposure time was 1.2 ms, the scanning speed of the motorized stage was 0.6014 nm/s, and the actual length of the scanning line was 200 mm. During the measurement, the camera was placed vertically and a black background was used to minimize the effects of neighboring shadows and reflections and a reflection coefficient of 0.03.

After completing the hyperspectral image acquisition of the Korla pears, the sugar content of the pears was measured using the ATAGO PAL-1 portable digital refractometer. Before the sugar content measurement, the peel of the pear sample at the measurement site was removed. Each sample was measured at three different positions (as shown in Figure 4). The average of the five measurements for each pear sample was taken as the reference sugar content value for that sample.



Figure 4. Measurement of brix at this location.

2.3. Spectral Preprocessing

To prevent noise in the hyperspectral images caused by dark current and uneven light intensity in the hyperspectral camera, black and white reference corrections were performed on the collected hyperspectral images. The corrected images were calculated using the following equation:

$$Ra = \frac{Ro - Rb}{Rw - Rb}$$

where

R_o is the original hyperspectral image of the pear.

R_b is the black reference image: The black reference image is recorded by photographing an object or background that is completely black. This image is used to correct for dark current and noise in the image to ensure image purity and contrast.

R_w is the white reference image: The white reference image is recorded by photographing a completely white object or background. The white reference image is used to correct for uneven brightness distribution in the image to ensure accurate brightness in the final image.

R_a is the corrected hyperspectral image.

The original spectrum data contain not only information from the sample itself but also noise, such as stray light and sample background, which interferes with the spectral information of the sample. To eliminate noise interference on the spectral curve, preprocessing methods such as multiplicative scatter correction (MSC), standard normal variate (SNV), and normalization were used in this study.

MSC algorithm: This algorithm normalizes data at each wavelength to eliminate the influence of multiplicative scatter, thus improving data quality and interpretability. The MSC algorithm formula is as follows: Calculate the average spectral vector $\bar{x}$ for each sample, which is the average value of all samples at the same wavelength:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N X_i$$

Calculate the average spectral vector X_j for each wavelength, which is the average value of all samples at the same wavelength:

$$X_j = \frac{1}{M} \sum_{j=1}^M X_{ij}$$

Calculate the standard deviation δ_i for each wavelength, which is the standard deviation of all samples at the same wavelength:

$$\delta_i = \sqrt{\frac{1}{M} \sum_{i=1}^M (X_{ij} - \bar{X}_j)^2}$$

Perform MSC correction for each sample using the following formula:

$$X'_{ij} = \frac{X_{ij} - \bar{X}_j}{k\delta_i} + \bar{X}$$

where X_{ij} is the spectral value of the i th sample at the j th wavelength, N is the number of samples, M is the number of wavelengths, k is an adjustment parameter (usually 1), X'_{ij} is the MSC-corrected spectral value, and $\bar{X}_j$ is the global average spectral vector.

SNV algorithm: This algorithm standardizes the data at each wavelength to conform to a standard normal distribution, eliminating fluctuations caused by spectral intensity differences and making the data closer to independently and identically distributed characteristics, thus improving feature interpretability. The SNV algorithm formula is as follows:

Calculate the average spectral vector $\bar{X}_j$ for each sample:

$$\bar{X}_j = \frac{1}{N} \sum_{i=1}^N X_{ij}$$

Calculate the standard deviation δ_i for each wavelength:

$$\delta_i = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (X_{ij} - \bar{X}_j)^2}$$

Standardize each sample using SNV:

$$X'_{ij} = \frac{X_{ij} - \bar{X}_j}{\delta_j}$$

where X_{ij} is the spectral value of the i th sample at the j th wavelength, N is the number of samples, M is the number of wavelengths, $\bar{X}_j$ is the global average spectral vector, and σ_j is the global standard deviation. All preprocessing methods were implemented using MATLAB R2024a.

2.4. Feature Wavelength Selection Methods

Currently, the successive projections algorithm (SPA) is increasingly used for extracting characteristic wavelengths. It effectively removes collinearity between wavelength variables and avoids information redundancy, allowing a small amount of information to represent most of the spectral information of the samples. Stepwise multivariate linear regression (SMLR) is a commonly used mathematical method for selecting regression variables in multivariate linear regression by choosing the optimal bands based on the correlation coefficients between quality parameters and wavelengths. The basic idea is to perform a significance test on each wavelength included in the regression equation at each step of the calculation to determine its significant impact on y . Insignificant wavelengths are removed, and only those significantly affecting y remain. New wavelengths are selected and validated for their significant characteristics, ensuring that the regression equation contains only the most influential wavelengths. In this study, SPA and SMLR methods, along with tenfold cross-validation and 100 iterations of Monte Carlo sampling, were used to determine the features with the lowest cross-validation RMSE.

2.5. Discriminant Modeling Methods

Convolutional neural networks (CNNs) are a deep learning technique widely applied in image and video recognition, recommendation systems, and image classification. CNNs simulate the human visual system, automatically and effectively identifying complex patterns and features in images. They consist of multiple layers, mainly including convolutional layers, pooling layers, and fully connected layers. Convolutional layers extract basic visual features from the input image using filters; pooling layers reduce feature dimensions and enhance model generalization; fully connected layers, at the final stage of the network, transform extracted features into the final output. In this study, we utilized CNNs to analyze and identify the maturity of Korla pears, training the model to recognize spectral images at different maturity stages for rapid and accurate maturity discrimination.

Additionally, we used a BP neural network model for maturity discrimination based on the sugar content characteristic wavelengths of Korla pears. We selected spectral features of 100 samples at different maturity stages as input vectors for the BP model, with 50 samples as test samples. The number of nodes in the input layer was equal to the number of input variables, with one hidden layer node and one output layer node. The transfer functions for the input and hidden layers were tangent sigmoid functions, while the output layer used a linear transfer function. The training method used was gradient descent with a momentum factor of 0.6, a training precision of 0.001, and a maximum of 5000 iterations.

We also designed a new modeling method based on convolutional neural networks, named CNN-S. This model employed depthwise separable convolution, applying filters to each input channel individually for depthwise convolution, followed by 1×1 pointwise convolution for feature combination, reducing the number of model parameters and computational complexity. CNN-S integrated batch normalization and ReLU activation functions in all convolutional layers, enhancing the model's learning ability and efficiency in processing hyperspectral image data. The stacked convolutional and pooling layers iteratively extracted features from pear image data, improving generalization and accelerating convergence. In the final stage, an adaptive average pooling layer simplified feature

processing before the fully connected layer, further improving the classification efficiency of different maturity levels of pears. The fully connected layer mapped the extracted features to maturity levels (immature, semimature, mature, overripe), assigning a numerical value (1, 2, 3, 4) for classification.

All hyperspectral images of pear samples were collected using the SpectralSENS software 17.0, and the collected hyperspectral image data were processed and analyzed using ENVI 5.6 and MATLAB R2024a software.

3. Results and Analysis

3.1. Hyperspectral Images of Pears at Different Maturity Stages

Figure 5 shows the hyperspectral images of Korla pears at different maturity stages (first stage, second stage, third stage, and fourth stage).

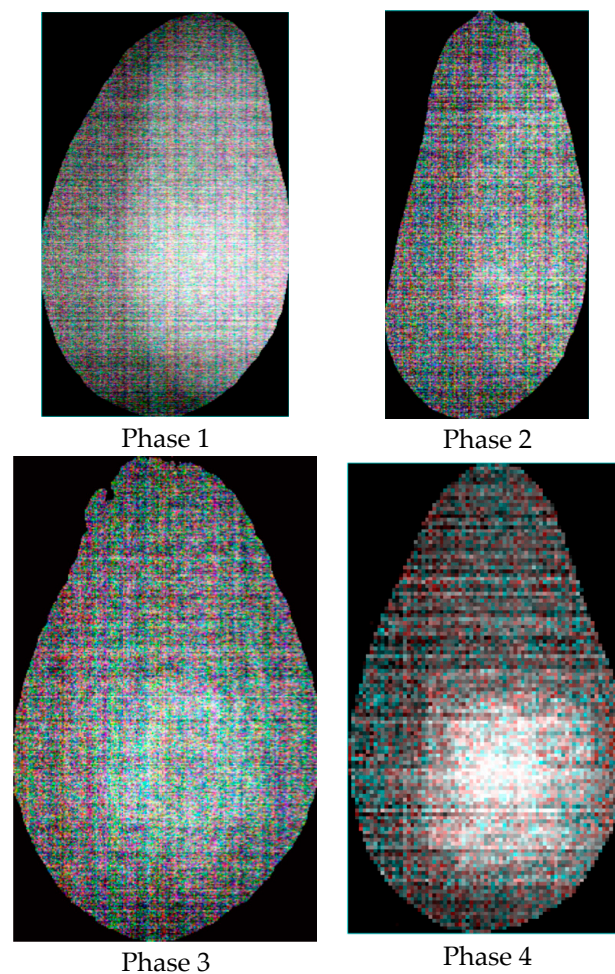


Figure 5. Hyperspectral images of balsam pear at different times.

3.2. Statistical Analysis

Analysis of Average Spectral Curves of Korla Balsam Pear at Different Ripening Stages

Figure 6 displays the average spectral curves of Korla pears at different maturity stages. It can be observed that unripe pears exhibit higher spectral reflectance, likely due to their physical surface properties. Specifically, the surface of unripe pears may be rougher or less smooth, which could be due to the fruit's incomplete development, harder skin, and lower water content, leading to increased light reflection and reduced absorption. The level of reflectance is closely related to the pear's maturity. High reflectance not only indicates that the pear is not fully ripe but also suggests that its taste, flavor, and nutritional value have not reached their optimal state. Conversely, mature or overripe pears exhibit lower

reflectance, typically due to thinner skin and higher water content, allowing for greater light absorption. Therefore, monitoring the spectral reflectance of pears can effectively determine their maturity.

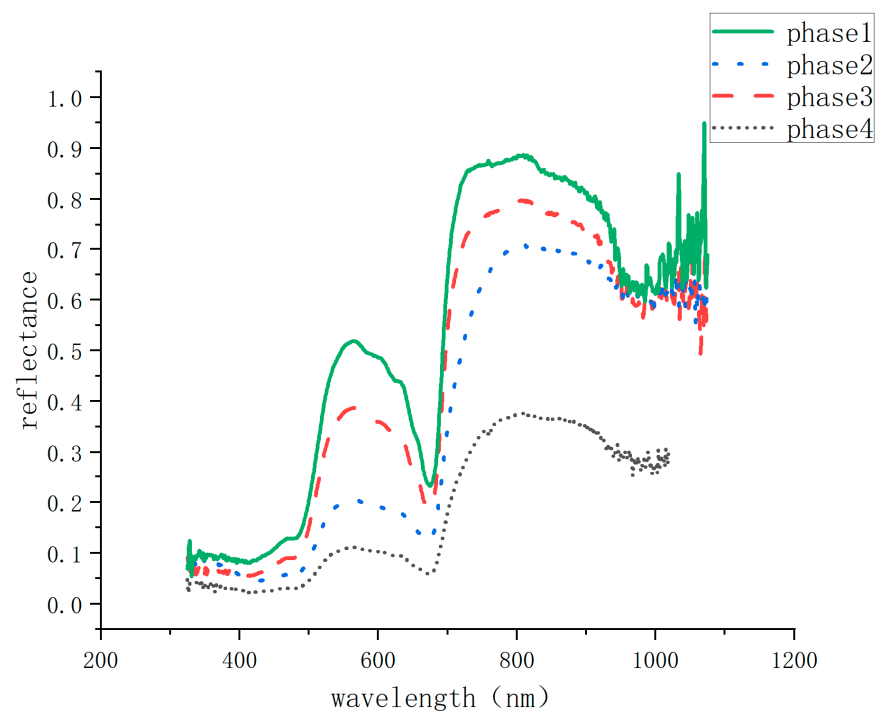


Figure 6. Average spectrograms for different phases.

3.3. Statistical Analysis of Soluble Sugar Content in Korla Pears at Different Maturity Stages

Four stages of balsam pear brix were evaluated by analysis of variance (ANOVA). The results in Table 2 show that the sweetness of balsam pear increased from 8.06% in the first stage to 13.6% in the fourth stage as the picking date progressed and ripened. The average values of brix at the four stages of balsam pear are shown in Figure 7.

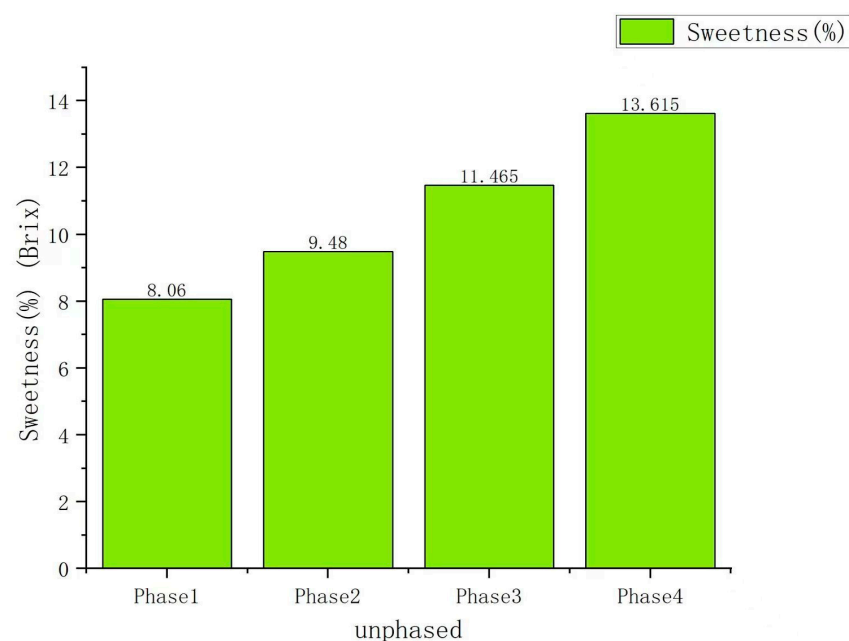


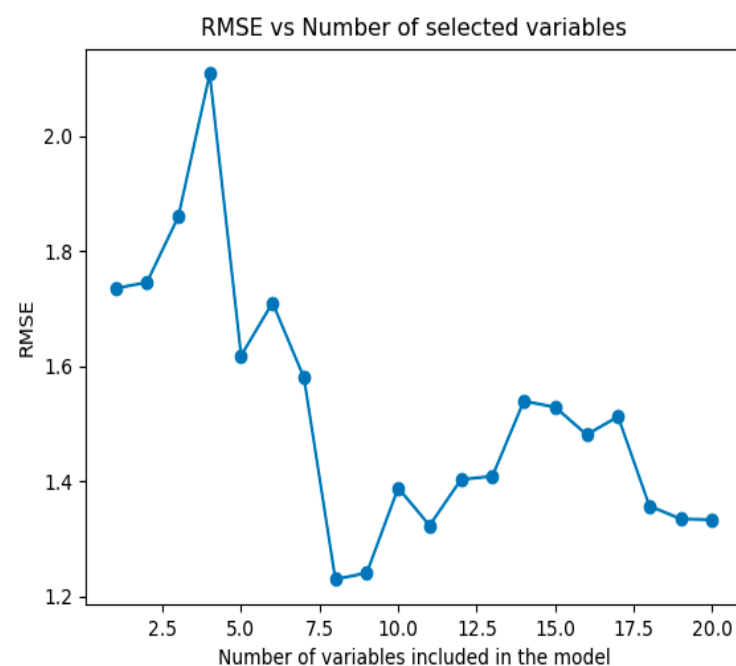
Figure 7. Average brix values of balsam pear at different stages.

Table 2. Statistical analysis of soluble sugar content in Korla balsam pears at different maturity stages.

Unphased	Sample Size	Sweetness (%)			
		Minimum	Maximum	Average	Standard Deviation
Phase 1	150	7.68	8.43	8.06	0.375
Phase 2	150	9.15	9.8	9.48	0.325
Phase 3	150	10.33	12.6	11.465	1.135
Phase 4	150	12.48	14.75	13.615	1.135

3.4. Selection of Characteristic Wavelengths Based on Soluble Sugar Content

In this experiment, the continuous projection method was used to extract the characteristic wavelengths, and the relative reflectance of the hyperspectra at full wavelengths was subjected to SPA operation. A total of 500 wavelengths of data between 450 and 950 nm were selected for the extraction of the characteristic wavelengths of the saccharinity in this experiment, because of the noise that may occur in the image acquisition. The number of wavelengths in this experiment was 500, and the total number of optimal spectra was determined based on the root mean square error (RMSE) of prediction of the calibration set. The number of wavelengths corresponding to the smallest RMSE is the total number of optimal spectral variables. When the selected characteristic wavelength is eight, the RMSE reaches the minimum, and the minimum value is 1.22954. Finally, eight characteristic wavelengths of balsam pear brix were selected by the continuous projection method, which are 460, 570, 594, 610, 670, 730, 810, and 890 nm, respectively. The results of the SPA run are shown in Figure 8.

**Figure 8.** Cont.

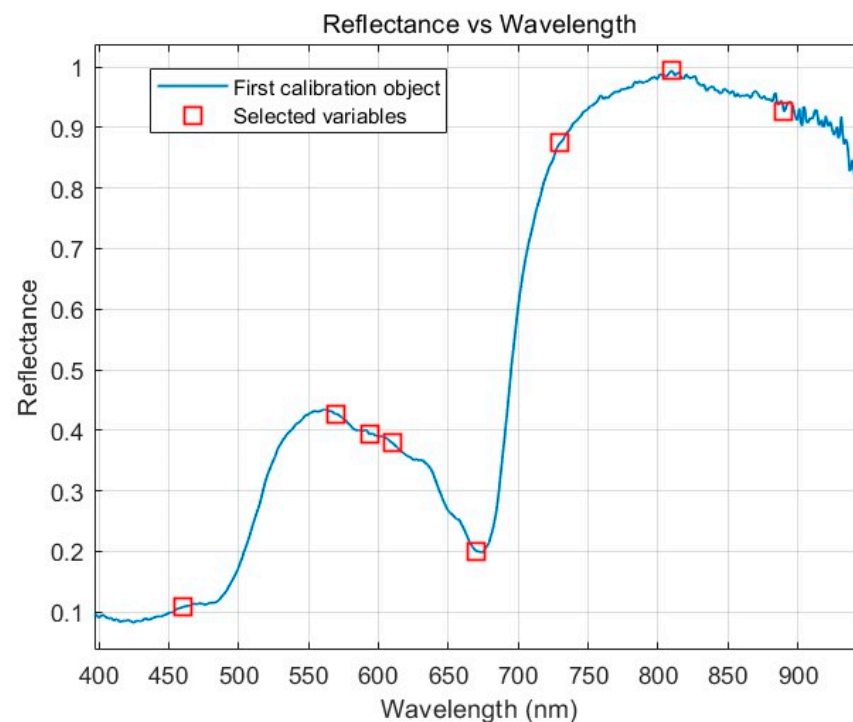


Figure 8. Successive projections algorithm to extract characteristic wavelength.

3.5. Effect of Different Spectral Preprocessing Methods on Raw Spectral Data

Due to the presence of noise in the raw reflectance spectra, preprocessing was performed using MSC, SNV, and normalization to enhance the accuracy and stability of the recognition model. Multiplicative scatter correction (MSC) can reduce sample differences by adjusting spectral data to minimize variations caused by particle size, shape, or different instrument settings. Additionally, MSC improves spectral quality by correcting distortions caused by multiple scattering, making the data more representative of true spectral signals. Standard normal variate (SNV) transformation can eliminate baseline drift and scale differences. SNV standardizes the spectral intensity of each sample, eliminating baseline drift and scale differences between samples. This standardization makes spectral data of the same substance more comparable across different environments or instruments. SNV also enhances the robustness of analysis, especially when dealing with data from various sources. As seen in Figure 9, the preprocessed spectral curves are smoother than the raw spectral curves, indicating that preprocessing has removed some noise and background interference.

3.6. Maturity Discrimination Model Based on Characteristic Wavelengths of Soluble Sugar Content

A BP neural network pear ripening discrimination model based on eight characteristic wavelengths of soluble sugar content was developed using the SPA method. The input vectors of the model were spectral features from 80 training samples at each ripening stage, of which 50 samples were used for testing and 30 samples for validation. The input layer has the same number of nodes as the number of input variables, with one hidden layer node and one output layer node. The transfer functions of the input and hidden layers are tangent sigmoid functions, and the transfer function of the output layer is linear. The learning method uses gradient descent with a momentum factor of 0.6, a training accuracy of 0.001, and a maximum of 5000 iterations. The specific results are shown in Table 3.

Table 3. Accuracy of BP model at different stages.

Discrimination Method	Maturity Stage	Recognition Rate			Comprehensive Accuracy
		Training Set Accuracy	Testing Set Accuracy	Validation Set Accuracy	
BP	Phase 1	1	0.95	0.9	0.95
	Phase 2	0.98	0.93	0.88	0.93
	Phase 3	1	0.96	0.98	0.98
	Phase 4	0.96	0.9	0.86	0.91

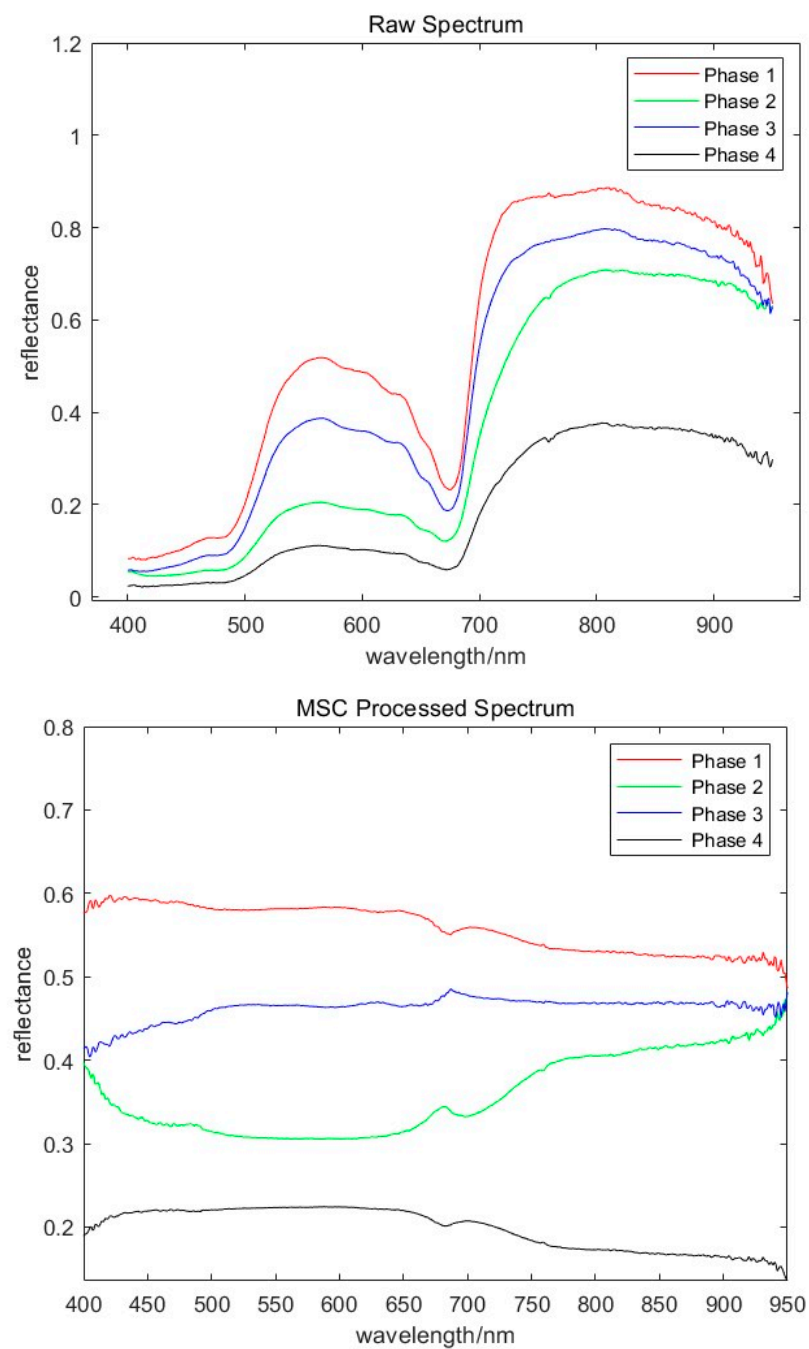


Figure 9. Cont.

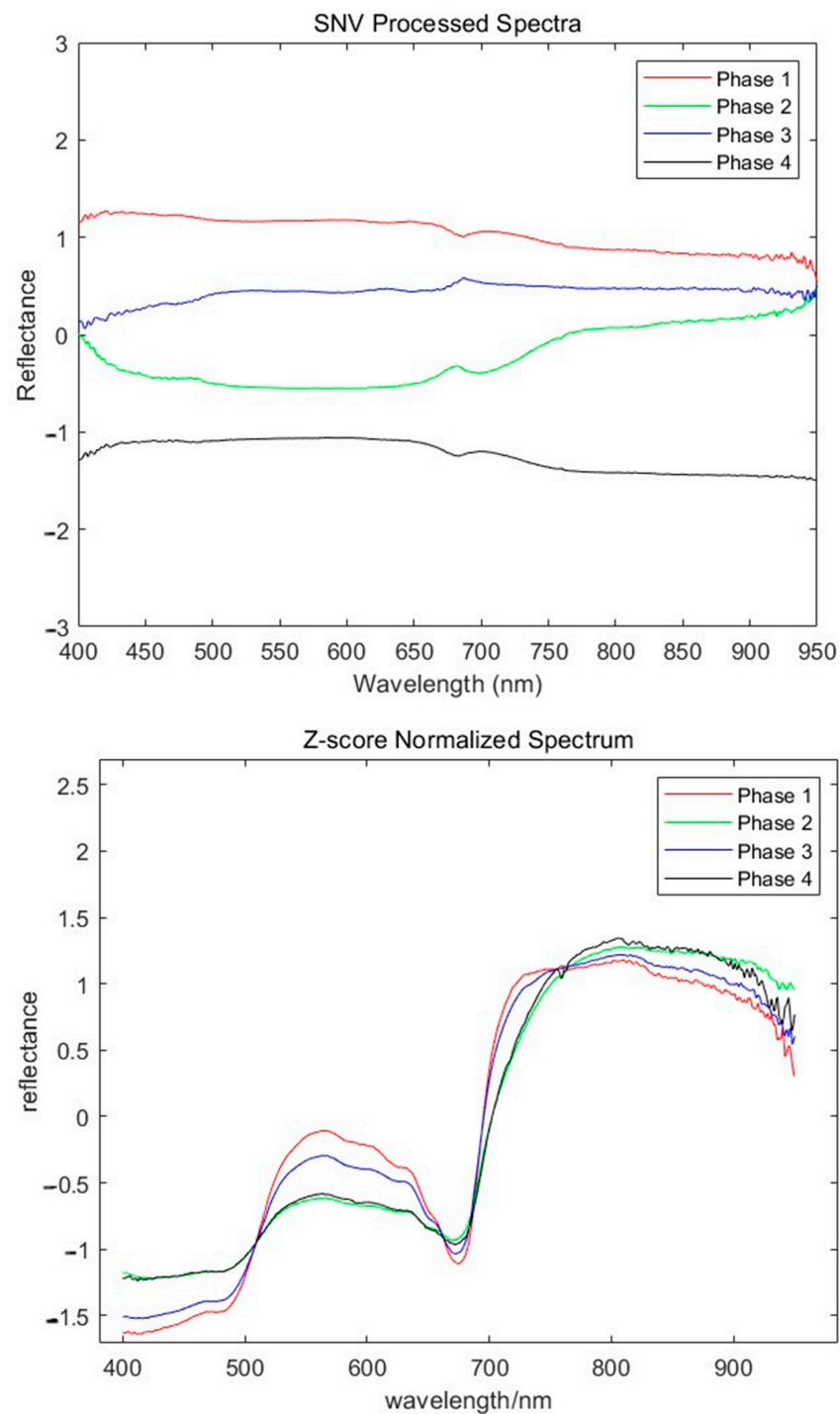


Figure 9. Effect of different spectral preprocessing methods on raw spectral data.

- First bul

The ripeness discrimination model of balsam pear established by BP based on the characteristic wavelength of sugar level is shown in the table. The integrated recognition rate of the training set is 98.5%, the integrated recognition rate of the test set is 93.5%, and the integrated recognition rate of the validation set is 90.5%. The integrated recognition accuracy for the four ripeness levels is 95%, 93%, 98%, and 91%, respectively. The prediction set of the BP model is shown in Figure 10.

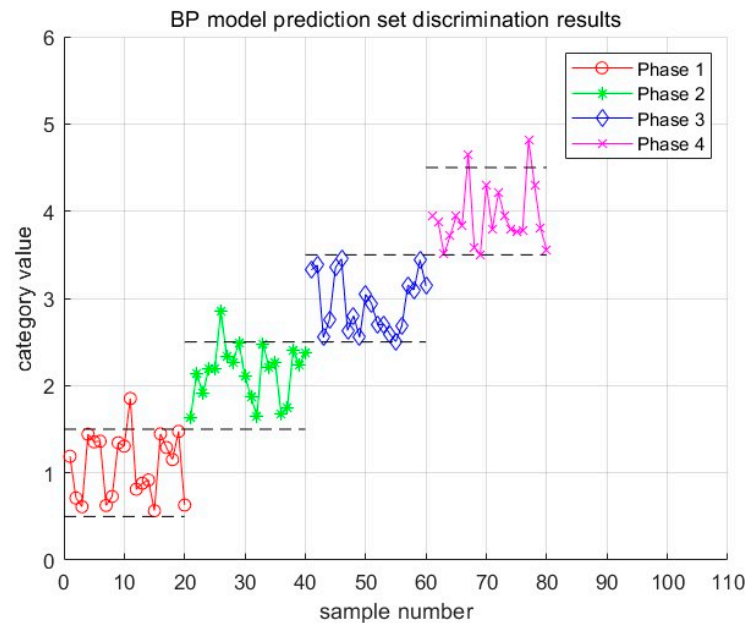


Figure 10. BP model prediction set discrimination results.

From Figure 10, it is evident that the BP model used for classifying the maturity of fragrant pears shows varying misclassification rates across different stages. Specifically, there was one misclassification in the first stage, one in the second stage, none in the third stage, and two in the fourth stage. The classification accuracy rates for the test samples at these four stages were 95%, 93%, 98% and 91%, respectively. These results indicate that the BP model, based on the characteristic wavelength related to sugar content, performs well in discriminating the maturity of fragrant pears. Therefore, it is feasible to use the characteristic wavelength of sugar content to determine the maturity of fragrant pears.

3.7. Recognition Results of Different Models for the Maturity of Fragrant Pears

As can be seen in Table 4, the CNN-S model performs well in all evaluation metrics, significantly outperforming the AlexNet and ResNet models. Specifically, CNN-S achieves an accuracy of 98% on the training set, and 92% and 96% on the test and validation sets, respectively, while maintaining a low loss rate on all datasets, showing its excellent learning ability and stability. In contrast, the AlexNet and ResNet models have lower accuracy on the test and validation sets; in particular, AlexNet's loss rate on the test set is as high as 35%, which may indicate its weak adaptability to unseen data. These data clearly demonstrate the advantages of the CNN-S model in terms of learning efficiency, generalization ability, and stability, highlighting its usefulness and efficiency in deep learning tasks, and providing strong evidence for future model optimization and application. To summarize, the comprehensive performance of the CNN-S model stands out in comparison, and its efficient network structure and training strategy performs superiorly in practical applications. In order to better visualize the results, the data is summarized as shown in Figure 11.

Table 4. Recognition results of balsam pear ripeness by different discriminant models.

Model	Training Set		Test Set		Validation Set	
	Training Set Accuracy	Training Set Loss	Test Set Accuracy	Test Set Loss	Validation Set Accuracy	Validation Set Loss
CNN-S	0.98	0.03	0.92	0.08	0.96	0.04
alexnet	0.93	0.07	0.85	0.35	0.88	0.25
resnet	0.96	0.04	0.89	0.30	0.91	0.18

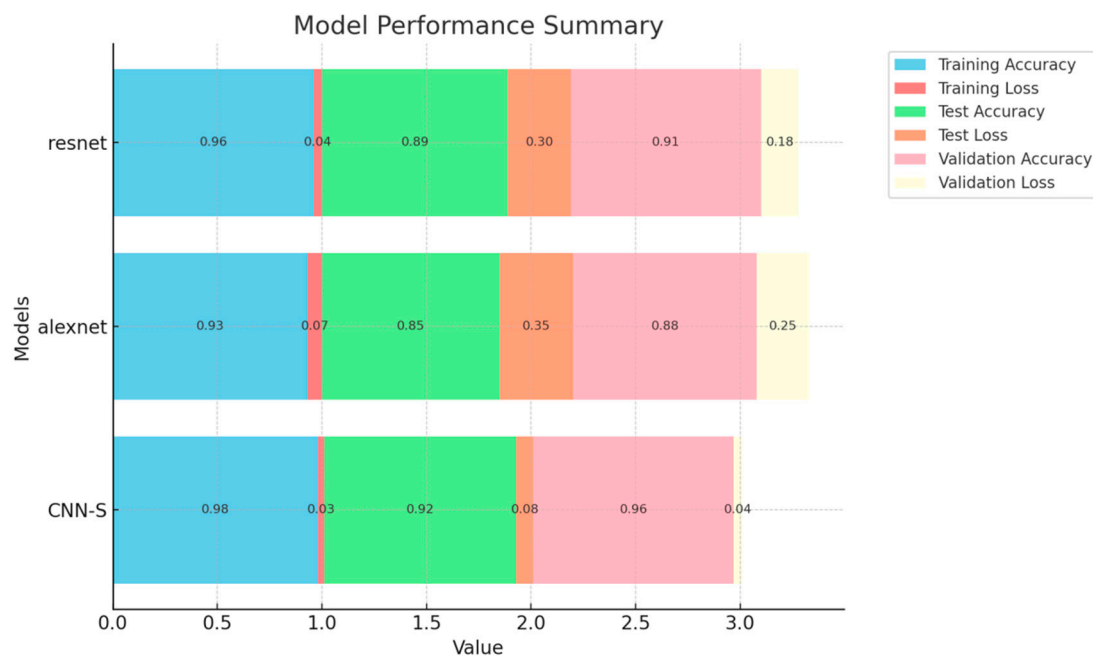


Figure 11. Model Performance Summary.

4. Discussion

The sample size of this study was relatively small, with only 150 samples of Korla balsam pear in each stage. This limitation may affect the generalizability and reliability of the results. In practical applications, the insufficient sample size may lead to poor performance of the model when dealing with unseen data, especially when dealing with pears of different varieties, different growing environments, or different picking times [21]. Therefore, future studies should consider increasing the sample size and including samples from different regions to improve the generalization ability and stability of the model. Obtaining more data from actual production by working with agricultural cooperatives and farms can enable the results to be validated and replicated in different regions and growing conditions.

In addition, the hyperspectral imaging technique and convolutional neural network model (CNN-S) used in this study were tested under laboratory conditions, which may be different from applications in real agricultural production environments. For example, factors such as lighting conditions, temperature variations, and external contamination of pears may affect the detection performance of the model. Therefore, further validation and optimization of the model is needed to ensure its robustness and reliability under different environmental conditions.

The experimental conditions were relatively homogeneous and mainly conducted in a controlled laboratory environment, which may have neglected the effects of different environmental factors (e.g., temperature, humidity) on pear ripeness detection. In real agricultural production, changes in environmental conditions can significantly affect fruit growth and ripening. Therefore, considering multivariate environmental conditions is crucial to improve the generalization ability of the model [22]. For example, how temperature and humidity variations affect the accuracy of hyperspectral imaging and neural network models is something that needs to be further explored.

In terms of feature selection, although a certain feature selection method was used in this study, future research could consider using other feature selection methods to improve the analysis. For example, common feature selection methods include principal component analysis (PCA), recursive feature elimination (RFE), and tree model-based importance assessment. Each of these methods has advantages and disadvantages:

Principal component analysis (PCA): PCA converts high-dimensional data into low-dimensional data through dimensionality reduction techniques, which can reduce data redundancy and improve the computational efficiency of the model. However, the results of PCA may be harder to interpret because principal components are linear combinations of original features.

Recursive feature elimination (RFE): RFE selects the optimal subset of features by recursively constructing the model and eliminating the least important features. RFE usually provides high accuracy but is computationally expensive, especially when dealing with a large number of features.

Importance assessment based on tree models such as random forest and gradient boosted tree: These methods perform feature selection by evaluating the contribution of each feature to the model performance. Tree modeling approaches typically strike a good balance between accuracy and computational efficiency and are easy to interpret.

Future research could compare the performance of these feature selection methods in terms of accuracy and computational efficiency to determine the best method that is most suitable for ripeness detection in Kuril balsam pear.

To overcome these limitations, future research efforts could be improved and extended in several ways. First, the sample size needs to be expanded by collecting more pear samples from different seasons, geographic locations, and varieties to improve the generalization ability of the model. This can be accomplished by working with agricultural cooperatives and farms to obtain more data from actual production. Second, multienvironmental testing is conducted to evaluate the model's performance under different environmental conditions (e.g., light intensity, temperature, humidity, etc.) to ensure its adaptability in real-world applications. Onsite tests can be conducted during the production and processing of pears to collect more actual data.

In addition, model optimization is key. The structure and parameters of the model are further optimized by introducing more advanced machine learning and deep learning techniques. For example, integrated learning methods or migration learning techniques can be used to significantly improve the accuracy and robustness of the model. Meanwhile, multi-indicator assessment should also be carried out, and more quality indicators (e.g., hardness, acidity, aroma, etc.) should be introduced for comprehensive assessment in addition to the characteristic wavelength of brix, in order to realize the comprehensive detection of pear quality.

Cross-regional studies are also very important. Studies need to be conducted in pear-growing areas in different regions to verify the applicability and stability of the model in different regions and to explore the feasibility of modeling in different regions. In addition, more environmental variables, such as temperature and humidity, should be added in subsequent studies. Experiments were conducted under different temperature and humidity conditions to assess the effects of these factors on hyperspectral imaging and neural network modeling. Specific measures include collecting samples under high-temperature, low-temperature, high-humidity, and low-humidity conditions to analyze the effects of temperature and humidity variations on pear spectral features and model accuracy. Data should also be collected under different light conditions (e.g., bright light and low light) to ensure the stability of the model under various light intensities.

To further improve the accuracy and robustness of the model, multivariate models can be introduced to optimize the model using multivariate machine learning methods by combining more environmental variables (e.g., soil composition, precipitation, etc.) for comprehensive analysis. For example, methods such as random forests and gradient boosted trees can be used to improve model performance. Factors such as temperature, humidity, and light are incorporated into the model through multivariate regression modeling or integrated learning methods to enhance the comprehensive assessment of pear ripeness.

Finally, field testing and validation are also essential steps. Field tests are conducted in real production environments to obtain measured data under different environmental conditions, and field experiments are conducted to verify the applicability and stability

of the model in real agricultural production [23]. There should be collaboration with agricultural cooperatives and farms to conduct long-term monitoring and data collection in different regions and environmental conditions to ensure the effectiveness of the model under various real conditions.

5. Conclusions

Analysis method similarity for Korla pear samples: This study utilized hyperspectral imaging technology to obtain data, which were significantly enhanced in quality after preprocessing with multiplicative scatter correction (MSC) and standard normal variate (SNV). The deep separable convolutional neural network model (CNN-S) established using these data outperformed traditional models in sample classification accuracy. It also demonstrated better generalization ability and lower loss rate, achieving an accuracy of 92% on the test set.

Model performance comparison: Experimental comparisons of different models on the same dataset demonstrated that the CNN-S model has significant advantages in handling high-dimensional, complex data. However, the limitations of this study include a small sample size and relatively uniform experimental conditions, which may restrict the broader applicability of the results. Future research could expand the sample size and include data from Korla pears grown in different environments and varieties to verify the model's generalization ability.

Implications for agricultural maturity detection: This finding supports the use of hyperspectral technology combined with advanced deep learning techniques, such as CNN-S, to achieve more precise and effective detection of agricultural product maturity. Future research could explore maturity detection for other types of agricultural products and the model's cross-regional generalization ability, providing more technological support for smart agriculture.

Author Contributions: Conceptualization, J.L. and H.M.; methodology, J.L. and H.M.; software, J.L.; validation, J.L. and H.M.; formal analysis, J.L.; investigation, H.M.; resources, H.M.; data curation, J.L.; writing—original draft preparation, J.L.; writing—review and editing, J.L. and H.M.; visualization, J.L.; supervision, H.M.; project administration, H.M.; funding acquisition, H.M. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National (Corps) Science and Technology Program Project Fund (Project No. 2021DB008).

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

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

References

1. Kong, D.; Gao, F.; Luo, H. Detection of minor damage in Korla pears based on hyperspectral transmission and reflection technology information fusion. *J. Tarim Univ.* **2023**, *35*, 77–83.
2. Zhao, D.; Li, A.; Guo, H.; Kang, L.; Ju, Y. Feasibility study on the identification of the origin of Korla pears using stable isotope technology. *J. Nucl. Agric. Sci.* **2020**, *34*, 37–42.
3. Chen, W. Discussion on the origin of Korla pears. *Xinjiang For.* **1999**, 37–38.
4. Zhao, D.; Ju, Y.; Ma, X.; Ding, C.; Zhang, W.; Zhao, D. Quality analysis and evaluation of Korla pears from Xinjiang. *J. Food Saf. Qual. Detect.* **2022**, *13*, 6637–6644. [\[CrossRef\]](#)
5. Liu, Y.; Wu, Y. Research progress on Korla pears. *Xinjiang Agric. Reclam. Technol.* **2015**, *38*, 23–26.
6. Le, S.; Wei, Q.; Lei, C. Comprehensive quality evaluation of Taoximi pomelo of different fruit sizes based on multivariate statistical methods. *J. Jiangxi Agric. Univ.* **2021**, *43*, 740–749. [\[CrossRef\]](#)
7. Lan, H.; Jia, F.; Tang, Y.; Zhang, Q.; Han, Y.; Liu, Y. Quantitative evaluation method for the maturity of Korla pears. *Trans. Chin. Soc. Agric. Eng.* **2015**, *31*, 325–330.
8. Lan, H.; Tang, Y.; An, J.; Liu, W.; Li, Y.; Li, F. Research on maturity evaluation method of Korla pears based on hardness and SSC. *J. Agric. Mech. Res.* **2013**, *35*, 193–196. [\[CrossRef\]](#)
9. Feng, Y.; He, J.; Cheng, Y.; Li, N.; Wang, J.; Guan, J. Effects of maturity on the quality of 'Hongxiangsu' pears during cold storage and shelf life. *J. Food Saf. Qual. Detect.* **2021**, *12*, 4513–4519. [\[CrossRef\]](#)

10. Sun, Y. Research on Disease Detection of Peach Fruit Based on Hyperspectral Imaging Technology. Ph.D. Thesis, Nanjing Agricultural University, Nanjing, China, 2018. [\[CrossRef\]](#)
11. Wang, C.; Wang, S.; He, X.; Dong, H. Study on beef breed identification method based on hyperspectral technology combined with image information. *Spectrosc. Spectr. Anal.* **2020**, *40*, 911–916.
12. He, H.-J.; Sun, D.-W. Hyperspectral imaging technology for rapid detection of various microbial contaminants in agricultural and food products. *Trends Food Sci. Technol.* **2015**, *46*, 99–109. [\[CrossRef\]](#)
13. Rajkumar, P.; Wang, N.; Elmasry, G.; Raghavan, G.S.V.; Gariepy, Y. Studies on banana fruit quality and maturity stages using hyperspectral imaging. *J. Food Eng.* **2020**, *108*, 194–200. [\[CrossRef\]](#)
14. Zhang, C.; Guo, C.; Liu, F.; Kong, W.; He, Y.; Lou, B. Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine. *Food Eng.* **2016**, *179*, 11. [\[CrossRef\]](#)
15. Ribera-Fonseca, A.; Noferini, M.; Jorquera-Fontena, E.; Rombolà, A.D. Assessment of technological maturity parameters and anthocyanins in berries of cv. Sangiovese (*Vitis vinifera* L.) by a portable vis/NIR device. *Sci. Hortic.* **2016**, *209*, 229–235. [\[CrossRef\]](#)
16. Munera, S.; Amigo, J.M.; Blasco, J.; Cubero, S.; Talens, P.; Aleixos, N. Ripeness monitoring of two cultivars of nectarine using VIS-NIR hyperspectral reflectance imaging. *J. Food Eng.* **2017**, *214*, 29–39. [\[CrossRef\]](#)
17. 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. [\[CrossRef\]](#)
18. Sun, J.; Ma, B.; Dong, J.; Yang, J.; Xu, J.; Wei, J.; Gao, Z. Study on the maturity discrimination of Hami melon using hyperspectral technology combined with feature wavelength selection and support vector machine. *Spectrosc. Spectr. Anal.* **2017**, *37*, 2184–2191.
19. Shao, Y.; Wang, Y.; Xuan, G.; Gao, C.; Wang, K. Visualization analysis and maturity detection of Feicheng peach quality based on hyperspectral imaging. *Trans. Chin. Soc. Agric. Mach.* **2020**, *51*, 344–350.
20. Jiang, H.; Zhang, C.; Liu, F.; Zhu, H.; He, Y. Strawberry maturity identification based on multi-spectral parameters of hyperspectral images. *Spectrosc. Spectr. Anal.* **2016**, *36*, 1423–1427. [\[CrossRef\]](#)
21. Zhao, J. Impact of Sample Size on Machine Learning Model Performance in Agriculture. *Precis. Agric.* **2019**, *20*, 1057–1074.
22. Zhang, C. Hyperspectral imaging and their applications in the nondestructive quality assessment of fruits and vegetables. *Appl. Spectrosc. Rev.* **2018**, *53*, 475–497.
23. Shafiee, S. Field-based hyperspectral imaging for predicting fruit quality. *Comput. Electron. Agric.* **2017**, *137*, 193–201.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.