agronomy

MDPI

Article

Prediction of Strawberry Quality during Maturity Based on Hyperspectral Technology

Li Fan [1], Jiacheng Yu [1], Peng Zhang [2] and Min Xie [2,*]

[1] College of Horticulture & Plant Protection, Inner Mongolia Agricultural University, Huhhot 010010, China; fanli@imau.edu.cn (L.F.); 18748296027@163.com (J.Y.)
[2] College of Agriculture, Inner Mongolia Agricultural University, Huhhot 010010, China; 2024zhangp@gmail.com
* Correspondence: xiemin@imau.edu.cn

Abstract: In a study aimed at developing a rapid and nondestructive method for testing the quality of strawberries, spectral data from four strawberry varieties at different ripening stages were collected using a geophysical spectrometer, primarily focusing on the 350–1800 nm band. The spectra were preprocessed using Savitzky–Golay (SG) filtering, and characteristic bands were extracted using Pearson correlation coefficient (PCC) analysis. Models for predicting strawberry quality were built using random forest (RF), support vector machine (SVM), partial least squares (PLS), and Gaussian regression (GPR). The results indicated that the SVM model exhibited relatively high accuracy in predicting anthocyanin, hardness, and soluble solids content in strawberries. For the test set, the SVM model achieved R^2 and RMSE values of 0.81, 0.87, and 0.89, and 0.04 mg/g, 0.33 kg/cm^2, and 0.72%, respectively. Additionally, the PLS model demonstrated relatively high accuracy in predicting the titratable acid content of strawberries, achieving R^2 and RMSE values of 0.85 and 0.03%, respectively, for the test set. These findings provided a solid foundation for strawberry quality modeling and a veritable guide for non-destructive assessment of strawberry quality.

Keywords: hyperspectral; strawberry quality; non-destructive monitoring

check for updates

Citation: Fan, L.; Yu, J.; Zhang, P.; Xie, M. Prediction of Strawberry Quality during Maturity Based on Hyperspectral Technology. *Agronomy* **2024**, *14*, 1450. https://doi.org/10.3390/agronomy14071450

Academic Editors: Ana Elisa Rato and Julio Nogales-Bueno

Received: 19 May 2024
Revised: 28 June 2024
Accepted: 1 July 2024
Published: 4 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

Strawberry (*Fragaria* × *ananassa* Duch.) is a small perennial berry of the Rosaceae family. It is rich in nutrients, including organic acids, protein, pectin, vitamin C, and phenolics [1,2]. Strawberries are distinctive for their sweet taste and vibrant colors and are used both for fresh food and for processing [3,4]. According to the Food and Agriculture Organization of the United Nations (FAO), global strawberry production was over 9.56 million tons in 2022 [5]. China is the world's largest producer of strawberries, accounting for one-third of the world's production and more than three times that of the United States, the world's second-largest strawberry producer [6].

Key factors influencing the quality of strawberry fruit encompass aspects such as color, texture, and the delicate balance between sweetness and acidity [7]. Among the predominant soluble solids found in strawberries are the total soluble sugars. These are low molecular weight carbohydrates, playing a crucial role in imparting sweetness, flavor, aroma, and enhancing the fruit's visual appeal [8,9]. Additionally, the firmness of strawberries emerges as a significant quality indicator. Detectable through touch—whether in the act of picking the fruit or chewing it—the firmness directly influences both the taste experience and the fruit's longevity in storage [10].

Anthocyanin plays a pivotal role in influencing the color and nutritional profile of strawberries and is one of the flavonoid pigments commonly present in plants [11]. This naturally occurring water-soluble pigment imparts vivid hues to plants, serving to attract insects for pollination and seed dispersal, while also providing protection against diverse biotic and abiotic stresses [12]. In human applications, anthocyanin pigments are utilized as

natural food colorants, recognized for their potential preventive effects against a spectrum of chronic diseases [13–16].

At present, strawberry quality detection mainly relies on traditional chemical methods, which are time-consuming and destructive detection techniques. Thus, there is less research conducted in nondestructive testing [17,18], necessitating a rapid nondestructive testing technique. Hyperspectral technology emerges as a cutting-edge, non-destructive, multi-band spectral processing technology, widely used for rapid non-destructive test of agricultural product quality and safety [19–21]. For example, Cho J. S et al. utilized hyperspectral imaging (HSI) to detect anthocyanin content in strawberries of different maturity levels and established a prediction model for anthocyanins using partial least squares (PLS) regression [22]. Siedliska A et al. collected spectral reflectance of strawberries using hyperspectral technology, analyzed the substances including soluble solids content (SSC), and established a characteristic wavelength prediction model, which demonstrated the potential of hyperspectral techniques for rapid and non-destructive detection of soluble solids content (SSC) in strawberry fruits [23]. Chen S et al. determined anthocyanin content in grapes using hyperspectral techniques, demonstrating the reliability and feasibility of using near-infrared hyperspectral technology for the prediction of anthocyanins in wine grapes in the spectral range of 900–1700 nm [24].

Preprocessing of hyperspectral data is essential because raw hyperspectral data may be affected by a variety of factors that may interfere with or confuse the results of the analyses. SG smoothing is a commonly used method for preprocessing hyperspectral data, and its full name is the Savitzky–Golay smoothing strategy, which was proposed by Abraham Savitzky and Marcel E. Golay in 1964 [25]. It works well to eliminate noise, improve signal quality and enhance features. For instance, Mo C et al. developed a partial least squares regression (PLSR) model for soluble solids content (SSC) mapping within apples by hyperspectral imaging, and the results showed higher accuracy and fewer errors for the model with SG smoothing [26].

Regarding model selection, Rahul Raj et al. collected hyperspectral reflectance of strawberries in the range of 350–2500 nm using a calibrated spectroradiometer. They developed models such as support vector machine (SVM), decision tree (DT), and multilayer perceptron (MLP) to predict the ripeness of strawberries, and the results showed that the ripeness of strawberries could be detected quickly and non-destructively using support vector machine (SVM) modeling [27]. Xie C et al. used hyperspectral imaging to determine metrics such as hardness of bananas, and they developed a partial least squares (PLS) model to predict the hardness of bananas. The results showed that hyperspectral imaging could be used for non-destructive testing of banana hardness [28]. Xu M et al. used hyperspectral techniques to detect SSC content. Three studies used PLSR, SVM, and PLSR, respectively. All three studies achieved good results, the best being the SVM model [29]. Fan S et al. used hyperspectral imaging to determine the soluble solids content (SSC) and hardness of pears. They developed a calibrated model using partial least squares (PLS) for the analysis. The results showed that the assay using hyperspectral technique was a rapid and promising method for the determination of SSC and hardness of pears [30].

The aim of this study was to assess the accuracy of the method of extracting spectral feature bands using principal correlation analysis for the detection of strawberry quality. The spectral data were collected and then subjected to SG smoothing. Strawberry anthocyanin, hardness, soluble solids and titratable acid content were modeled and predicted by feature extraction. Four different estimation models, namely random forest (RF), support vector machine (SVM), partial least squares (PLS) and Gaussian process regression (GPR) were tested. The performance of these four models was compared and evaluated, from which the best model for predicting strawberry quality was selected. The results could provide a theoretical basis for non-destructive testing of strawberry quality in Inner Mongolia.

2. Materials and Methods

2.1. Research Materials and Research Sites

The experiment was conducted in January 2024 in Saihan District, Hohhot City, Inner Mongolia. The strawberries of four varieties, including Fu-kunoka Syougo, Ganlu, Miao Xiang 7 and Ssanta, were grown in the greenhouses. The cultivation environment in the greenhouses were maintained the same temperature, moisture, light period and soil type. Sampling started on 9 January 2024 and lasted one month.

2.2. Strawberry Sample Collection and Processing

Samples of four strawberry varieties were collected at three stages of strawberry ripening. During sampling activities, 20 samples of similar shape and size were selected at the strawberry ripening stages of color change (s1), semi-red (s2), and full red (s3) without any disturbing factors such as breakage, decay, or surface foreign matter. A total of 60 samples of each strawberry variety were selected and transported back to the laboratory in refrigerated boxes. All samples were individually cleaned, wiped, and numbered after picking, and then stored at 17–19 °C to ensure that the final strawberries used for the experiments were in optimal condition. A total of 240 samples were collected during the sampling events.

2.3. Hyperspectral Data

Strawberry near-end spectral reflectance was collected using an SVC HR-1024i spectrometer (HR-1024, SVC, Manufactured by Sloan Valve Company (svc), located in Franklin Industrial Park near Chicago, IL, USA). In the present study, we performed the measurements under clear, windless, dry air conditions. Before collecting spectral data, the instrument was calibrated by scanning a white board. A black absorbent cloth was placed under the sample to set the appropriate distance, and then the probe was pointed at the strawberry sample facing the sun to start the measurement. Care was taken not to let shadows cover the sample during the measurement to avoid affecting the results. After scanning the sample, the spectral data were resampled and data were cleaned for later analysis.

2.4. Measurement of Strawberry Indicators

2.4.1. Anthocyanin Content

A total of 2.0 g of strawberry pulp was weighed into the mortar. Then, 2 mL of 4 °C pre-cooled 1% HCL-methanol solution was added quickly, and the strawberry samples were ground in an ice bath. The resulting solution was transferred to a 25 mL volumetric flask, and the mortar was rinsed with 1% HCL-methanol solution. The mixture was diluted to 25 mL with 1% HCL-methanol solution, then extracted at 4 °C, avoiding light for 20 min, during which it was shaken several times. Afterward, the mixture was filtered. The 1% HCL-methanol solution was used as a blank control, and the absorbance values of the filtrate were determined at 600 nm and 530 nm, respectively, and repeated three times. The anthocyanin content was expressed as the difference between the absorbance values at 530 nm and 600 nm for each gram of fresh strawberry tissue. That is, $U = (OD530 - OD600)/g$ [31].

2.4.2. Strawberry Hardness

The GY-1 hardness tester (GY-1, manufactured by Top Instruments, located in West Park, West Lake Science and Technology Park, Hangzhou, Zhejiang Province, China) probe was pressed vertically onto the strawberry. When the probe reached the 10 mm mark inside the strawberry, it was stopped, and the value of the outer circle was read and recorded as the equatorial hardness value of the entire strawberry. After each measurement, the pointer was reset to zero, and the next measurement was conducted [32].

2.4.3. Soluble Solids Content

Strawberry soluble solids content was determined using a WYT-32 hand-held refractometer (WYT-32 hand-held refractometer, manufactured by Quanzhou Optical Instrument Factory, located at Xinhua West Road, Quanzhou, Fujian Province, China), where the strawberry's extruded homogenate was dripped into the measuring trough to measure its soluble solids content [33].

2.4.4. Titratable Acids

In all, 2.5 g of strawberries were weighed, ground in a mortar, and transferred to a 25 mL volumetric flask. The ground strawberry samples were diluted to 25 mL with distilled water and centrifuged at 4000 r for 10 min. A total of 5 mL of the supernatant was taken, 1% phenolphthalein indicator was added, and titration was carried out with sodium hydroxide. Distilled water was used as a control [34].

$$X = V \times c \times (V1 - V0) \times f \times Vs \times m \times 100\% \quad (1)$$

In this equation, X represents the titratable acid content, V represents the total volume of sample extract (mL), c represents the titration concentration of NaOH (mol/L), V1 represents the volume of NaOH solution consumed in the titration of the filtrate (mL), V0 represents the volume of NaOH solution consumed in the titration of the distilled water (mL), f represents the conversion factor (g/mmol), vs. represents the volume of filtrate taken during the titration (mL), and m represents the mass of the sample (g).

2.5. Data Analysis Methods

2.5.1. Hyperspectral Preprocessing, Feature Band Extraction and Dataset Partitioning

In order to improve the accuracy of the model, data preprocessing was performed on the raw hyperspectral reflectance data. In this experiment, a SavGol filter with a window size of 5 and a polynomial order of 2 was used as a preprocessing step before analysis. This filter reduced noise and variability in the dataset, improving the predictive power and accuracy of the model (the deep learning framework TensorFlow 2.1 for Python (version 3.7.16) was used in this study. The computer was equipped with a GeForce GTX 4060 graphics card with 16 GB of graphics memory and an Intel(R) Core(TM) i7-12650H processor at 2.3 GHz).

Feature band extraction was utilized as a technique to eliminate redundant spectral information, reduce the likelihood of model overfitting, and enhance the speed of model runs. In this study, the Pearson correlation coefficient (PCC) method was employed for data dimensionality reduction. The PCC method reflected the strength of the linear relationship between two variables and filtered out the feature bands.

Regarding dataset partitioning, the combined dataset was divided into two parts, namely the training set and the test set, with the training data accounting for 80% and the test data accounting for 20%. This classification method ensured a large training dataset for model development and critical evaluation of the model.

2.5.2. Model Establishment

Random forest (RF) is designed for small datasets. The RF method is known for its fast-training speed, generalization ability, and robustness. Large-scale datasets and high-dimensional data are two of its areas of strength [35].

The SVM machine learning algorithm is based on the principle of structural risk minimization. It achieves strong generalization ability and reduces the complexity of the learning machine while maintaining the training accuracy. SVM is commonly used in regression problems because it could effectively solve problems with few samples, nonlinearities, and high dimensionality [36]. The SVM kernel function adopted a radial basis function kernel. The cross-validation method considered the parameters δ (variance

of the RBF kernel function) and C (penalty factor) and was a method to determine the optimal value. The modeling parameters in this study were C = 2 and δ = 0.7.

The PLS method helped to refine important spectral information and was particularly suitable for inverse modeling of datasets with a limited number of samples. The PLS method combined the advantages of multiple linear regression, typical correlation analysis, and principal component analysis. It took into account the ability of the independent variables to explain the dependent variable and provided a linear regression model [37].

GPR is a nonparametric model for regression analysis of data using the Gaussian process. Based on the convenience of the Gaussian process and its kernel function, GPR is usually used for low dimensionality and small sample regression problems [38].

2.5.3. Model Performance

A prediction model was developed with strawberry quality content as the dependent variable and the coefficient of determination (R^2) and root mean square error (RMSE) were used to evaluate the model. The larger the R^2 is, the closer it is to 1. This means that the model is more accurate, and the smaller the RMSE is, the closer it is to 0, indicating that the model is more accurate. The formulas for these two evaluation coefficients are shown in Equations (2) and (3) [39].

$$R^2 = \frac{\sum(\hat{y}_i - \overline{y})^2}{\sum(y_i - y)^2} \quad (2)$$

$$RMSE = \frac{\sqrt{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}}{\dot{n}} \quad (3)$$

where y_i is the actual value; $\hat{y}_i$ is the estimated value; y is the mean actual value of the sample; and n is the number of samples.

3. Results

3.1. Strawberry Quality Content Analysis

As seen in Figure 1, anthocyanin content and soluble solids content increased with maturity, with s1 being the lowest and s3 the highest. Hardness and titratable acid content decreased with increasing ripening stage, with s3 being the lowest and s1 the highest.

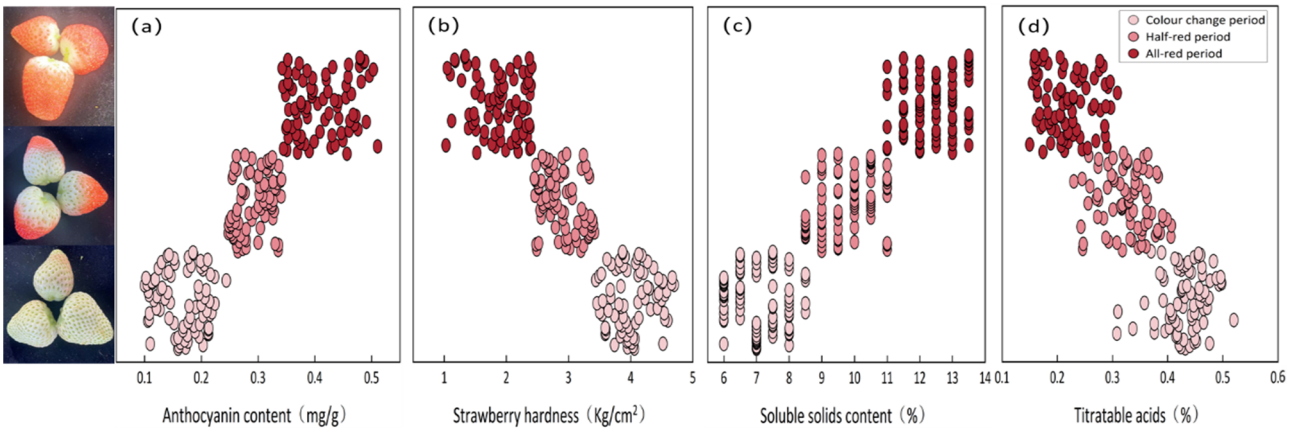


Figure 1. Distribution of strawberry anthocyanins (**a**), hardness (**b**), soluble solids (**c**) and titratable acids (**d**) at different stages of ripening.

Analysis of the quality content data of harvested strawberries (Table 1) revealed that anthocyanin content ranged from 0.10 to 0.51, hardness from 1.03 to 4.70, soluble solids from 6.0 to 13.5, and titratable acid from 0.15 to 0.52.

Table 1. Statistical analysis of quality content of strawberries at different stages of ripening.

Strawberry Quality	Mature Stage	Maximum	Minimum	Mean	Standard Deviation
Anthocyanin content (mg/g)	Color change period	0.51	0.34	0.41	0.04
	Half-red period	0.34	0.24	0.29	0.02
	All-red period	0.24	0.10	0.16	0.03
Strawberry hardness (kg/cm^2)	Color change period	2.41	1.03	1.87	0.38
	Half-red period	3.42	2.45	2.90	0.29
	All-red period	4.70	3.44	4.02	0.33
Soluble solids content (%)	Color change period	13.5	11.0	12.39	0.78
	Half-red period	11.0	8.5	9.73	0.80
	All-red period	8.5	6.0	7.09	0.78
Titratable acids (%)	Color change period	0.30	0.15	0.21	0.04
	Half-red period	0.40	0.23	0.32	0.04
	All-red period	0.52	0.30	0.43	0.03

3.2. Hyperspectral Data Preprocessing Analysis

The hyperspectral reflectance of strawberries is shown in Figure 2. As the spectral data above 1800 nm were noisy, the data between 350 and 1800 nm were selected.

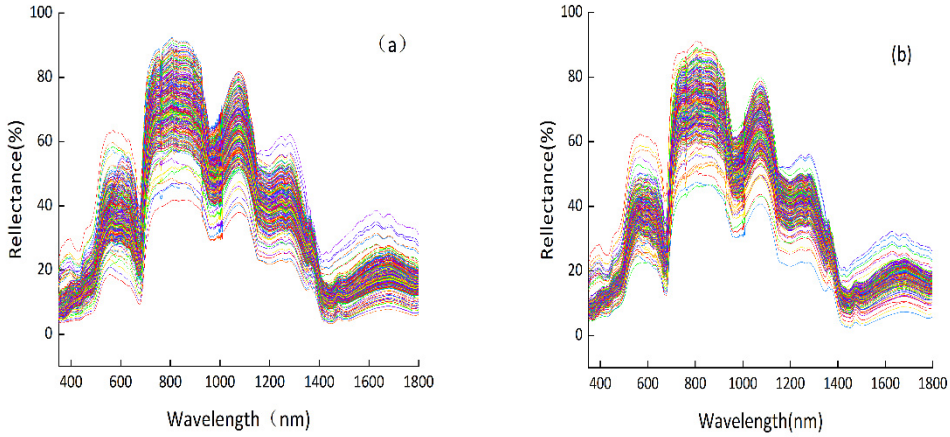


Figure 2. Raw spectral reflectance image (**a**) and SG preprocessed image (**b**).

3.3. Hyperspectral Data Feature Extraction

The Pearson correlation coefficient (PCC) method was used to extract the feature bands. The spectral bands that were highly correlated with strawberry quality were extracted from the preprocessed data. Among them, wavelengths with a correlation greater than 0.5 and that ranked in the top 20 were used instead of the original bands. This approach reduced the complexity of the model and shortened the modeling time. Table 2 showed the results of feature band extraction.

Table 2. Extracted characteristic wavelengths.

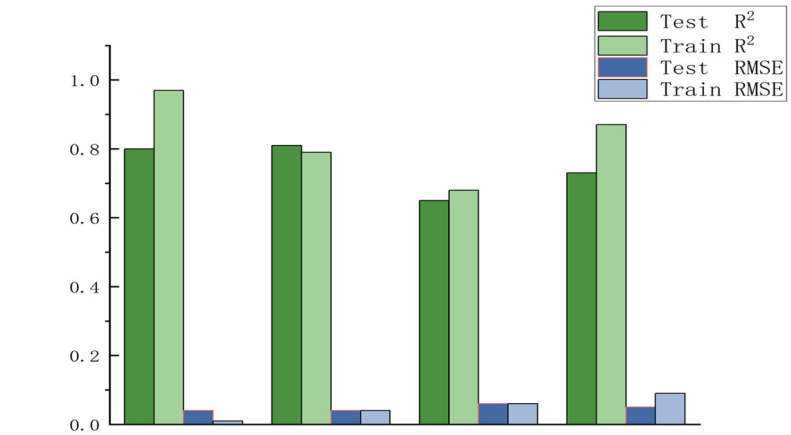
Strawberry Indicators	Preprocessing Method	Number of Feature Bands	Maximum Correlation Coefficient
Anthocyanin content	SG	20	0.77
Strawberry hardness	SG	20	0.78
Soluble solids content	SG	20	0.78
Titratable acids	SG	20	0.57

3.4. Performance of Strawberry Quality Prediction Models

In this study, we compared the performance of four strawberry quality prediction models: random forest (RF), support vector machine (SVM), partial least squares (PLS), and Gaussian process regression (GPR). The SG smoothing preprocessing method and the Pearson's correlation coefficient (PCC) method were also used for feature extraction. To improve the accuracy of the model, the dataset was divided into 192 training sets and 48 test sets in an 8:2 ratio. The model evaluation coefficients included R^2 and RMSE, and by comparing the values of R^2 and RMSE, the best inversion results for each model were listed in the table.

3.4.1. Anthocyanins Model Prediction Results

The prediction results of anthocyanins by each model were shown in Figure 3, in which the R^2 of the training set of RF, SVM, and GPR were all greater than 0.8. The main comparison of the R^2 of the test set showed that the R^2 of the spectral test set based on SVM was 0.81, while the R^2 of the spectral test set of RF and GPR were 0.80 and 0.73, respectively, which were smaller than that of SVM. Additionally, the RMSE of the spectral test set of SVM was comparatively small among all the models. It could be seen that, in the strawberry anthocyanin content monitoring model, the performance of the strawberry anthocyanin content monitoring model constructed using the SVM method was relatively stable and accurate.

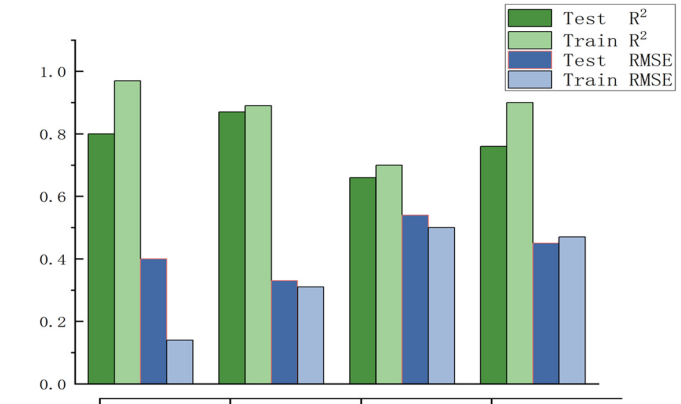


	RF	SVM	PLS	GPR
Test R^2	0.80	0.81	0.65	0.73
Train R^2	0.97	0.80	0.68	0.87
Test RMSE	0.04	0.04	0.06	0.05
Train RMSE	0.01	0.04	0.06	0.09

Figure 3. Model prediction of strawberry anthocyanins.

3.4.2. Hardness Model Prediction Results

The prediction results of each model for hardness were shown in Figure 4, where the R^2 of the spectral training set for RF, SVM, and GPR were all greater than 0.8, and their R^2 values were not significantly different. Thus, upon comparing the R^2 of the test sets, it was observed that the R^2 of the spectral test set of SVM was 0.87, while the R^2 of the spectral test set of RF and GPR were 0.80 and 0.76, respectively, which were smaller than the R^2 of the spectral test set of SVM. Additionally, the RMSE of the spectral test set of SVM was smaller than that of the other models. Therefore, in the strawberry hardness detection model, the strawberry hardness monitoring model constructed using SVM exhibited more stable and accurate results.

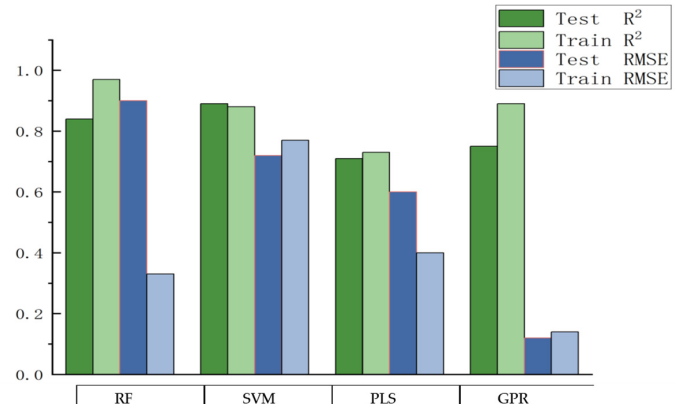


	RF	SVM	PLS	GPR
Test R^2	0.80	0.87	0.66	0.76
Train R^2	0.97	0.89	0.70	0.90
Test RMSE	0.40	0.33	0.54	0.45
Train RMSE	0.14	0.31	0.50	0.47

Figure 4. Results of model prediction of strawberry hardness.

3.4.3. SCC Model Prediction Results

The prediction results of soluble solids by each model were shown in Figure 5, where the R^2 of the spectral training set of RF, SVM, and GPR were all above 0.8. Upon comparing the test set R^2, it was found that the R^2 of the spectral test set of SVM was 0.89, while the R^2 of the test set of RF and GPR were 0.84 and 0.75, respectively, all of which were smaller than that of SVM. Additionally, the test set RMSE values for SVM are less than the spectral RMSE values for RF and GPR. Therefore, the performance of the strawberry soluble solids model constructed using the SVM method was more stable and accurate in the soluble solids detection model.

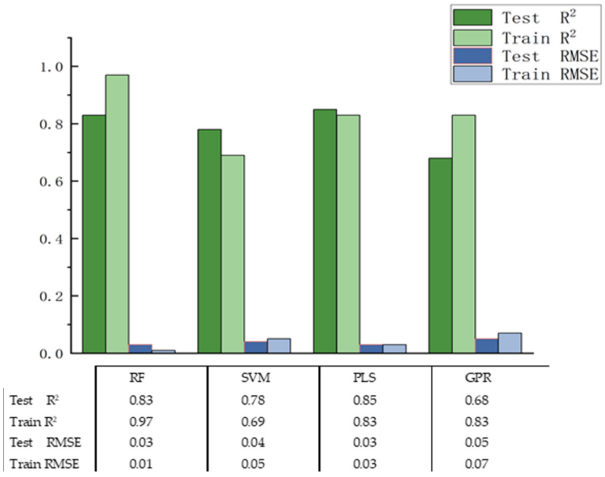


	RF	SVM	PLS	GPR
Test R^2	0.84	0.89	0.71	0.75
Train R^2	0.97	0.88	0.73	0.89
Test RMSE	0.90	0.72	0.6	0.12
Train RMSE	0.33	0.77	0.4	0.14

Figure 5. Model prediction of strawberry soluble solids content.

3.4.4. Titratable Acids Model Prediction Results

The prediction results of each model for titratable acid were shown in Figure 6, where the spectral training set R^2 of RF, PLS, and GPR were all greater than 0.8. Primarily, the comparison of the test set R^2 was carried out, revealing that the PLS spectral test set R^2 was 0.85, while the spectral test set R^2 of RF and GPR were 0.83 and 0.68, respectively. Additionally, the RMSE of the PLS spectral test set was comparatively small among all the models. Therefore, the performance of the strawberry titratable acid model constructed using the PLS method was relatively stable and accurate in the detection model of titratable acid.



	RF	SVM	PLS	GPR
Test R^2	0.83	0.78	0.85	0.68
Train R^2	0.97	0.69	0.83	0.83
Test RMSE	0.03	0.04	0.03	0.05
Train RMSE	0.01	0.05	0.03	0.07

Figure 6. Strawberry titratable acid model predictions.

3.5. Model Confirmation

Through the above analysis, the optimal solutions of the models for each index of strawberry were determined. As seen from Figure 7, the optimal model for anthocyanin, hardness, and soluble solids content were all SVM models, with the R^2 of the spectral test set being 0.81, 0.89, and 0.89, respectively. From Figure 6, it was evident that among all the prediction models for titratable acid, the accuracy of the SVM model in predicting titratable acid was satisfactory. However, the optimal model for titratable acid was the PLS model, with an R^2 of 0.85 for the spectral test set. Therefore, choosing the SVM model for predicting anthocyanins, hardness, and soluble solids of strawberries, and selecting the PLS model for predicting titratable acid of strawberries was the best choice for predicting the quality of strawberries.

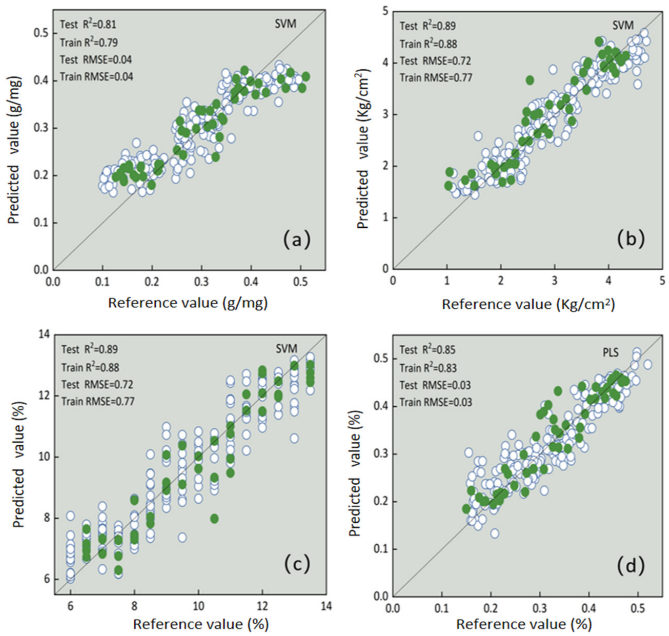


Figure 7. Validation results of the strawberry quality content regression model. Each fitting result is for white dots indicating training set data and green dots indicating test set data. The strength of the model can be summarized based on its deviation from the standard line. These plots show the R2 and RMSE. The best choices for strawberry quality modeling were (**a**) the anthocyanin prediction model by SVM method, (**b**) the hardness prediction model by SVM method, (**c**) the soluble solids predictive model by SVM method, and (**d**) the titratable acid prediction model by PLS method.

4. Discussion

In this study, spectral data obtained from raw absorbance spectral data were analyzed by PLS, SVM, RF, and GPR. The SG smoothing method was used to preprocess the spectral data, and the PCC method was used for feature extraction. The best predictive model for strawberry quality analysis was finally obtained.

Preprocessing the spectra strengthened the predictive ability of the model. It was shown that spectral data were processed using SG smoothing and modeled based on the processed data, which improved the accuracy of the model [40]. Guo C [41] also argued that spectral preprocessing improved the prediction of fruit quality. Feature band extraction was also used to help improve model prediction. In this study, PCC was used for data downscaling and feature band extraction to improve the speed of the model run, consistent with the findings of Ouyang [42]. In the prediction of anthocyanins in blueberries, Liu M et al. [43] found that the SVM model was more suitable for small sample determination and had the smallest RMSE value, consistent with this study. Basak J K et al. predicted the total soluble solids (TSS) content of strawberries using machine learning, and the results showed the optimal performance of the SVM model, with a test set R^2 of 0.84 obtained, in agreement with the results obtained in this experiment [44]. Sarkar S [45] predicted the soluble solids of kiwifruit using spectral data, and in most cases, the SVM algorithm yielded better model predictions. In this experiment, 4 models were combined with 4 strawberry indicators for a total of 16 models. Among them, the SVM model was better in predicting strawberry anthocyanins, hardness, and soluble solids, while the PLS model was better in predicting strawberry titratable acid content. It is worth mentioning that the SVM model also performed well in predicting titratable acid, which was second only to the PLS model. This may be due to the fact that the SVM model based on SG spectral preprocessing is more suitable for data at high latitude and with small samples. The model provided a theoretical basis for predicting strawberry quality content using small sample data.

It was worth noting that although four strawberry varieties were selected for this experiment, strawberry quality content measurements and spectra were only collected for one year, limiting the applicability of the model. Future studies should have increased the number of years of indicator collection as a way to increase the generality of the model. In addition, only one preprocessing method, SG smoothing, was used in this study for spectral preprocessing, which did not allow for better model comparisons. It made sense to explore various preprocessing methods to improve the accuracy of the model in subsequent studies. Generally, increasing the sample size was a way to improve the model's accuracy and applicability. Therefore, in future research, expanding the sample size to train the model was essential to enhance its predictive ability and stability.

5. Conclusions

In this study, we conducted a complete workflow for predicting strawberry quality based on hyperspectral. We performed 240 strawberry samples (4 varieties and 3 periods), and after sampling, we first acquired hyperspectral data and strawberry quality, where we preprocessed the spectral data using SG smoothing. The feature bands were extracted using the PCC method for data reduction in order to increase the speed of the model run. Then, the dataset division was performed to divide the dataset into two parts, i.e., the training set and the test set, where the training data accounted for 80% and the test data accounted for 20%. Strawberry quality was also predicted and compared by four modeling methods to determine the best model. The results showed that a comparison of the four modeling methods revealed that using the SVM model was the best predictive model for strawberry anthocyanins, hardness, and soluble solids, with R^2 and RMSE values of 0.81, 0.87, 0.89, and 0.04 mg/g, 0.33 kg/cm^2, 0.72%, respectively. Prediction of strawberry titratable acids using the PLS model had the highest accuracy and good robustness. Its R^2 and RMSE values were 0.85 and 0.03%. The above results show that it is feasible to detect the quality content of strawberries using hyperspectral technology, which provides a theoretical basis for non-destructive testing of strawberry quality in western Inner Mongolia.

Author Contributions: Conceptualization, M.X. and L.F.; methodology, P.Z.; software, P.Z.; invetigation, J.Y.; resources, L.F.; data curation, J.Y.; writing—original draft preparation, J.Y.; writing—review and editing, M.X. and L.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (32360788), Inner Mongolia Natural Science Foundation Project (2021MS03013), Western Young Scholar of Chinese Academy of Sciences, Basic Research Funds of Inner Mongolia Universities (BR22–13-11).

Data Availability Statement: Data are contained within the article.

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

References

1. Milosavljević, D.M.; Mutavdžić, D.R.; Radotić, K.; Milivojević, J.M.; Maksimović, V.M.; Maksimović, J.J.D. Phenolic profiling of 12 strawberry cultivars using different spectroscopic methods. *J. Agric. Food Chem.* **2020**, *68*, 4346–4354. [CrossRef] [PubMed]
2. Liu, Z.; Liang, T.; Kang, C. Molecular bases of strawberry fruit quality traits: Advances, challenges, and opportunities. *Plant Physiol.* **2023**, *193*, 900–914. [CrossRef] [PubMed]
3. Henning, S.M.; Seeram, N.P.; Zhang, Y.; Li, L.; Gao, K.; Lee, R.-P.; Wang, D.C.; Zerlin, A.; Karp, H.; Thames, G.; et al. Strawberry consumption is associated with increased antioxidant capacity in serum. *J. Med. Food* **2010**, *13*, 116–122. [CrossRef] [PubMed]
4. Skrovankova, S.; Sumczynski, D.; Mlcek, J.; Jurikova, T.; Sochor, J. Bioactive compounds and antioxidant activity in different types of berries. *Int. J. Mol. Sci.* **2015**, *16*, 24673–24706. [CrossRef]
5. Zacharaki, A.K.; Monaghan, J.M.; Bromley, J.R.; Vickers, L.H. Opportunities and challenges for strawberry cultivation in urban food production systems. *Plants People Planet* **2024**, *6*, 611–621. [CrossRef]
6. Hernández-Martínez, N.R.; Blanchard, C.; Wells, D.; Salazar-Gutiérrez, M.R. Current state and future perspectives of commercial strawberry production: A review. *Sci. Hortic.* **2023**, *312*, 111893. [CrossRef]
7. Shamaila, M.; Baumann, T.; Eaton, G.; Powrie, W.; Skura, B. Quality Attributes of Strawberry Cultivars Grown in British Columbia. *J. Food Sci.* **1992**, *57*, 696–699. [CrossRef]
8. Pădureț, S.; Oroian, M.; Gutt, G.; Amariei, S. Evaluation of strawberry texture in close relation with their anisotropy. *Int. J. Food Prop.* **2016**, *20*, 247–259. [CrossRef]
9. Jouquand, C.; Chandler, C.; Plotto, A.; Goodner, K. A sensory and chemical analysis of fresh strawberries over harvest dates and seasons reveals factors that affect eating quality. *J. Am. Soc. Hortic. Sci.* **2008**, *133*, 859–867. [CrossRef]
10. Kopjar, M.; Pilizota, V.; Hribar, J.; Tiban, N.N.; Subaric, D.; Babic, J.; Pozrl, T. Influence of Trehalose Addition On Instrumental Textural Properties of Strawberry Pastes. *Int. J. Food Prop.* **2008**, *11*, 646–655. [CrossRef]
11. Tanaka, Y.; Sasaki, N.; Ohmiya, A. Biosynthesis of plant pigments: Anthocyanins, betalains and carotenoids. *Plant J.* **2008**, *54*, 733–749. [CrossRef] [PubMed]
12. Liu, Y.; Tikunov, Y.; Schouten, R.E.; Marcelis, L.F.M.; Visser, R.G.F.; Bovy, A. Anthocyanin Biosynthesis and Degradation Mechanisms in Solanaceous Vegetables: A Review. *Front. Chem.* **2018**, *6*, 52. [CrossRef] [PubMed]
13. Navas, M.J.; Jiménez-Moreno, A.M.; Bueno, J.M.; Sáez-Plaza, P.; Asuero, A.G. Analysis and Antioxidant Capacity of Antho-cyanin Pigments. Part IV: Extraction of Anthocyanins. *Crit. Rev. Anal. Chem.* **2012**, *42*, 313–342. [CrossRef]
14. Landi, M.; Tattini, M.; Gould, K.S. Multiple functional roles of anthocyanins in plant-environment interactions. *Environ. Exp. Bot.* **2015**, *119*, 4–17. [CrossRef]
15. Lev-Yadun, S.; Gould, K.S. Role of Anthocyanins in Plant Defence. In *Anthocyanins: Biosynthesis, Functions, and Applications*; Springer: New York, NY, USA, 2008; pp. 22–28. [CrossRef]
16. Wang, L.S.; Stoner, G.D. Anthocyanins and their role in cancer prevention. *Cancer Lett.* **2008**, *269*, 281–290. [CrossRef]
17. Baltacıoğlu, C.; Keskin, O.; Baltacıoğlu, H.; Ağçam, E. Encapsulation and drying methods in the production of powdered red cabbage (*Brassica oleracea* L.): Chemometrics and Fourier transform infrared spectroscopy. *Food Sci. Technol. Int.* **2024**, 10820132241238261. [CrossRef]
18. Taghavi, T.; Patel, H.; Akande, O.E.; Galam, D.C.A. Total anthocyanin content of strawberry and the profile changes by extraction methods and sample processing. *Foods* **2022**, *11*, 1072. [CrossRef]
19. Steele, M.R.; Gitelson, A.A.; Rundquist, D.C.; Merzlyak, M.N. Nondestructive estimation of anthocyanin content in grapevine leaves. *Am. J. Enol. Vitic.* **2009**, *60*, 87–92. [CrossRef]
20. Gutiérrez, S.; Tardaguila, J.; Fernández-Novales, J.; Diago, M. On-the-go hyperspectral imaging for the in-field estimation of grape berry soluble solids and anthocyanin concentration. *Aust. J. Grape Wine Res.* **2019**, *25*, 127–133. [CrossRef]
21. Zhu, M.; Huang, D.; Hu, X.-J.; Tong, W.-H.; Han, B.-L.; Tian, J.-P.; Luo, H.-B. Application of hyperspectral technology in detection of agricultural products and food: A Review. *Food Sci. Nutr.* **2020**, *8*, 5206–5214. [CrossRef]
22. Cho, J.S.; Lim, J.H.; Park, K.J.; Choi, J.H.; Ok, G.S. Prediction of pelargonidin-3-glucoside in strawberries according to the postharvest distribution period of two ripening stages using VIS-NIR and SWIR hyperspectral imaging technology. *LWT* **2021**, *141*, 110875. [CrossRef]

23. Siedliska, A.; Baranowski, P.; Zubik, M.; Mazurek, W.; Sosnowska, B. Detection of fungal infections in strawberry fruit by VNIR/SWIR hyperspectral imaging. *Postharvest Biol. Technol.* **2018**, *139*, 115–126. [CrossRef]
24. Chen, S.; Zhang, F.; Ning, J.; Liu, X.; Zhang, Z.; Yang, S. Predicting the anthocyanin content of wine grapes by NIR hyperspectral imaging. *Food Chem.* **2015**, *172*, 788–793. [CrossRef] [PubMed]
25. Savitzky, A.; Golay, M.J. Soothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
26. Mo, C.; Kim, M.S.; Kim, G.; Lim, J.; Delwiche, S.R.; Chao, K.; Lee, H.; Cho, B.-K. Spatial assessment of soluble solid contents on apple slices using hyperspectral imaging. *Biosyst. Eng.* **2017**, *159*, 10–21. [CrossRef]
27. Raj, R.; Cosgun, A.; Kulić, D. Strawberry water content estimation and ripeness classification using hyperspectral sensing. *Agronomy* **2022**, *12*, 425. [CrossRef]
28. Xie, C.; Chu, B.; He, Y. Prediction of banana color and firmness using a novel wavelengths selection method of hyperspectral imaging. *Food Chem.* **2018**, *245*, 132–140. [CrossRef]
29. Xu, M.; Sun, J.; Yao, K.; Wu, X.; Shen, J.; Cao, Y.; Zhou, X. Nondestructive detection of total soluble solids in grapes using VMD-RC and hyperspectral imaging. *J. Food Sci.* **2021**, *87*, 326–338. [CrossRef]
30. Fan, S.; Huang, W.; Guo, Z.; Zhang, B.; Zhao, C. Prediction of soluble solids content and firmness of pears using hyperspectral reflectance imaging. *Food Anal. Methods* **2015**, *8*, 1936–1946. [CrossRef]
31. Pirie, A.; Mullins, M.G. Changes in anthocyanin and phenolics content of grapevine leaf and fruit tissues treated with sucrose, nitrate, and abscisic acid. *Plant Physiol.* **1976**, *58*, 468–472. [CrossRef]
32. Harker, F.; Elgar, H.; Watkins, C.B.; Jackson, P.J.; Hallett, I.C. Physical and mechanical changes in strawberry fruit after high carbon dioxide treatments. *Postharvest Biol. Technol.* **2000**, *19*, 139–146. [CrossRef]
33. Kader, A.A. Fruit maturity, ripening, and quality relationships. In Proceedings of the International Symposium Effect of Pre- & Postharvest Factors in Fruit Storage 485, Warsaw, Poland, 3–7 August 1997; pp. 203–208. [CrossRef]
34. Toacute, V. Sequential determination of titratable acidity and tartaric acid in wines by flow injection spectrophotometry. *Analyst* **1998**, *123*, 661–664. [CrossRef]
35. Yue, J.; Yang, G.; Feng, H. Comparative of remote sensing estimation models of winter wheat biomass based on random forest algorithm. *Trans. Chin. Soc. Agric. Eng.* **2016**, *32*, 175–182. [CrossRef]
36. Vines, L.L.; Kays, S.E.; Koehler, P.E. Near-infrared reflectance model for the rapid prediction of total fat in cereal foods. *J. Agric. Food Chem.* **2005**, *53*, 1550–1555. [CrossRef] [PubMed]
37. Li, Z.; Song, J.; Ma, Y.; Yu, Y.; He, X.; Guo, Y.; Dou, J.; Dong, H. Identification of aged-rice adulteration based on near-infrared spectroscopy combined with partial least squares regression and characteristic wavelength variables. *Food Chem. X* **2023**, *17*, 100539. [CrossRef] [PubMed]
38. Rasmussen, C.E. Gaussian processes in machine learning. In *Summer School on Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 63–71.
39. Bruning, B.; Liu, H.; Brien, C.; Berger, B.; Lewis, M.; Garnett, T. The development of hyperspectral distribution maps to predict the content and distribution of nitrogen and water in wheat (*Triticum aestivum*). *Front. Plant Sci.* **2019**, *10*, 1380. [CrossRef] [PubMed]
40. Zhang, J.; Liu, L.; Chen, Y.; Rao, Y.; Zhang, X.; Jin, X. The Nondestructive Model of Near-Infrared Spectroscopy with Different Pretreatment Trans-formation for Predicting "Dangshan" Pear Woolliness Disease. *Agronomy* **2023**, *13*, 1420. [CrossRef]
41. Guo, C.; Zhang, J.; Cai, W.; Shao, X. Enhancing Transferability of Near-Infrared Spectral Models for Soluble Solids Content Prediction across Different Fruits. *Appl. Sci.* **2023**, *13*, 5417. [CrossRef]
42. Ouyang, A.-G.; Xie, X.-Q.; Zhou, Y.-R.; Liu, Y.-D. Partial least squares regression variable screening studies on apple soluble solids NIR spectral detection. *Spectrosc. Spectr. Anal.* **2012**, *32*, 2680–2684. [CrossRef]
43. Liu, M.; Liu, M.; Bai, L.; Shang, W.; Ren, R.; Zhao, Z.; Sun, Y. Establishing a Berry Sensory Evaluation Model Based on Machine Learning. *Foods* **2023**, *12*, 3502. [CrossRef]
44. Basak, J.K.; Madhavi, B.G.K.; Paudel, B.; Kim, N.E.; Kim, H.T. Prediction of total soluble solids and pH of strawberry fruits using RGB, HSV and HSL colour spaces and machine learning models. *Foods* **2022**, *11*, 2086. [CrossRef]
45. Sarkar, S.; Basak, J.K.; Moon, B.E.; Kim, H.T. A Comparative study of PLSR and SVM-R with various preprocessing techniques for the quantitative determination of soluble solids content of hardy Kiwi fruit by a portable Vis/NIR spectrometer. *Foods* **2020**, *9*, 1078. [CrossRef]

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.