

## Article

# Rapid Detection of Tannin Content in Wine Grapes Using Hyperspectral Technology

Peng Zhang <sup>1</sup>, Qiang Wu <sup>2</sup>, Yanhan Wang <sup>1</sup>, Yun Huang <sup>1</sup>, Min Xie <sup>1,\*</sup> and Li Fan <sup>1,\*</sup>

<sup>1</sup> College of Agriculture, Inner Mongolia Agricultural University, Huhhot 010010, China; 2024zhangp@gmail.com (P.Z.)

<sup>2</sup> Inner Mongolia Academy of Agriculture and Animal Husbandry, Huhhot 010031, China; imauwq@emails.imau.edu.cn

\* Correspondence: xiemin@imau.edu.cn (M.X.); fanli@imau.edu.cn (L.F.)

**Abstract:** Wine grape quality is influenced by the variety and growing environment, and the quality of the grapes has a significant impact on the quality of the wine. Tannins are a crucial indicator of wine grape quality, and, therefore, rapid and non-destructive methods for detecting tannin content are necessary. This study collected spectral data of Pinot Noir and Chardonnay using a geophysical spectrometer, with a focus on the 500–1800 nm spectrum. The spectra were preprocessed using Savitzky–Golay (SG), first-order differential (1D), standard normal transform (SNV), and their respective combinations. Characteristic bands were extracted through correlation analysis (PCC). Models such as partial least squares (PLS), support vector machine (SVM), random forest (RF), and one-dimensional neural network (1DCNN) were used to model tannin content. The study found that preprocessing the raw spectra improved the models' predictive capacity. The SVM–RF model was the most effective in predicting grape tannin content, with a test set  $R^2$  of 0.78, an RMSE of 0.31, and an RE of 10.71%. These results provide a theoretical basis for non-destructive testing of wine grape tannin content.

**Keywords:** hyperspectral; wine grape; tannin; non-destructive detection



**Citation:** Zhang, P.; Wu, Q.; Wang, Y.; Huang, Y.; Xie, M.; Fan, L. Rapid Detection of Tannin Content in Wine Grapes Using Hyperspectral Technology. *Life* **2024**, *14*, 416. <https://doi.org/10.3390/life14030416>

Academic Editors: Rongsheng Zhu and Joel Tellinghuisen

Received: 3 February 2024

Revised: 5 March 2024

Accepted: 18 March 2024

Published: 21 March 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

The study of phenolic compounds has become increasingly important in recent years due to their significant impact on the sensory properties of red wine [1,2]. Anthocyanins and tannins had a great influence on the quality of red wine. They influenced the organoleptic properties of wine by interacting with other components [3]. Among them, tannins are found in grape skins and seeds and are an important phenolic compound in grapes [4,5]. These compounds impart astringency to wine, which influences its taste and flavor [6]. Additionally, tannins contribute to the color changes observed during wine aging [7]. The tannin content in wine grapes increases with maturation. However, the increase in tannin content is influenced by various factors, such as grape varieties, the growth environment, and cultivation practices [8–10]. Therefore, monitoring tannin levels is crucial for timely adjustments in cultivation management, intervention in grape growth, and determining the optimal harvest period [11]. This process is essential for producing high-quality red wine.

Tannin extraction is predominantly conducted using chemical methods. Spectrophotometry and high-performance liquid chromatography (HPLC) represent the standard techniques for determining tannin content [12]. However, these methods are characterized by lengthy extraction cycles and cumbersome procedures [13]. Hyperspectral technology has emerged as a powerful analytical tool due to its continuous evolution. Hyperspectral technology has gained widespread use in monitoring key plant indicators to assess growth status due to its rapid and non-destructive nature [14]. Numerous researchers have utilized hyperspectral imaging to monitor grape phenotypic characteristics, soluble solids content,

and anthocyanin content [15–17]. For instance, Zhang et al. employed hyperspectral imaging to detect tannin content in grains and established a predictive model using both full and characteristic wavelengths. The study found that hyperspectral technology enables quick and non-invasive evaluation of tannin levels in grains [18]. Maria Inês Rouxinol and her colleagues used a portable infrared spectrometer to collect spectral data on wine grapes, analyzing various components, including tannins. They then modeled the data using partial least squares regression (PLS). The results showed significant potential for accurately predicting tannin content [19].

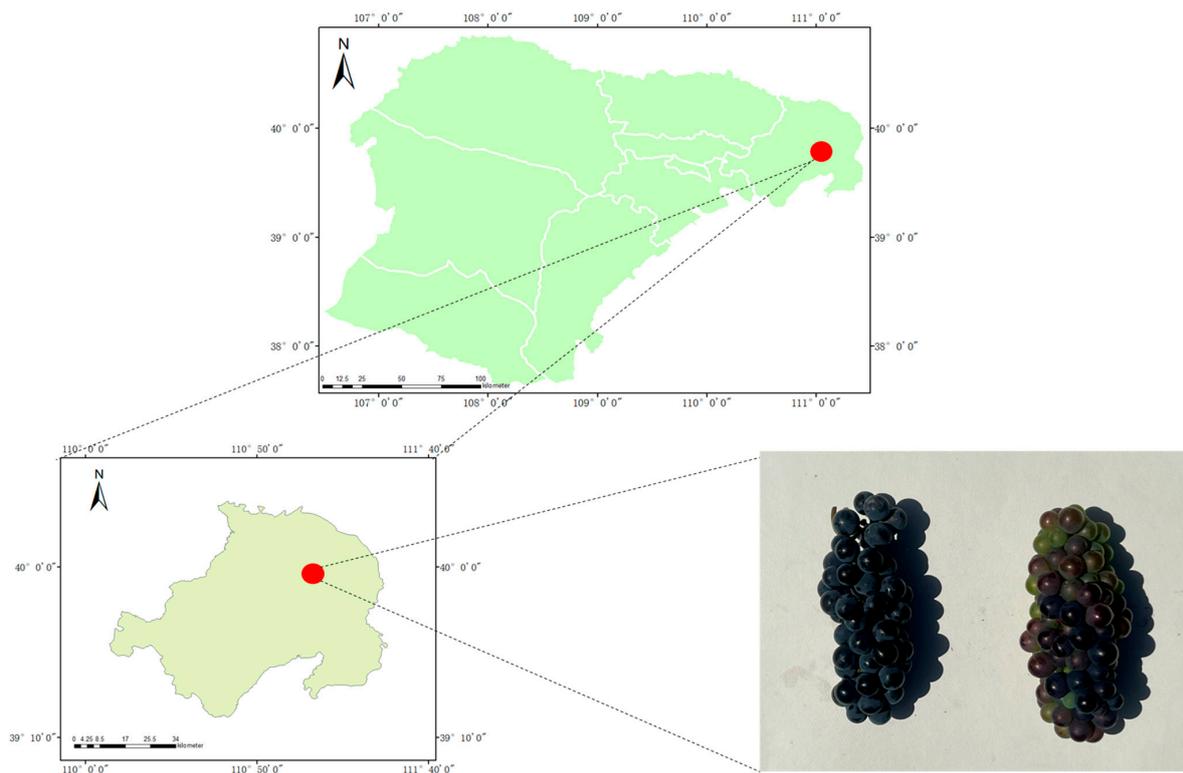
Chen et al. used hyperspectral imaging to measure the tannin content in drying persimmons. They applied seven preprocessing methods (SG, SNV, 1D, and 2D) to prepare the data and developed models to assess the effectiveness of these methods. The results showed that SG1D and SG2D were the most effective, with  $R^2$  values of 0.742 and 0.857, respectively. This highlights the importance of choosing appropriate preprocessing techniques to improve model performance [20]. Gao Sheng et al. analyzed red tip berries using hyperspectral imaging and employed preprocessed hyperspectral data for modeling. They developed predictive models, such as partial least squares regression (PLS), least squares support vector machine (LSSVM), and random forest (RF), to predict the Brix and hardness of red grapes. The study found that the RF model was the most effective in predicting both Brix and hardness, with  $R^2$  values of 0.928 and 0.932, respectively. This highlights the importance of selecting an appropriate model to improve accuracy in predicting these parameters [21]. Julio Nogales-Bueno et al. used spectroscopic equipment to scan harvested grapes and preprocessed the hyperspectral data using multiplicative scattering correction (MSC), standardized normal variable (SNV), and detrending. The study showed that hyperspectral techniques can be used to quickly and non-destructively detect polyphenol content in grapes. It also predicted the extractable polyphenol content in red grape skins [22].

Specifically, this study focused on analyzing wine grapes from Ordos (Zhungeer Banner). Spectral data was collected using a portable geophysical spectrometer (SVC HR-1024i) and processed using various techniques, including SG, SNV, 1D, SG1D, SG1D-SNV, and RAW. The study used spectral feature extraction integrated with principal correlation analysis (PCC) to model grape tannin content. Four different estimation models were employed: random forest (RF), support vector machine (SVM), partial least squares (PLS), and 1-dimensional neural network (1DCNN). The accuracies of the models were compared to determine the most effective one for estimating tannin content.

## 2. Materials and Methods

### 2.1. Sample Preparation

The experiment was carried out in Zhungeer Banner, Ordos City, from 2021 to 2022 using Pinot Noir and Chardonnay as experimental varieties (Figure 1). Both experimental varieties were harvested at the same time. The harvest dates in 2021 were 28 August, 4 September, 11 September, and 18 September. The harvest dates in 2022 were 25 August, 1 September, 8 September, and 15 September.



**Figure 1.** Location of the research areas and experimental designs.

## 2.2. Spectral Acquisition

In this study, a portable geophysical spectrometer (HR-1024, SVC, Manufactured by Sloan Valve Company (svc), located in Franklin Industrial Park near Chicago, IL, USA) was used with a detection range of 500–1800 nm. The built-in CPU provided data processing capacity, and the personal digital assistant (PDA) enabled real-time information transmission through remote Bluetooth technology. Prior to collecting spectral data, the instrument was calibrated by scanning a whiteboard. Data was collected by scanning the wine grape berries. During scanning, ten plants of each variety were selected. Two spikes were taken from each plant, and five grains were taken from each spike, totaling 100 grains (20 spikes). Every 10 grains were divided into a group. A total of 160 samples were collected for further analysis by making four measurements before harvesting. The spectrometer automatically adjusted the integration time according to changes in light intensity for optimal scanning. After completing the scanning process, we measured the reflectance of the grape samples using the companion software of the instrument called DARWinSP (version 1.10.8).

## 2.3. Software and Model Evaluation

The study utilized TensorFlow 2.1, a deep-learning framework, in Python (version 3.7.16). The computer was equipped with a GeForce GTX 1650 graphics card with 6 GB video memory and an Intel(R) Core (TM) i7-9750H processor operating at 2.59 GHz.

## 2.4. Measurements of Tannin Content

Tannin content was determined by the Folin–Denis method. A total of 10 g of grapes were placed in a triangular flask and 50 mL of distilled water was added. The grapes were filtered in a water bath at 60 °C for 12 h. The supernatant was extracted in a water bath at 80 °C for 20 min and filtered. A total of 2 mL of the sample filtrate was aspirated and centrifuged at 8000 r/min for 4 min and the supernatant was set aside. Then, 1 mL of 0 g/L, 20 g/L, 40 g/L, 60 g/L gallic acid standard use solution was sucked up, and 5 mL of distilled water, 1 mL of a sodium tungstate–sodium molybdate mixture and 3 mL of

sodium carbonate solution were added; the concentration of gallic acid standard solution was 0 g/L, 2 g, 4 g/L, 6 g/L, respectively, and the color was developed and left for 2 h, and then 0 g/L was used as the blank of the standard curve. The absorbance was measured at 760 nm using a spectrophotometer and the standard curve was plotted. Pipette 1 mL of sample supernatant, add 5 mL of water, 1 mL of a sodium tungstate–sodium molybdate mixed solution and 3 mL of sodium carbonate solution, respectively, develop the color and leave it for 2 h, and measure the absorbance at 760 nm using the standard curve 0 g/L as blank. The tannin content of the samples was calculated as in Equation (1).

$$X = C \times (1/V) \times 250 \times (1/1000) \times (1/m) \times 1000 \quad (1)$$

In this equation, 'X' represents the tannin content (g/L), 'C' is the absorbance of the sample on the standard curve (mg), 'V' is the volume of the test solution (mL), '250' is the total volume of the extract (mL), and 'm' is the mass of the weighed sample (g).

## 2.5. Data Analysis Methods

### 2.5.1. Hyperspectral Preprocessing

To enhance model accuracy, we conducted data preprocessing on the raw hyperspectral reflectance data. We employed six preprocessing methods in this experiment: SG, 1D, SNV, SG1D, SG1D-SNV and Raw. Hyperspectral preprocessing methods can reduce or eliminate the impact of unimportant data on spectral data, reduce background noise interference, and highlight spectrally valid information, thereby improving spectral sensitivity.

### 2.5.2. Data Dimensionality

Data dimensionality reduction is a technique that eliminates redundant spectral information, reducing the likelihood of model overfitting and improving the speed of model operation. In this study, we utilized the Pearson Correlation Coefficient (PCC) method for data dimensionality reduction. The PCC method reflects the strength of the linear relationship between two variables, allowing for the screening out of characteristic bands. The calculation formula is presented below as Equation (2).

$$r_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \quad (2)$$

In this equation, 'r' represents the correlation coefficient, 'X' the spectral wavelength, 'Y' the grape tannin content, 'cov' the covariance, and 'σ' the standard deviation.

### 2.5.3. Model Establishment

The data on grape tannins were split into a training set and a test set in an 8:2 ratio using Scikit-learn in Python. We developed a mathematical model for predicting grape tannin content using hyperspectral non-destructive testing and subsequently validated and evaluated the predictive ability of each model for accuracy.

SVM is a machine learning algorithm based on the principle of structural risk minimization. It reduces the complexity of the learning machine to achieve good generalization ability while ensuring training accuracy. SVM is effective in addressing issues with small samples, nonlinearities, and high dimensions, making it widely applicable in regression problems [23]. The kernel function in SVM used the radial basis function kernel. The method to find the optimal parameters was to utilize the cross-validation method, which included the parameter C (penalty factor) and the parameter  $\delta$  (variance in the RBF kernel function). In this study, the paper mentioned that  $C = 2$  and  $\delta = 0.7$  were used as parameters for modeling.

Random forest (RF) is another machine learning algorithm proposed by Breiman in 2001, which is tailored for small-scale data. The random forest algorithm is known for its robustness and strong generalization capabilities, as well as its fast training speed. It is particularly effective in handling high-dimensional data and large-scale datasets with

high accuracy [24]. This study modeled 200 decision trees, and the number of independent variables required to create branches was set to 'auto'.

The PLS model is particularly suitable for inverse modeling of datasets with small sample sizes and is conducive to refining key spectral information. PLS merges the benefits of principal component analysis, canonical correlation analysis, and multiple linear regression. This approach offers a many-to-many linear regression model and considers the explanatory power of the independent variables for the dependent variable [25].

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (3)$$

where  $a_0$  in the equation is the intercept of the regression coefficient,  $a_i$  is the regression coefficient,  $x_i$  are the independent variables 1 to  $n$ .

The 1DCNN model consists of an input module, a convolution module, a fully connected layer, and a regression output layer. The model's input parameters include the spectral data corresponding to each sample and the measured tannin values. This model demonstrates strong generalization and nonlinear capabilities, making it suitable for the conditions and requirements of this experiment [26]. The convolutional layers of the convolutional module consisted of convolutions with 16 kernels, a size of  $3 \times 3$ , and steps of one. The number of convolutional layers was determined by the number of specially acquired features.

#### 2.5.4. Model Performance

A prediction model was developed with tannin content as the dependent variable and the model was evaluated using the coefficient of determination ( $R^2$ ) and root mean square error (RMSE). The larger  $R^2$  is closer to 1. It means that the model is more accurate. A smaller RMSE indicates that the accuracy of the model is more robust. The two evaluation coefficients were formulated as in Equations (4) and (5).

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (5)$$

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

### 3. Results

#### 3.1. Analysis of the Tannin Content of Grapes

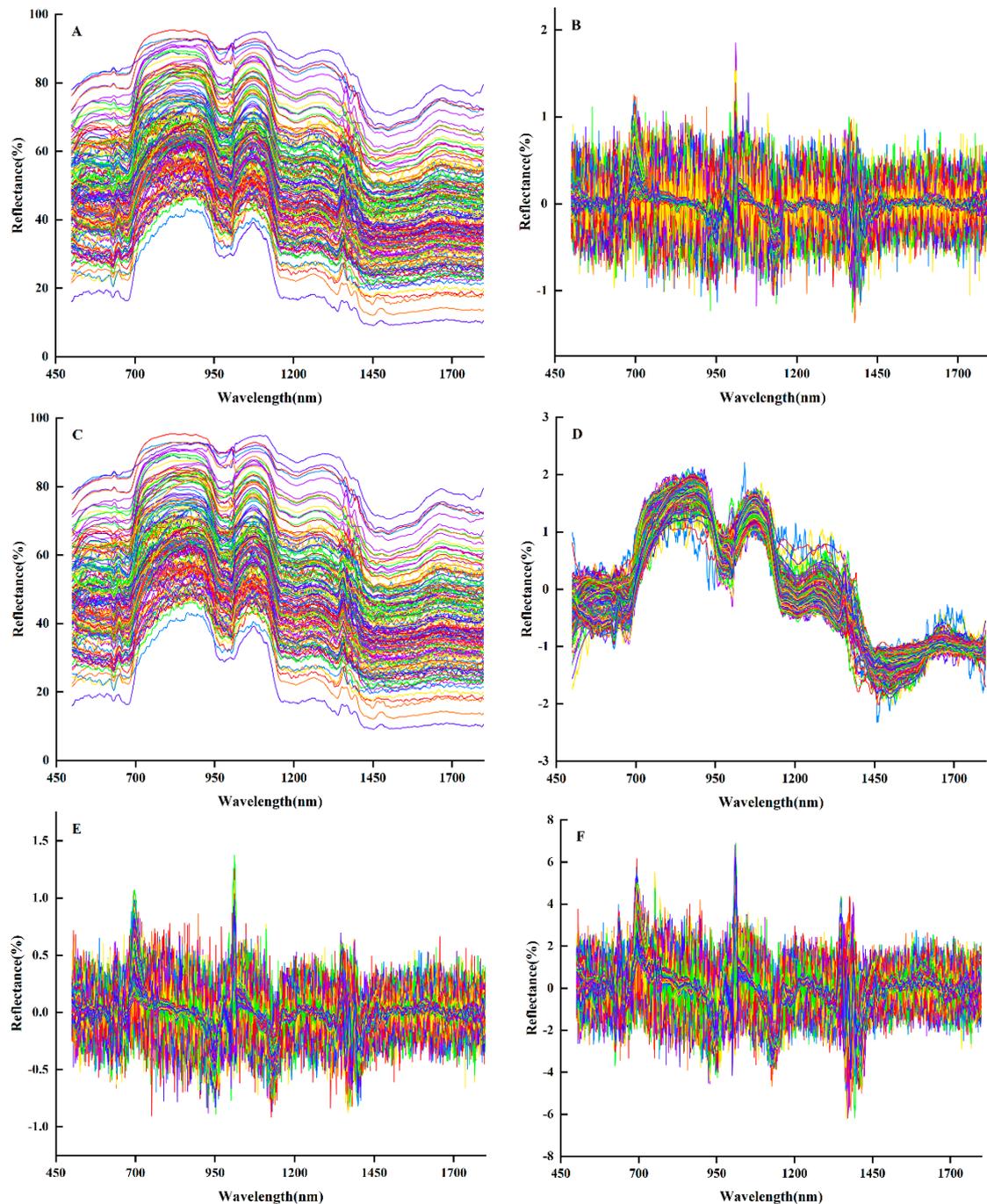
Analysis of the tannin content data for the two grape varieties from the 2021–2022 harvest (Table 1) showed that the tannin content of the grapes ranged from 1.06 to 3.92, with Chardonnay having a tannin content of 1.09–3.85 and Pinot Noir having a tannin content of 1.06–3.92. In addition, in 2021 and 2022, the average tannin content of Chardonnay was 2.01 and 2.32, respectively, and the average tannin content of Pinot Noir was 2.35 and 2.66, respectively, which was higher than the average tannin content of Chardonnay in both years.

**Table 1.** Statistical analysis of tannin content in two grape varieties.

Variety	Measuring Time	Sample Size	Minimum (g/L)	Maximum (g/L)	Mean (g/L)	Standard Deviation (g/L)
Chardonnay	2021	40	1.12	3.19	2.01	0.46
	2022	40	1.09	3.85	2.32	0.59
Pinot Noir	2021	40	1.06	3.31	2.35	0.70
	2022	40	1.08	3.92	2.66	0.86

### 3.2. Hyperspectral Data Preprocessing Analysis

As shown in Figure 2A, the spectral reflectance curves of grape berries showed three peaks at 920 nm, 1070 nm, and 1350 nm, which were related to the vibration of N-H and C-H groups in the samples. The troughs at 950 nm, 1130 nm, and 1400 nm were related to the C-H, N-H, and O-H of the tannins in the samples, which indicated that there was a close correlation between the spectral reflectance of tannins and the tannin content.



**Figure 2.** Hyperspectral image and preprocessing image. (A) The raw average reflectance image. (B) First derivative (1D) preprocessing image. (C) SG smoothing preprocessing image. (D) SNV preprocessing image. (E) SG1D preprocessed image. (F) SG1D-SNV preprocessed image.

Given the broad range of spectral data for grape berries and the extensive number of measurement periods, the external environment can influence the spectral reflectance. The first-order derivative (1D) was selected to enhance the convergence speed of the model (Figure 2B). SG smoothing was applied for data smoothing (Figure 2C), and the standard normalized variable (SNV) was employed to eliminate the gap (Figure 2D). Furthermore, the three preprocessing methods were integrated as SG1D (Figure 2E) and SG1D-SNV (Figure 2F) for spectral data preprocessing.

### 3.3. Data Dimension Reduction

The experiment utilized the Pearson Correlation Coefficient (PCC) method to extract characteristic bands. Only bands highly correlated with grape tannins were extracted from the preprocessed data. Wavelengths with a correlation greater than 0.5 and ranking in the top 20 were used instead of the original bands. This approach reduced model complexity and shortened the modeling time. Table 2 displays the results of feature band extraction.

**Table 2.** Extracted characteristic wavelengths.

Preprocessing Method	Number of Feature Bands	Maximum Correlation Coefficient
1D	19	0.63
SG	20	0.72
SNV	20	0.86
SG1D	19	0.60
SG1DSNV	20	0.62
RAW	20	0.73

### 3.4. Performance of Models for Tannin Content Estimation

In this study, tannin content prediction was performed using SVM, RF, PLS, and 1DCNN with various preprocessing methods. To improve model accuracy, the dataset was divided into 128 training sets and 32 test sets, in an 8:2 ratio. Model evaluation coefficients included  $R^2$ , RMSE and RE. The table highlights the best inversion results for each model by comparing the values of  $R^2$ , RMSE, and RE.

#### 3.4.1. SVM Model Prediction Results

The prediction results of the SVM model were shown in Table 3, the  $R^2$  of the spectral training set based on 1D, SG, and SNV were all greater than 0.80, the  $R^2$  of the training set of the three was not much different, and the three were mainly compared from the test set  $R^2$ . Among them, the spectral test set  $R^2 = 0.77$  based on SNV, and the spectral test set  $R^2$  of SG and 1D were 0.75 and 0.66, respectively, which were lower values than those of SNV. The test sets RMSE and RE of SG and 1D were larger than those of SNV, which indicated that, when using the SVM model to monitor the tannin content of grapes, the preprocessing of the spectra by choosing SNV could effectively improve the predictive ability and stability of the model.

**Table 3.** Training and testing results of SVM model to estimate tannin content.

Preprocessing Method	Train			Test		
	$R^2$	RMSE	RE (%)	$R^2$	RMSE	RE (%)
Raw	0.78	0.33	11.56	0.71	0.35	12.07
1D	0.80	0.31	9.63	0.66	0.38	13.61
SG	0.78	0.33	11.56	0.71	0.35	12.07
SNV	0.78	0.33	11.13	0.77	0.31	10.77
SG1D	0.64	0.43	13.90	0.43	0.49	18.39
SG1DSNV	0.59	0.45	13.86	0.37	0.52	19.53

### 3.4.2. RF Model Prediction Results

The prediction results of the RF model were shown in Table 4; the  $R^2$  of the spectral training set with different preprocessing showed that SNV was the highest, and 1D was the second highest, with  $R^2$  of 0.97 and 0.96, respectively. There was not much difference in the  $R^2$  of the training set of the two, and the comparison was mainly drawn from the  $R^2$  of the test set. From the table, it could be seen that the  $R^2$  of the spectral test set based on SNV and 1D were 0.78 and 0.56, respectively, and the SNV spectra were better than the 1D spectra. In addition, the spectral test sets RMSE and RE of SNV were smaller than those of 1D. Therefore, when using the RF model to monitor the tannin content of grapes, choosing SNV for spectral preprocessing can effectively improve the accuracy and robustness of the model.

**Table 4.** Training and testing results of RF model to estimate tannin content.

Preprocessing Method	Train			Test		
	$R^2$	RMSE	RE (%)	$R^2$	RMSE	RE (%)
Raw	0.96	0.14	4.72	0.80	0.29	9.64
1D	0.95	0.16	5.63	0.56	0.43	16.16
SG	0.96	0.14	4.77	0.81	0.28	9.29
SNV	0.97	0.12	4.25	0.78	0.31	10.71
SG1D	0.92	0.20	7.19	0.40	0.51	20.92
SG1DSNV	0.92	0.20	7.17	0.39	0.51	19.97

### 3.4.3. PLS Model Prediction Results

The prediction results of the PLS model were shown in Table 5, the  $R^2$  of the spectral training sets with different preprocessing was greater than 0.5, the training set  $R^2$  did not differ much, and the comparison was mainly drawn from the test set  $R^2$ . From the table, it could be seen that the test set  $R^2$  of 1D-based spectra was 0.69, which was much higher than the test set  $R^2$  of the other preprocessed spectra. In addition, the test set RMSE and RE of the spectra of 1D were 0.36 and 13.10%, which were smaller than the other preprocessed spectra. This suggested that when using the PLS model to monitor the tannin content of grapes, choosing 1D to preprocess the spectra could effectively improve the predictive ability and stability of the model.

**Table 5.** Training and testing results of PLS model to estimate tannin content.

Preprocessing Method	Train			Test		
	$R^2$	RMSE	RE (%)	$R^2$	RMSE	RE (%)
Raw	0.67	0.41	15.07	0.57	0.43	16.66
1D	0.68	0.40	15.41	0.69	0.36	13.10
SG	0.67	0.41	15.07	0.57	0.43	16.66
SNV	0.78	0.33	11.13	0.77	0.31	10.77
SG1D	0.55	0.48	17.94	0.59	0.42	16.51
SG1DSNV	0.53	0.49	17.62	0.47	0.47	18.75

### 3.4.4. 1DCNN Model Prediction Results

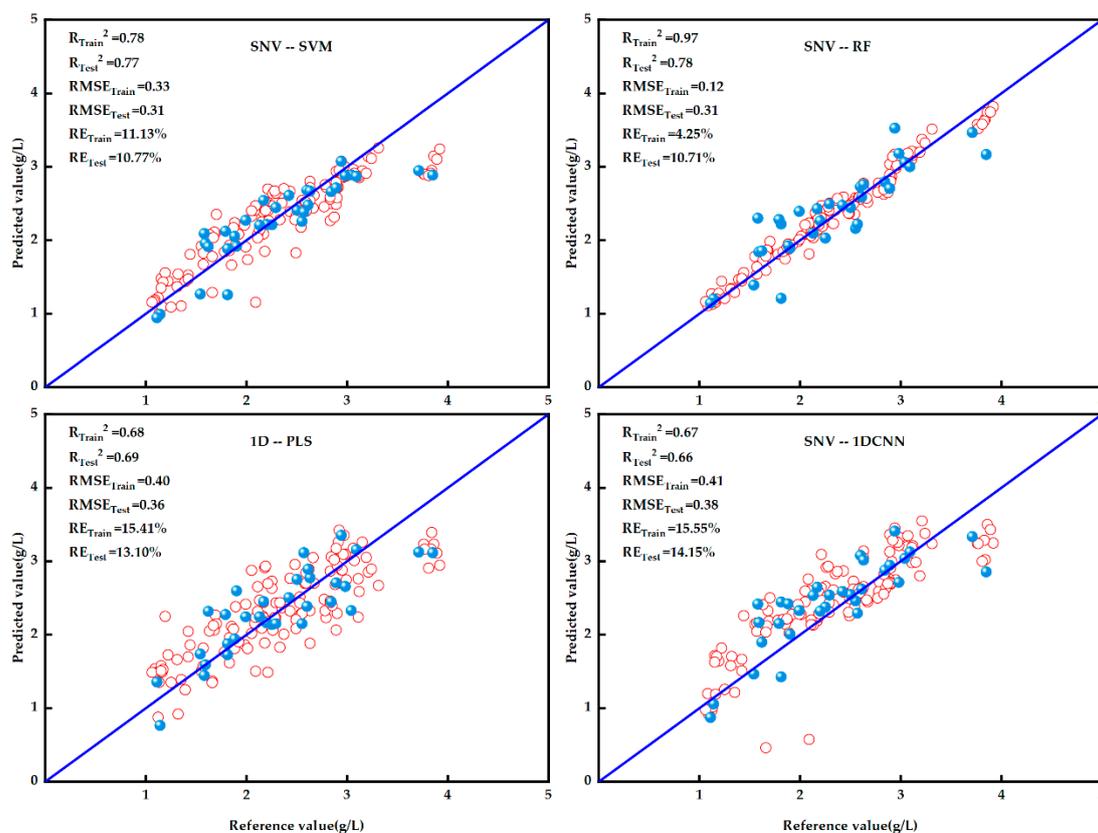
The prediction results of the 1DCNN model were shown in Table 6; the spectral training set based on SG1DSNV had a higher  $R^2$  of 0.87, followed by 1D and SNV spectra with 0.79 and 0.71, respectively. From the table, it could be seen that the spectral test set based on SNV had the highest  $R^2$ , followed by 1D and SG1D spectra with  $R^2$  of 0.70, 0.63, and 0.50, respectively. The other preprocessed spectral test sets all had  $R^2$  less than 0.5 and had poor prediction results. In addition, the spectral test sets RMSE and RE of SNV were smaller than 1D and other preprocessed spectra. This suggested that when using the 1DCNN model to monitor the tannin content of grapes, choosing SNV to preprocess the spectra could effectively improve the accuracy and robustness of the model.

**Table 6.** Training and testing results of 1DCNN model to estimate tannin content.

Preprocessing Method	Train			Test		
	R <sup>2</sup>	RMSE	RE (%)	R <sup>2</sup>	RMSE	RE (%)
Raw	0.31	1.60	27.82	0.23	1.53	24.12
1D	0.76	0.35	10.84	0.31	0.54	21.34
SG	0.43	1.03	23.53	0.27	0.81	17.03
SNV	0.67	0.41	15.55	0.66	0.38	14.15
SG1D	0.66	0.41	14.56	0.50	0.46	19.12
SG1DSNV	0.87	0.26	8.84	0.21	0.58	23.31

### 3.4.5. Selection of Optimal Model for Tannin Content Estimation

As depicted in Figure 3, the four modeling methods—SVM, RF, PLS, and 1DCNN—were compared. The models exhibiting the best prediction performance were selected to create independent validation scatter plots, showing both measured and predicted tannin content. The most effective predictive models were identified as SNV-SVM, SNV-RF, 1D-PLS, and SNV-1DCNN, respectively. Notably, the sample distributions in the validation and test sets of the SNV-RF model showed minimal deviation from the 1:1 line, especially when compared to those of the SNV-SVM, 1D-PLS, and SNV-1DCNN models. This distribution was essentially linear along the 1:1 line, suggesting that the prediction accuracy of the SNV-RF model surpasses that of the other three models overall. Consequently, the SNV-RF model was selected for detecting grape tannin content, as it could further enhance the accuracy and stability of the prediction results.



**Figure 3.** Validation results of the regression model for tannin content. Each fit is plotted for the training and test sets, with the red blobs indicating the training set data and the blue blobs indicating the test set data. The degree of model strength can be summarized based on the deviation of the model from the standard line, and these plots show R<sup>2</sup>, RMSE, and RE.

#### 4. Discussion

In recent years, spectroscopic techniques have been widely used for the rapid monitoring of fruit substance content, among other applications [27]. Visible-near-infrared spectroscopy has been demonstrated by numerous researchers as feasible for predicting grape composition [28,29]. It is worth noting that the majority of these studies utilized raw spectral data without implementing spectral data preprocessing and feature band extraction. The use of redundant and complex spectral data resulted in decreased model prediction accuracy and operational speed. To address this issue, spectral data preprocessing is performed, followed by feature band extraction based on the preprocessed spectral data. This approach reduces the dimensionality of the spectral data and retains the feature bands that are highly correlated with the samples. Thus, this methodology effectively addresses the issues of decreased predictive ability and operational speed of the model. In this study, six preprocessing methods were used to process the spectral data, aiming to eliminate noise and enhance spectral variability, thereby improving spectral quality. Furthermore, principal component analysis (PCA) was used for feature band extraction to achieve data downscaling and simplification. The feasibility of four distinct modeling methods for the prediction of tannin content in wine grapes has also been investigated.

In this study, the raw spectral data were preprocessed, and modeling based on the preprocessed data resulted in improved model accuracy. This result is consistent with the results of the study [30]. The modeling prediction of the spectral data preprocessed by SNV was the best. This may be due to the fact that SNV standardizes and normalizes the raw data to further improve the accuracy of the spectral data and make the differences between different spectra more significant, thus enabling PCC to extract the characteristic bands more accurately. This is what leads to the higher accuracy of the model built based on the spectral data preprocessed by SNV. In addition, in this study, PCC was used for data downscaling and feature band extraction, replacing 1300 variables with 19–20 variables, which improved the running speed of the model. The results showed that the extracted feature bands were feasible for estimating grape tannin content. This is consistent with the results of the study [20]. In this study, SVM, RF, PLS and 1DCNN were used to develop an accurate quantitative model for wine grape tannin content. In this study, each of the four models was modeled based on six pre-processed spectral data, for a total of 24 combined models. The optimal models corresponding to each modeling method were SNV-SVM, SNV-RF, 1D-PLS, and SNV-1DCNN, respectively. Comparison of these four models revealed that the SNV-RF model had the strongest predictive ability, which may be due to the fact that the RF model based on SVM spectral preprocessing is more suitable for data with small sample sizes and high dimensionality of variables. The model provides a theoretical basis for the prediction of tannin content in wine grapes using small sample data.

Considering the limitations of this study, it is crucial to note that, despite examining the tannin content at the maturity of different wine grape varieties over two consecutive years, the small number of varieties selected limits the model's generalizability. Future research could benefit from increasing the number of varieties and ecological zones, thereby enhancing the model's generalizability. Furthermore, while PCC is employed for feature band extraction, this method may inadvertently exclude some crucial bands due to their marginally lower correlation, leading to the loss of significant bands. In subsequent research, exploring various methods for feature band extraction to achieve data dimensionality reduction and enhance model accuracy will be valuable. It should be noted that deep learning models typically demonstrate greater applicability to larger datasets. Thus, in future studies, expanding the sample size and training the model with these accumulated samples will be crucial for enhancing its predictive capability and stability.

#### 5. Conclusions

In this study, we describe the complete workflow for predicting tannin content in grapes based on hyperspectral detection of tannin content in grapes, using six preprocessing methods for spectral data preprocessing. The extraction of feature bands was performed

using PCC, and the data were downsampled to improve the speed of model running. And four modeling methods were used to increase the comparability of the models by pairing with the six preprocessing to determine the best model. The results show that the use of SNV for spectral data preprocessing can effectively improve the predictive ability and stability of the model when using SVM, RF and 1DCNN models for prediction. A comparison of the optimal models of the four modeling approaches revealed that the SNV-RF model had the highest accuracy and good robustness in predicting grape tannin content. Its  $R^2 = 0.78$ , RMSE = 0.21, and RE = 10.71%. These results indicate that it is feasible to utilize hyperspectral technology for the detection of tannin content in wine grapes and to provide a theoretical basis for the rapid non-destructive detection of tannin content in wine grapes.

**Author Contributions:** Conceptualization, M.X. and L.F.; methodology, Q.W.; software, Q.W.; investigation, Y.W. and Y.H.; resources, L.F.; data curation, P.Z.; writing—original draft preparation, P.Z.; 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 Central Government Guides Local Science and Technology Development Fund Projects (grant No: 2021ZY0021), Inner Mongolia Science and Technology Program (2023YFHH0056), Basic Research Funds of Inner Mongolia Universities (BR22-13-11).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Limitations applied to the dataset: the data provided in this study are available upon request from the corresponding author; the data are not publicly available due to privacy.

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

## References

1. Todorov, S.D.; Alves, V.F.; Popov, I.; Weeks, R.; Pinto, U.M.; Petrov, N.; Ivanova, I.V.; Chikindas, M.L. Antimicrobial compounds in wine. *Probiotics Antimicrob. Proteins* **2023**, *ahead of print*. [CrossRef]
2. Gabrielli, M.; Ounaissi, D.; Lançon-Verdier, V.; Julien, S.; Le Meurlay, D.; Maury, C. Hyperspectral imaging to assess wine grape quality. *JSEA Rep.* **2023**, *3*, 452–462. [CrossRef]
3. Cheynier, V.; Duenas-Paton, M.; Salas, E.; Maury, C.; Souquet, J.M.; Sarni-Manchado, P.; Fulcrand, H. Structure and properties of wine pigments and tannins. *Am. J. Enol. Vitic.* **2006**, *57*, 298–305. [CrossRef]
4. Molino, S.; Francino, M.P.; Henares, J.Á.R. Why is it important to understand the nature and chemistry of tannins to exploit their potential as nutraceuticals? *Food Res. Int.* **2023**, *173*, 113329. [CrossRef]
5. Wimalasiri, P.M.; Harrison, R.; Hider, R.; Donaldson, I.; Kemp, B.; Tian, B. Development of Tannins and Methoxypyrazines in Grape Skins, Seeds, and Stems of Two Pinot Noir Clones during Ripening. *J. Agric. Food Chem.* **2023**, *71*, 15754–15765. [CrossRef]
6. Zhao, Q.; Du, G.; Zhao, P.; Guo, A.; Cao, X.; Cheng, C.; Liu, H.; Wang, F.; Zhao, Y.; Liu, Y.; et al. Investigating wine astringency profiles by characterizing tannin fractions in Cabernet Sauvignon wines and model wines. *Food Chem.* **2023**, *414*, 135673. [CrossRef] [PubMed]
7. Pérez-Gil, M.; Pérez-Lamela, C.; Falqué-López, E. Comparison of Chromatic and Spectrophotometric Properties of White and Red Wines Produced in Galicia (Northwest Spain) by Applying PCA. *Molecules* **2022**, *27*, 7000. [CrossRef]
8. Song, X.; Yang, W.; Qian, X.; Zhang, X.; Ling, M.; Yang, L.; Shi, Y.; Duan, C.; Lan, Y. Comparison of Chemical and Sensory Profiles between Cabernet Sauvignon and Marselan Dry Red Wines in China. *Foods* **2023**, *12*, 1110. [CrossRef]
9. Morgani, M.B.; Fanzone, M.; Peña, J.E.P.; Sari, S.; Gallo, A.E.; Tournier, M.G.; Prieto, J.A. Late pruning modifies leaf to fruit ratio and shifts maturity period, affecting berry and wine composition in *Vitis vinifera* L. cv. ‘Malbec’ in Mendoza, Argentina. *Sci. Hortic.* **2023**, *313*, 111861. [CrossRef]
10. Stavrakaki, M.; Doudoumi, T.; Daskalakis, I.; Bouza, D.; Biniari, K. Effect of different viticultural techniques on the qualitative and quantitative characters of cv. Xinomavro under vineyard conditions in Naoussa. In *BIO Web of Conferences*; EDP Sciences: Les Ulis, France, 2023; Volume 56, p. 01023. [CrossRef]
11. Van Truong, N.; Khanh, T.Q. The Impact of Technology and Automation in Enhancing Efficiency, Quality, and Control in Modern Vineyards and Wineries. *J. Comput. Soc. Dyn.* **2023**, *8*, 1–14. Available online: <https://vectorial.org/index.php/JCSD/article/view/40> (accessed on 17 December 2023).
12. Guaita, M.; Bosso, A. Polyphenolic characterization of grape skins and seeds of four Italian red cultivars at harvest and after fermentative maceration. *Foods* **2019**, *8*, 395. [CrossRef]

13. Gomes, V.; Mendes-Ferreira, A.; Melo-Pinto, P. Application of hyperspectral imaging and deep learning for robust prediction of sugar and pH levels in wine grape berries. *Sensors* **2021**, *21*, 3459. [[CrossRef](#)] [[PubMed](#)]
14. Wang, D.; Cao, W.; Zhang, F.; Li, Z.; Xu, S.; Wu, X. A review of deep learning in multiscale agricultural sensing. *Remote Sens.* **2022**, *14*, 559. [[CrossRef](#)]
15. Zhang, N.; Liu, X.; Jin, X.; Li, C.; Wu, X.; Yang, S.; Ning, J.; Yanne, P. Determination of total iron-reactive phenolics, anthocyanins and tannins in wine grapes of skins and seeds based on near-infrared hyperspectral imaging. *Food Chem.* **2017**, *237*, 811–817. [[CrossRef](#)] [[PubMed](#)]
16. Gao, S.; Xu, J.H. Hyperspectral image information fusion-based detection of soluble solids content in red globe grapes. *Comput. Electron. Agric.* **2022**, *196*, 106822. [[CrossRef](#)]
17. Benelli, A.; Cevoli, C.; Ragni, L.; Fabbri, A. In-field and non-destructive monitoring of grapes maturity by hyperspectral imaging. *Biosyst. Eng.* **2021**, *207*, 59–67. [[CrossRef](#)]
18. Zhang, J.; Lei, Y.; He, L.; Hu, X.; Tian, J.; Chen, M.; Huang, D.; Luo, H. The rapid detection of the tannin content of grains based on hyperspectral imaging technology and chemometrics. *J. Food Compos. Anal.* **2023**, *123*, 105604. [[CrossRef](#)]
19. Rouxinol, M.I.; Martins, M.R.; Murta, G.C.; Mota Barroso, J.; Rato, A.E. Quality Assessment of Red Wine Grapes through NIR Spectroscopy. *Agronomy* **2022**, *12*, 637. [[CrossRef](#)]
20. Chen, X.; Jiao, Y.; Liu, B.; Chao, W.; Duan, X.; Yue, T. Using hyperspectral imaging technology for assessing internal quality parameters of persimmon fruits during the drying process. *Food Chem.* **2022**, *386*, 132774. [[CrossRef](#)] [[PubMed](#)]
21. Gao, S.; Wang, Q.; Fu, D.; Li, Q. Nondestructive detection of red grape sugar content and hardness by hyperspectral imaging. *J. Opt.* **2019**, *10*, 355–364.
22. Nogales-Bueno, J.; Baca-Bocanegra, B.; Rodríguez-Pulido, F.J.; Heredia, F.J.; Hernández-Hierro, J.M. Use of near infrared hyperspectral tools for the screening of extractable polyphenols in red grape skins. *Food Chem.* **2015**, *172*, 559–564. [[CrossRef](#)] [[PubMed](#)]
23. 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](#)]
24. 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.
25. 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.* **2023**, *17*, 100539. [[CrossRef](#)] [[PubMed](#)]
26. Zheng, C.; Abd-Elrahman, A.; Whitaker, V.M.; Dalid, C. Deep learning for strawberry canopy delineation and biomass prediction from high-resolution images. *Plant Phenomics* **2022**, *2022*, 9850486. [[CrossRef](#)]
27. Matteoli, S.; Diani, M.; Massai, R.; Corsini, G.; Remorini, D. A spectroscopy-based approach for automated nondestructive maturity grading of peach fruits. *IEEE Sens. J.* **2015**, *15*, 5455–5464. [[CrossRef](#)]
28. Fadock, M.; Brown, R.B.; Reynolds, A.G. Visible-near infrared reflectance spectroscopy for nondestructive analysis of red wine grapes. *Am. J. Enol. Vitic.* **2016**, *67*, 38–46. [[CrossRef](#)]
29. Zhou, X.; Liu, W.; Li, K.; Lu, D.; Su, Y.; Ju, Y.; Fang, Y.; Yang, J. Discrimination of Maturity Stages of Cabernet Sauvignon Wine Grapes Using Visible–Near-Infrared Spectroscopy. *Foods* **2023**, *12*, 4371. [[CrossRef](#)]
30. Saad, A.; Azam, M.M.; Amer, B.M. Quality analysis prediction and discriminating strawberry maturity with a hand-held Vis–NIR spectrometer. *Food Anal. Methods* **2022**, *15*, 689–699. [[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.