applied sciences

MDPI

*Article*

# Nondestructive Determination of Leaf Nitrogen Content in Corn by Hyperspectral Imaging Using Spectral and Texture Fusion

**Ling Wu [1], Yuanjuan Gong [1], Xiaoping Bai [2], Wei Wang [1,*] and Zhuo Wang [2,3,*]**

[1] College of Engineering, Shenyang Agricultural University, Shenyang 110866, China
[2] Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China
[3] Xiong'an Innovation Research Institute, Chinese Academy of Sciences, Baoding 071899, China
* Correspondence: syww@syau.deu.cn (W.W.); zwang@sia.cn (Z.W.)

**Abstract:** The nitrogen content is an important indicator affecting corn plants' growth status. Most of the standard hyperspectral imaging-based techniques for nondestructive detection of crop nitrogen content use a single feature as the input variable of the model, which reduces the generalization ability of the prediction model. To this end, a prediction model for the nitrogen content of corn leaves based on the fusion of image and spectral features is proposed. In this study, corn leaves at the modulation stage were studied, samples with different nitrogen levels were numbered, and their hyperspectral data in the wavelength range of 400~1100 nm were collected. The average spectrum of the models was used as valid spectral information. First-order derivatives, standard normal variables transformation (SNV), Savitzky-Golay (S-G) smoothing, and normalization were selected to preprocess the spectral features. The CARS-SPA algorithm was used to screen sensitive spectral variables. The gray level co-currency matrix (GLCM) was chosen to extract the texture image features of the test samples. Corn leaf spectral and texture image features were fused and modeled as target features. Partial least squares regression (PLSR) and support vector machine regression (SVR) were used to predict corn leaves' nitrogen content. The results showed that the image and spectral-based fusion models improved the prediction performance to some extent compared to the univariate models. The PLSR model based on feature fusion predicted the best results, in which the $R_P^2$ and RMSEP were 0.987 and 0.047. This method provides a reliable theoretical basis and technical support for developing nondestructive and accurate detection of nitrogen content in corn leaves.

**Keywords:** hyperspectral; fusion features; corn nitrogen content; machine learning; nondestructive detection

check for updates

**Citation:** Wu, L.; Gong, Y.; Bai, X.; Wang, W.; Wang, Z. Nondestructive Determination of Leaf Nitrogen Content in Corn by Hyperspectral Imaging Using Spectral and Texture Fusion. *Appl. Sci.* **2023**, *13*, 1910. https://doi.org/10.3390/app13031910

Received: 16 December 2022
Revised: 11 January 2023
Accepted: 18 January 2023
Published: 1 February 2023

**Copyright:** © 2023 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

With the continuous development of breeding technology, it has become difficult for traditional breeding methods to meet the demand for yield increase in food crops, and phenomics is gradually gaining attention. Phenotype assesses genotypic and environmentally determined agronomic traits [1]. Plant phenotyping is becoming an essential aspect of crop breeding as a critical technology for detecting plant growth and predicting yield [2]. Nitrogen, a crucial phenotypic trait of corn plants, is a component of vital organic compounds such as amino acids, proteins, nucleic acids, and chlorophyll. Nitrogen deficiency affects the synthesis of amino acids, chlorophyll, and other substances. It reduces their photosynthetic capacity [3], thus affecting the growth, quality, and yield of corn plants, so real-time monitoring of nitrogen in corn plants plays a vital role in their growth and development. The current chemical methods for detecting leaf nitrogen content (LNC) in corn are very mature. However, there are still problems, such as long detection time, tedious operations, and significant errors [4], which make it challenging to ensure the timeliness and practicality of the detection.

With the continuous development of non-destructive testing technology, hyperspectral imaging has been widely used to detect plant LNC. The absorption and reflection of LNC to the waveband affect the material composition and structural information inside the crop. Hyperspectral-based plant nitrogen detection obtains the spectral characteristics of plants and canopies by using hyperspectral sensors without damaging the crop tissues and structures and then obtains the nitrogen content surplus or deficit in the plant quickly and accurately by analyzing the spectral information of plant leaves or canopies [5]. Zhu et al. [6] quantitatively predicted the LNCs of winter wheat by constructing several LNC spectral indices (LNCSIs), and the results confirmed the validity of LNCSIs for winter wheat LNC prediction. Sun et al. [7] developed a regression model for rice LNC prediction by measuring its active and passive spectra. Raj et al. [8] proposed a LNC prediction model based on UAV hyperspectral images of the corn canopy. Li et al. [9] constructed a winter wheat LNC estimation model based on a multi-angle composite vegetation index by acquiring spectral data of multiple leaf inclination angles. The above studies demonstrate the feasibility of the inversion of plant nutrient elements based on spectral features. However, the single spectral feature has limitations for detecting the distribution pattern of plant nitrogen content. It cannot dynamically predict the internal plant nitrogen content, which has limitations for further improving the robustness of the prediction model.

Hyperspectral imaging is a combination of imaging and spectroscopic techniques. Although both spectral data and texture features can improve the accuracy of nitrogen nondestructive detection to some extent, a single consideration of image texture variables cannot predict the nitrogen content inside the plant. In contrast, a single consideration of spectral variables cannot characterize the overall spatial distribution of nitrogen [10,11]. Combining spectral and image texture information from optical hyperspectral data on plant leaf nitrogen can improve the accuracy of nitrogen prediction models and the generalization ability [12,13]. Yan et al. [14] constructed a chlorophyll content prediction model based on the fusion of spectral and texture features by comparing the back propagation neural network (BPNN) and support vector regression (SVR) algorithms, and its prediction set $R^2$ was 0.9571. In summary, the fusion of spectral and texture features is beneficial to alleviate the disadvantage of the low sensitivity of spectral analysis techniques and to improve the accuracy and robustness of nitrogen prediction models [15].

Algorithms play a decisive role in plant nitrogen prediction. The partial least squares (PLS) regression algorithm is an extension of the multiple linear regression model, which can effectively reduce the covariance between data variables and thus is applied extensively. On the other hand, the SVR algorithm can find the set of variables containing minor redundant information in the spectral information to minimize the covariance among variables, which is widely used in regression problems. Tan et al. [16] proposed a binary particle swarm optimization-support vector regression (BPSO-SVR) machine learning method, which was based on a binary particle swarm optimization algorithm, to predict soil nitrogen content with good prediction results. Bai et al. [17] proposed a PLS regression-based model for predicting the leaf nitrogen content of winter wheat with a validation set prediction accuracy of $R^2$ of 0.84. Wang et al. [18] used an aerial hyperspectral spectral imaging technique, and a partial least squares regression (PLSR)-based nitrogen content prediction model for corn plants was constructed with a prediction accuracy of $R^2$ of 0.85. Therefore, constructing plant prediction models based on PLSR and SVR algorithms has a certain reliability and research significance.

Few studies have been conducted to build plant inversion models based on feature fusion. In this study, we combined spectral imaging and image analysis techniques to predict LNC in corn accurately. We successively processed the spectral data using first-order derivatives, standard normal variance (SNV), and Savitzky-Golay (S-G) smoothing and normalization and extracted feature wavelengths using an improved fusion algorithm with a competitive adaptive reweighting algorithm (CARS) and a successive projection algorithm (SPA). The predictive regression model of the corn LNC was developed by fusing the spectral features and image texture features from hyperspectral data. In addition, we also compared the

application of PLSR and SVR modeling methods in corn LNC monitoring. Therefore, the fusion based on texture features and spectral features can break through the limitations of single-feature prediction, effectively obtain more comprehensive phenotypic characteristics of plants, improve the credibility of data, further enhance the accuracy of prediction models, and provide a new idea and method for detecting the growth pattern of plants.

## 2. Materials and Methods

### 2.1. Experimental Area

In this study, corn trials were conducted at the experimental base of the Shenyang Institute of Automation, Chinese Academy of Sciences. The site is located between 41°48′11.75″ N and 123°25′31.18″ E, with a temperate continental monsoon climate. This trial was conducted using the high-yielding and early maturing corn of Zhengdan 958 and was tested at the jointing stage. Corn leaf samples with the same leaf position (top three leaves), no disease on the leaf surface, no broken leaves, and good growth conditions were randomly selected from the test area. The number of samples was 82, as shown in Figure 1, collected in sealed food bags. All samples were numbered uniformly to avoid errors caused by human intervention and sent to the laboratory immediately for hyperspectral data collection.



**Figure 1.** The samples after drying treatment.

### 2.2. Experimental Data

The data collection experiment was conducted on 22 July 2022, which was a windless, bright day and corresponded to the corn elongation stage. The specific data collection time was extended from 10:00 a.m. to 2:00 p.m. The acquired data of the experiment consisted of hyperspectral data measured at close range and LNC data.

#### 2.2.1. Hyperspectral Data

The hyperspectral imaging system was responsible for acquiring hyperspectral data from the corn leaves, consisting of a hyperspectral camera (Specim Fx10e), a three-axis stage, a constant light source, and a computer and control software, as shown in Figure 2. The whole spectral system was in a closed dark box to reduce the influence of environmental light sources. During the experiments, the hyperspectral images were acquired using the Fx10e hyperspectral camera from Specim, Finland, with a spectral band range of 400~1000 nm, containing 224 bands, a spectral resolution of 5.5 nm, and a spatial pixel count of 1024 px. The camera was mounted on a three-axis platform to obtain stable and accurate hyperspectral data. The acquisition software was the Lumo Recorder 2020 software provided by the manufacturer. In order to acquire clear images, the hyperspectral camera focal length, the distance between the lens and the halogen lamp and between the halogen lamp and the measured object, the scanning speed, and the exposure time needed to be adjusted. After several pre-experiments, the distance between the lens and the halogen lamp was 25 cm, the distance between the halogen lamp and the object to be measured was 30 cm, the scanning speed was 20 mm/s, and the exposure time was 20 ms.

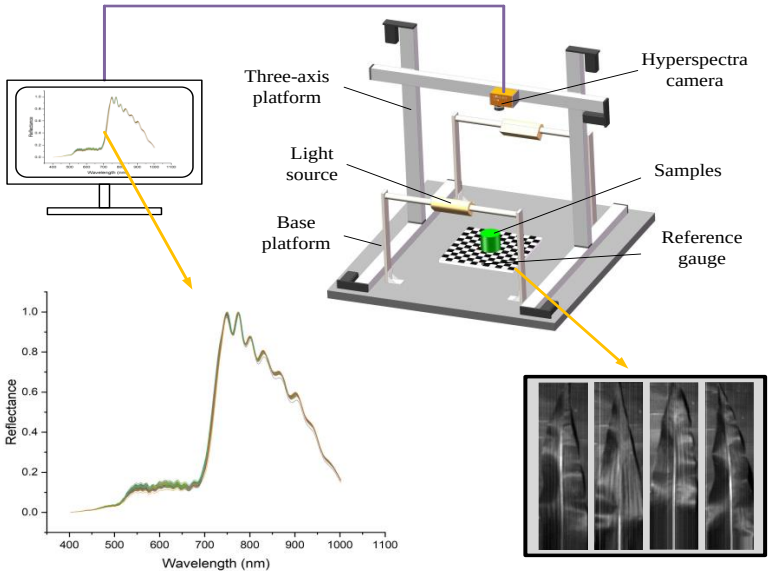



**Figure 2.** The structure of the hyperspectral imaging system.

In order to reduce the impact of camera dark current and light intensity variations on the hyperspectral data [19], the acquired data were corrected for black and white to achieve reflectance correction, and the correction equation is:

$$R = \frac{I_{\text{raw}} - I_{dark}}{I_{white} - I_{dark}} \quad (1)$$

where $R$ is the corrected relative reflectance image, $I_{raw}$ is the sample image, $I_{dark}$ is the all-black image, and $I_{white}$ is the white image obtained from a standard white plate with 99% reflectance.

### 2.2.2. LNC Acquirement

After the spectral data measurement of each corn leaf, the test samples were dried in an oven at 105 °C for 30 min and then at 80 °C for 24 h to completely remove the water from the leaves to reach a constant weight. The dried samples were then weighed, crushed, and measured with a Kjeldahl apparatus to obtain the true values of leaf nitrogen content in corn, as shown in Figure 3.

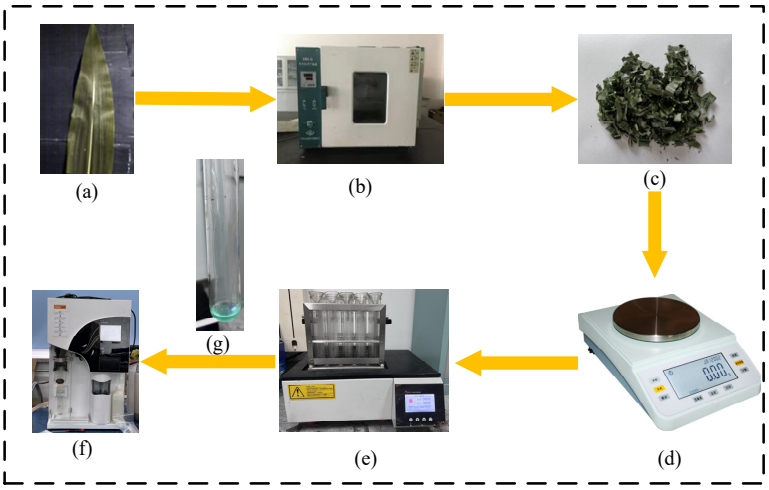



**Figure 3.** (**a**) Corn leaf samples; (**b**) dryer; (**c**) grinded sample shreds; (**d**) precision electronic balance; (**e**) graphite disintegrator; (**g**) dissolved test solutions; (**f**) Kjeldahl.

### *2.3. Research Methods*

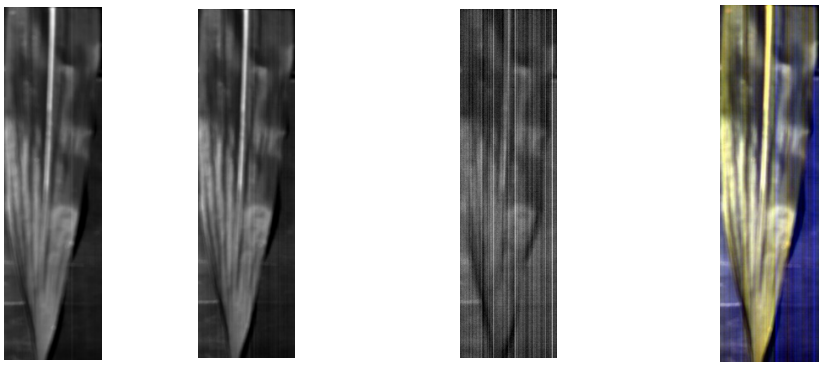
#### 2.3.1. Preprocessing of Hyperspectral Data

The changes of LNC in corn induce fluctuations in their sharp spectral curves [20]. The black- and white-corrected hyperspectral images were used to select the actual area of leaf nitrogen content using ENVI 5.3 software, and the average spectral curve was derived for the pixel points of the entire sample leaf. In order to eliminate the effects of scattered light, noise, and baseline drift during the acquisition of leaf spectral information [21], the hyperspectral data were preprocessed. Moreover, they were processed sequentially using first-order derivatives, SG smoothing [22], SNV [23], and standardization. The best preprocessing method was also modeled by the leave-one-out cross-validation method. The cross-prediction set coefficient of determination ($R^2$) and the cross-prediction set root mean square error (RMSE) were used as evaluation indicators to screen the best preprocessing method.

In addition, the first derivative could deduct the effect of the environmental background or drift on the signal and improve the spectral resolution. SG smoothing uses polynomials to decompose the data within the moving window of the original spectrum and fit the data with least squares to eliminate random noise in the spectral signal and improve the signal-to-noise ratio of the sample signal. SNV is mainly used to eliminate the effects of uneven particle distribution and scattered light generated by different particle sizes on the spectrum. Standardization is used to eliminate the undesirable effects of excessive differences in data scales.

#### 2.3.2. Analysis of Corn Image Features

The gray-level images at a wavelength of 700.69 nm for red light (R), 546.39 nm for green light (G), and 435.59 nm for blue light (B) were selected to be fused into one RGB image in ENVI 5.3 by combining the degree of spectral separation and band correlation methods, as shown in Figure 4. The texture features were extracted from the fused RGB images using GLCM [24], shown in Figure 5.



(**a**) (**b**) (**c**) (**d**)

**Figure 4.** Grayscale image and fused color image R, G, B wavelength; (**a**) R, grayscale image; (**b**) G, grayscale image; (**c**) B, grayscale image; (**d**) fused color image.

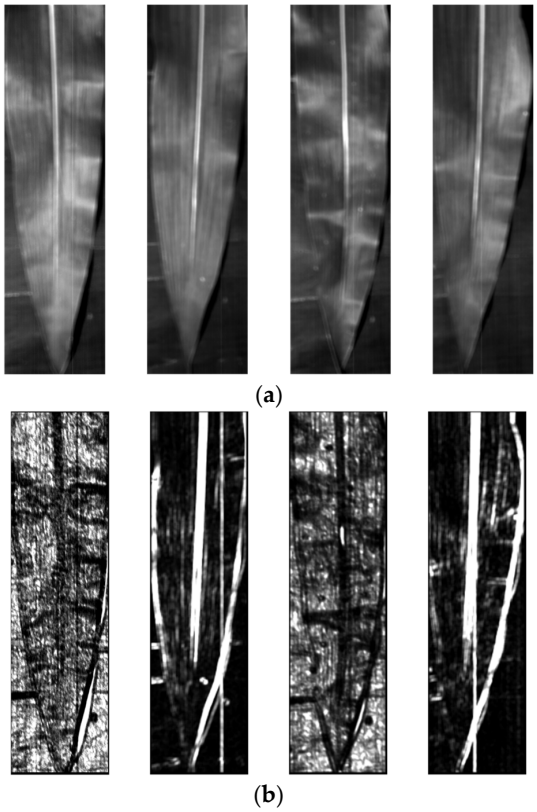
This texture extraction method was based on a gray-level spatial correlation matrix, representing the joint distribution of two gray pixel levels with a spatial location relationship. The grayscale co-occurrence matrix is the probability of simultaneous occurrence of pixels $(x_2, y_2)$ with a distance of $\alpha$, an azimuth of $\theta$, and a grayscale value of $j$, starting from a pixel $(x_1, y_1)$ with a grayscale of $i$ on the image $P(i, j, \alpha, \theta)$.

$$P(i, j, \alpha, \theta) = \{(x_1, y_1), (x_2, y_2) | f(x_1, y_1) = i, f(x_2, y_2) = j\} \quad (2)$$

where $i$ is the gray value of the image at pixels $(x_1, y_1)$, $j$ is the gray value of the image at pixels $(x_2, y_2)$, $P$ is the matrix $N \times N$, and $\alpha$ represents the generation step, which is the

distance between two pixels points, and $\theta$ represents the generation direction, which is usually taken as 0°, 45°, 90°, 135°.

The grayscale co-occurrence matrix visually describes the texture features of the samples by extracting some statistical quantities. In this study, we selected four unrelated statistical feature quantities: entropy, energy, contrast, and inverse different moment, shown in Table 1.



(**a**)

(**b**)

**Figure 5.** Results after GLCM: (**a**) original data; (**b**) feature images.

**Table 1.** Summary of texture features.

| Feature | Description | Equation |
| --- | --- | --- |
| Entropy | A measure of the image's randomness reflects the texture's complexity or non-uniformity. | $Ent = -\sum_i \sum_j P(i,j,\alpha,\theta) \log P(i,j,\alpha,\theta)$ |
| Energy | Reflecting the uniformity of the image grayscale distribution and the degree of texture coarseness. | $Energy = \sum_i \sum_j P(\mathrm{i},\mathrm{j},\alpha,\theta)$ |
| Contrast | Reflecting the clarity of the image and the depth of the texture grooves. | $Con = \sum_i \sum_j (i-j)^2 P(i,j,\alpha,\theta)$ |
| Inverse Different Moment | Reflecting the degree of local variation of image texture. | $Idm = \sum_i \sum_i \frac{P(i,j,\alpha,\theta)}{1+(i-j)^2}$ |

A correlation analysis of four image eigenvalues with the total LNC of corn was performed using IBM SPSS Statistics 23 software.

### *2.4. Establishment and Evaluation of Regression Models*

In this study, the spectral features of corn leaf samples and image texture features were fused to design LNC prediction models using PLSR and SVR algorithms. The prediction performance of the two algorithms and the three models based on different variable features were compared. The specific process is as follows: (1) the correlation analysis between LNC and spectral reflectance of each band is performed by the CARS-SPA algorithm to select the LNC-sensitive band; (2) the image texture features of corn leaf samples are extracted using the GLCM method; (3) the sensitive band, image texture features, and LNC are normalized between [−1,1], respectively, to eliminate the error caused by the difference in the size of variables; (4) the normalized spectral features, image texture features, and fusion features are used as the inputs for PLSR and SVR methods, respectively, to optimize each model parameter; (5) each LNC prediction model is established using the optimal parameters; (6) the models are validated based on the validation samples, and the optimal LNC prediction model is selected, as shown in Figure 6.

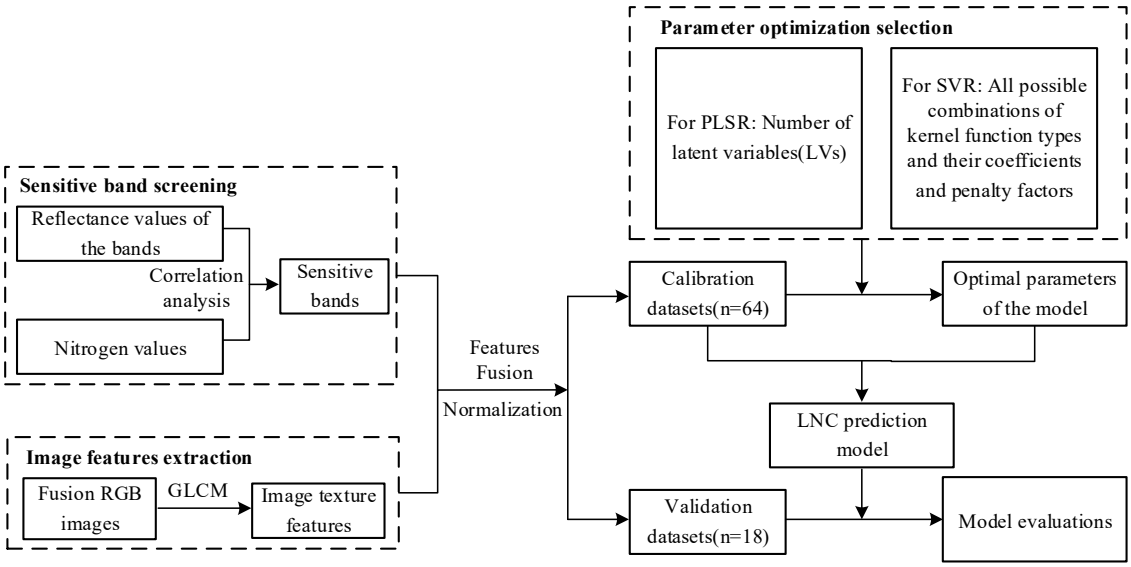



**Figure 6.** Flow chart of the LNC prediction model design based on spectral and image fusion methods.

#### 2.4.1. Normalization

Two types of data with different properties, spectral information, and texture features of corn leaves were normalized as shown in Figure 7, so that they were scaled to a particular proportion, and the fused data were used as input variables for the PLSR and SVR prediction models.

#### 2.4.2. Sensitive Bands Selection

The fusion algorithm of CARS and SPA was used to extract feature wavelengths for corn leaf samples to reduce the influence of redundant information and noise in the spectral data on model accuracy.

The CARS algorithm [25] combines Monte Carlo sampling and PLS model regression coefficients to select the feature variables, following the principle of "survival of the fittest." The wave points with short wavelengths are selected by setting the number of cyclic iterations for cyclic calculation. The optimal combination of variables is filtered according to the subset with the smallest root mean square error of interaction verification (RMSECV). The number of loop sampling is set to 30 in this experiment to screen out the variables with small covariance and significant weight.

The SPA algorithm [26], widely used in screening and extracting feature wavelengths, is a method for an iterative search of forwarding feature variables, thereby enabling the

screening of variable combinations with the least redundant information and the least covariance. The SPA algorithm uses the projection analysis of vectors and projects wavelengths onto other wavelengths and, by comparing the magnitude of the projection vectors, selects the wavelength with the most significant projection vector as the wavelength. The final feature wavelength is selected based on the correction model. The SPA uses RMSE as an evaluation index and determines the number of final feature variables based on the smaller RMSE value.

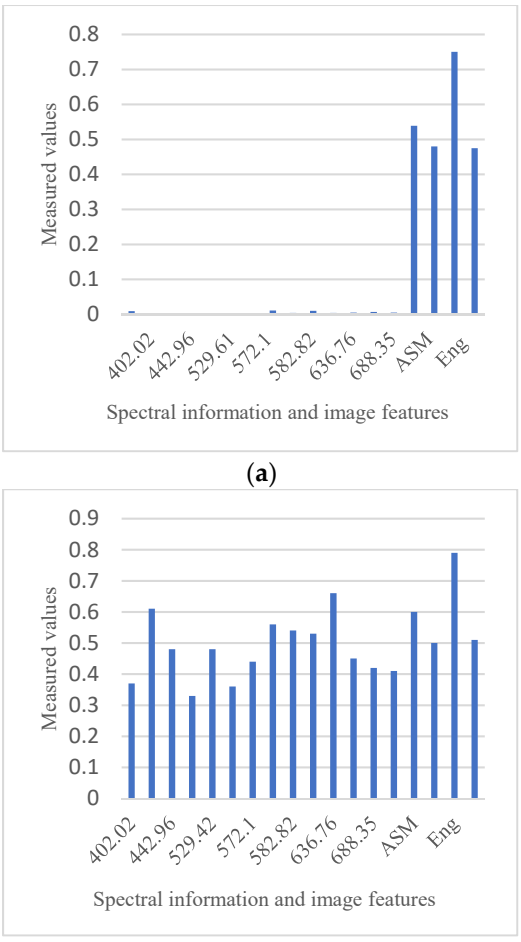



**Figure 7.** Data fusion comparison: (**a**) before normalization; (**b**) after normalization.

The fusion of CARS and SPA algorithms can effectively eliminate the spectral space vector covariance, reduce the redundant information between feature variables, and select representative feature wavelength variables for modeling. Furthermore, it can improve the accuracy and efficiency of the model, and can more intuitively reflect the response of spectral information to the change of leaf nitrogen content.

#### 2.4.3. PLSR

The PLSR algorithm [27] is a novel approach to multivariate statistics that combines the features of multiple linear regression analysis, typical correlation analysis, and principal component analysis. PLSR provides a regression modeling approach for studying multiple dependent variables on multiple independent variables and is used to address the interdependence between multiple correlated variables.

It is assumed that $p$ variable s $y_1, y_2, y_3, \ldots, y_p$ and $q$ independent variables $x_1, x_2, x_3, \ldots, x_q$ are standardized variables. The $n$ standardized observations of the dependent and independent variables are denoted as

$$F_0 = \begin{pmatrix} y_{11} & \cdots & y_{1p} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{np} \end{pmatrix}, E_0 = \begin{pmatrix} x_{11} & \cdots & x_{1q} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nq} \end{pmatrix} \tag{3}$$

The r components $t_1, t_2, t_3, \ldots, t_r$ $(r \leq q)$ are extracted from the set of independent variables, and the partial least squares regression is represented by establishing the regression equation of $y_1, y_2, y_3, \ldots, y_p$ with $t_1, t_2, t_3, \ldots, t_r$, which in turn is expressed as the regression equation of $y_1, y_2, y_3, \ldots, y_p$ with the independent variable $x_1, x_2, x_3, \ldots, x_q$. That is the partial least squares regression equation for the dependent variable:

$$y_j = a_{j1}x_1 + \ldots + a_{jq}x_q, (j = 1, 2, 3, \ldots, q) \tag{4}$$

2.4.4. SVR

The SVR algorithm [28] is a supervised learning model for regression analysis. The mapping of sample data onto a high-dimensional data features space by determining a suitable kernel function enables an excellent linear regression feature of the independent and dependent variables in the high-dimensional data to feature space with good generalization ability.

SVR minimizes the total deviation of the sample data to the hyperplane by constructing a hyperplane, which is calculated as follows:

$$f(x) = w^T x + b \tag{5}$$

$$\begin{cases} \min \frac{1}{2}\|w\|^2 \\ s.t. |y_i - (wx_i + b)| \leq \varepsilon, i = 1, 2, 3 \ldots n. \end{cases} \tag{6}$$

where $w$ is the normal vector, which determines the direction of the hyperplane; $b$ is the displacement term, which determines the distance between the hyperplane and the origin; and $y_i$ is the actual sample value. Therefore, the optimal $f(x)$ is obtained by minimizing the normal vector $w$ so that the error of the predicted value of the sample data is less than $\varepsilon$.

SVR selected the sensitive wavelength variables for the CARS-SPA algorithm to build a predictive nitrogen model. The kernel function of SVR used a radial basis function, and the penalty factor and kernel parameters were optimized by ten-fold cross-validation.

2.4.5. Statistical Analysis Methods

The accuracy and stability of the model were determined by using the leave-one-out cross-validation method of modeling, with the coefficient of determination $R^2$ and root mean square error (RMSE) as evaluation indicators. Ref. [29] is calculated as follows:

$$R^2 = \left( \sum_{i=1,j=1}^{n} y_i - \overline{y} \right)^2 \Big/ \left( \sum_{i=1}^{n} x_i - \overline{y} \right)^2 \tag{7}$$

$$RMSE = \sqrt{\sum_{i=1,j=1}^{n} (x_i - y_i)^2 / n} \tag{8}$$

where $x_i$ is the measured value of leaf nitrogen content in corn, $y_i$ is the predicted value of nitrogen content, $\overline{y}$ is the average nitrogen content, and $n$ is the number of samples.

Meanwhile, to better evaluate the accuracy and stability of the model, ten repetitive simulation tests are used in this test, and the final results of the model evaluation are the average of ten repetitive tests.

## 3. Results and Discussion

### *3.1. Nitrogen Measurement*

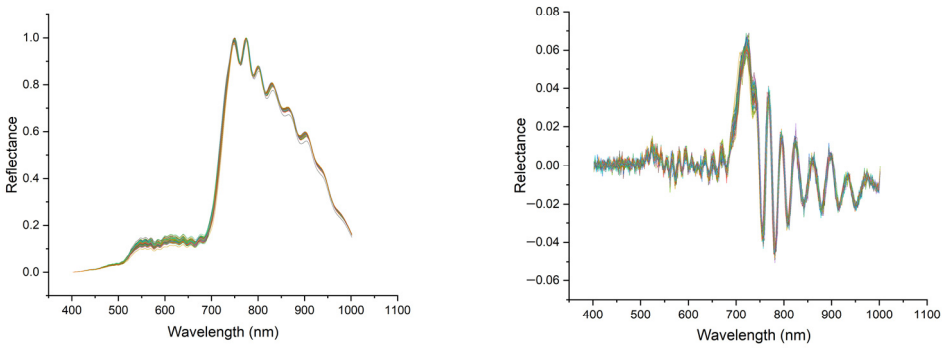
In this experiment, we collected 82 corn leaf samples, randomly divided into two data sets: 80% (64) as the training set and the remaining 20% (18) as the test set. Table 2 shows the total statistics of measured LNC values. The LNC data of the calibration dataset ranged from 0.65 to 2.57, with a mean of 1.63 and a standard deviation of 0.43. Similarly, the statistical parameters of the test dataset were 1.12~2.69, 1.69, and 0.4, respectively.

**Table 2.** Statistical analysis of LNC data.

| Dataset | Sample | Max | Min | Mean | SD |
|---|---|---|---|---|---|
| Calibration dataset | 64 | 2.57 | 0.65 | 1.63 | 0.43 |
| Validation dataset | 18 | 2.69 | 1.12 | 1.69 | 0.40 |

### *3.2. Spectral Features and Preprocessing*

The raw spectral data of all samples are shown in Figure 8a. Figure 8a shows that the degree of response of the spectral curves of corn with different leaf nitrogen contents varied. Still, the overall trend of changes showed the same regularity. Corn is sensitive to changes in nitrogen content, and leaves' spectral absorption characteristics are mainly influenced by various pigments and cellular tissue structures in the wavelength region [30]. Due to background noise, scattered light, and baseline drift during the acquisition process, there was a certain amount of noise in the spectral profile. To screen the optimal preprocessing method, a quantitative analysis of each preprocessing method was performed, and the preprocessing was modeled and analyzed using the partial least squares method. It is shown in Table 3 that the first-order derivative preprocessing had the best processing effect with wavelength range from 400 to 1000 nm, the $R_P^2$ of the PLSR model was 0.84, and the RMSEP was 0.17. As seen in Figure 8b, the spectra after the first-order derivative processing eliminated the interference of the baseline and background and improved the spectral resolution.




(**a**) (**b**)

**Figure 8.** (**a**) Original spectral curves; (**b**) first derivative curves.

The selection of spectrally sensitive variables affects the results and performance of the model, with the ultimate goal of selecting the few combinations of characteristic wavelengths from the original bands that are informative, less correlated, and have good class separability [31]. A joint algorithm of CARS and SPA was used for the pre-processed spectral data to filter the sensitive wavelengths from the full spectrum of 224 variables. CARS sets the number of Monte Carlo sampling runs to 50, while tenfold cross-validation is used to evaluate the effect of each subset, and the spectral feature variables are selected for the first-order derivative preprocessed data, as shown in Figure 9. From Figure 9a, it can be seen that the number of variables decreases rapidly in 25 sampling runs due to the

exponential decay function EDP, and then gradually slows down and stabilizes, which indicates that the CARS algorithm has "coarse selection" and "selection" in the feature variable selection. This process indicates that the CARS algorithm has two processes: "coarse selection" and "selection" in feature selection. As can be seen in Figure 9b, in the initial stage, the ten-fold cross-validation RMSECV values of the individual PLS models became progressively smaller as the number of iterations increased due to the exclusion of a large number of variables that were not relevant to the prediction of nitrogen content in corn leaves. When the RMSECV reached the minimum value, the corresponding number of sampling was 24. With a further increase in the number of sampling, the RMSECV also gradually increased, indicating that some crucial variables in the spectrum were eliminated. Therefore, it can be seen in Figure 9c that when the variables obtained in 24 iterations are identified as the characteristic variables for predicting the nitrogen content of corn leaves, a total of 24 variables.

**Table 3.** Comparison of partial least squares regression (PLSR) model effects of pretreatment methods.

| Spectral Wavelength/nm | Pretreatment Method | $R_p^2$ | RMSEP |
|---|---|---|---|
| 400~1000 | Original spectrum | 0.80 | 0.19 |
| | First derivative | 0.84 | 0.17 |
| | S-G | 0.72 | 0.23 |
| | SNV | 0.80 | 0.19 |
| | Normalization | 0.75 | 0.22 |

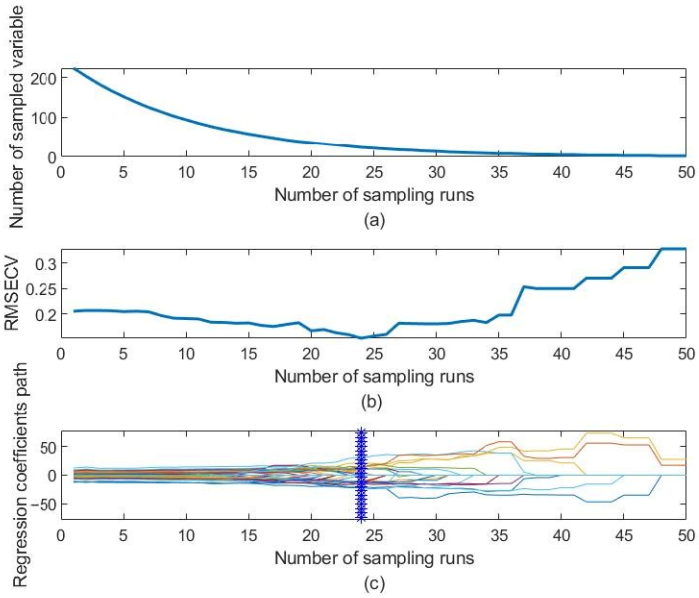



**Figure 9.** CARS method screening process of characteristic wavelength. (**a**) The change rule of sample variable quantity under different sampling times; (**b**) Variation of RMSECV under different sampling times; (**c**) Change trend of regression coefficient of each wavelength with sampling times.

Based on the CARS screening wavelengths, another feature variable screening was performed using the SPA algorithm. The SPA algorithm screened the 14 reflected wavelengths with the lowest covariance of leaf nitrogen content as feature wavelengths. The screened wavelengths were correlated with maize LNC as shown in Figure 10. The results showed that the wavelengths at 505.61 nm, 572.1 nm, and 732.01 nm showed a weak correlation with LNC, while the other wavelengths showed a strong correlation. Moreover, most of the screened feature wavelengths were distributed in the nitrogen-sensitive band range, further indicating the accuracy of the CARS-SPA algorithm based on the extraction of sensitive bands of the spectrum, which showed strong feasibility for predicting maize LNC

using spectral information and could well invert the intrinsic relationship between spectral information and LNC. Hence, selecting feature wavelengths is significant for predicting corn LNC.

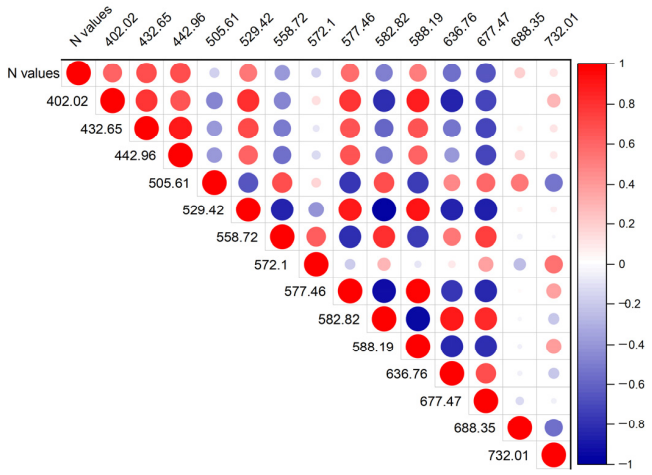



**Figure 10.** Correlation between selected sensitive spectral bands and *n* values.

### *3.3. Results of the Description Models on Spectral Features*

The 14 characteristic wavelengths and full spectral bands screened by the CARS-SPA method were inputted as parameters into the SVR and PLSR algorithms to construct a prediction model for the corn LNC, and the results are shown in Figure 11.

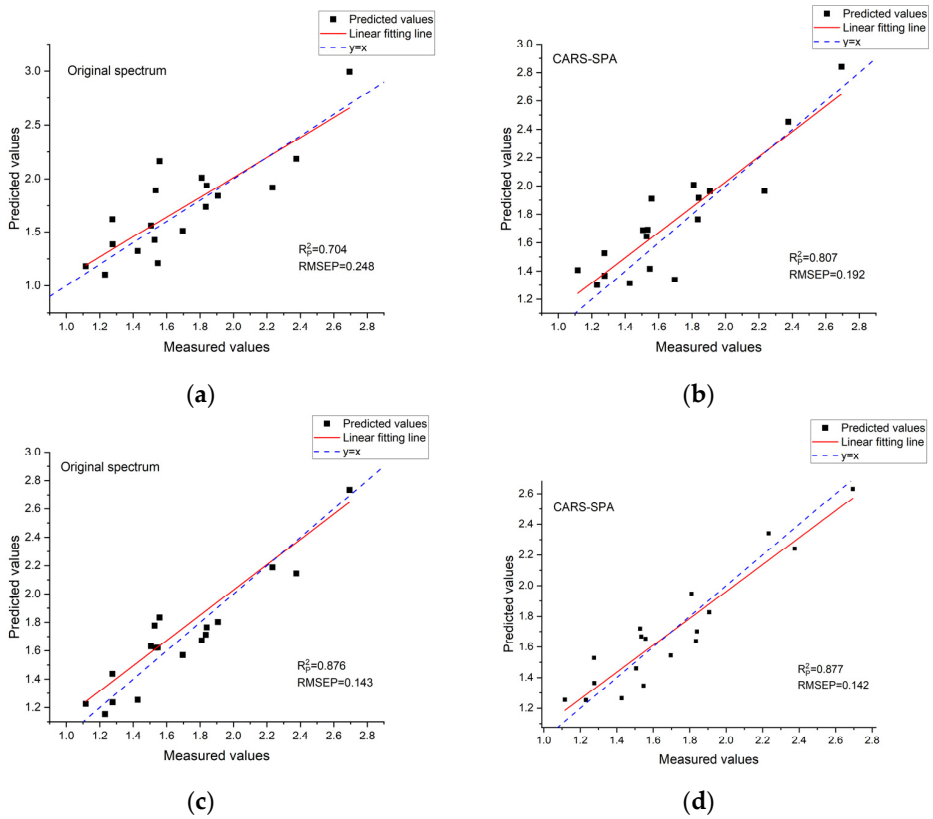



**Figure 11.** Accuracy of the measured and predicted values of the validation set: (**a**) full spectrum -SVR model; (**b**) characteristic wavelength-SVR model; (**c**) full spectrum-PLSR model; (**d**) Characteristic wavelength-PLSR model.

As seen in Figure 11, in the SVR model, the $R_P^2$ of the full-spectrum prediction model was 0.704, and the RMSEP was 0.248; the $R_P^2$ of the prediction model with the characteristic wavelength as the input variable was 0.807 and the RMSEP is 0.192. In the PLSR model, the $R_P^2$ of the full-spectrum model was 0.876, and the RMSEP was 0.143. The accuracy of nitrogen content prediction in both models was higher than that of the full-spectrum model because the use of the CARS-SPA algorithm could effectively reduce data redundancy and model processing complexity, remove the interference of a large amount of useless information on the correlation features, better represent the degree of light absorption by nitrogen groups in corn leaves, and improve the robustness of the model.

Figure 11b,d shows the results of the prediction models based on sensitive wavelengths. It can be seen that the PLSR model has a better fit between the predicted and actual values with an improved $R_P^2$ of 0.1 and an improved RMSEP of 0.05, so the prediction based on the PLSR model was better than the SVR model. Table 4 shows the training results based on different features and models. It can be seen that the PLSR model outperforms the SVR model in terms of the complete spectral band as well as the characteristic wavelength as input variables, indicating that the PLSR model can further eliminate the influence of covariance variables on the model prediction and improve the model generalization ability.

**Table 4.** Training results of various models based on full waveband and characteristic wavelength.

| Algorithm | Feature Types | Calibration Set ($n$ = 64) | | Validation Set ($n$ = 18) | |
|---|---|---|---|---|---|
| | | $R^2_C$ | RMSEC | $R^2_P$ | RMSEP |
| SVR | Ref | 0.979 | 0.063 | 0.704 | 0.248 |
| | CARS-SPA | 0.906 | 0.132 | 0.807 | 0.192 |
| PLSR | Ref | 0.882 | 0.149 | 0.876 | 0.143 |
| | CARS-SPA | 0.905 | 0.134 | 0.877 | 0.142 |

### *3.4. Texture Features*

Hyperspectral images were a combination of image and spectral techniques that reflect not only the spectral information of the target but also the image information [32]. Grayscale images at R, G, and B bands were obtained by measuring the brightness of each pixel within a single electromagnetic spectrum [33], and the nitrogen affected the absorption of light by the corn leaves. It led to different grayscale distributions in the pictures of corn obtained by the camera with different leaf nitrogen content. Therefore, there was an internal link between the picture characteristics of corn and their leaf nitrogen content.

As seen in Figure 12, the correlation coefficients of these four feature values with the nitrogen content were all above 0.8 and highly correlated.

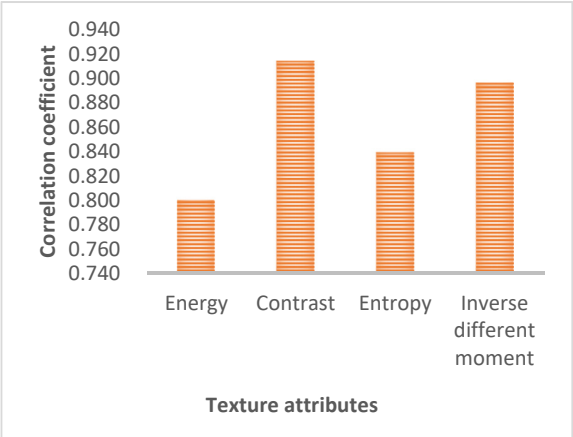
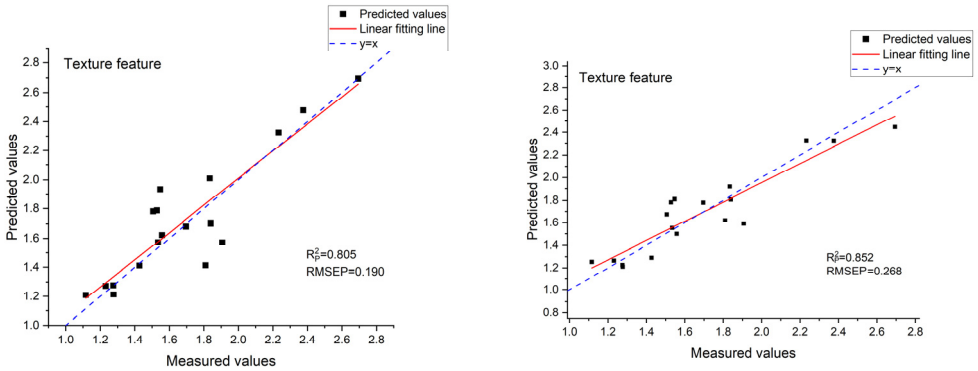



**Figure 12.** Correlation coefficient between corn leaf texture attributes and nitrogen values.

### *3.5. Results of the Description Models on Textural Features*

The results of the prediction model based on texture features and different machine learning algorithms in the prediction set are shown in Figure 13. Among them, the $R_P^2$ of the SVR prediction model based on image features was 0.805, and the RMSEP was 0.19; the $R_P^2$ of the PLSR model was 0.852, and the RMSEP was 0.268. To a certain extent, the prediction effect based on image features was slightly lower than that of the prediction model based on spectral information. It was because the spectral information has higher sensitivity and more obvious absorption for nitrogen in the green band and red-edge region. From the prediction results, although the prediction effect based on image features is not as good as the spectral data, the $R_P^2$ was above 0.80. Moreover, the texture image features were able to compensate for the deficiency of spectral information to a certain extent, which strongly proved image texture's feasibility in predicting corn LNC and has some research values.



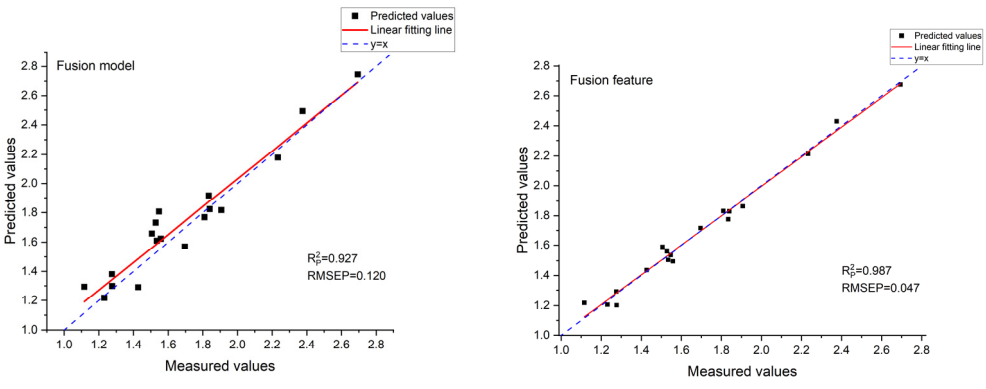

(**a**) (**b**)

**Figure 13.** Prediction of corn leaf nitrogen content by models SVR (**a**) and PLSR (**b**) based on image texture features.

### *3.6. Results of the Description Models on Feature Data Fusion*

The collected spectral data and image features were normalized to the range of 0~1, and the spectral data and image features with different properties were transformed into pure quantities for data fusion. The fusion of spectral and texture data were able to improve not only the model's convergence speed but also the model's prediction accuracy.

The spectral features and texture features screened by CARS-SPA were fused as input variables of the SVR and PLSR models to build a prediction model about the LNC of corn. The model was run iteratively to obtain the final results, shown in Figure 14.




(**a**) (**b**)

**Figure 14.** Leaf nitrogen content of corn predicted by the models SVR (**a**) and PLSR (**b**) based on fusion features.

As can be seen in Figure 14, the $R_P^2$ of the prediction set of the SVR regression model based on fusion features was 0.927, the RMSEP was 0.12, the $R_P^2$ of the PLSR model was 0.987, and the RMSEP was 0.047. The validation sample of the SVR model was basically around the 1:1 line except for a few points of deviation, while the data points of the PLSR model were almost close to the 1:1 line. The PLSR model had a better prediction performance.

We can see from Table 5 that the $R_P^2$ of the prediction set of both models based on fused features was significantly improved. In particular, the PLSR regression model based on fused features improved the prediction set $R_P^2$ by 0.11 compared to the model with single spectral information and 0.135 compared with the model with a single texture feature. The model has improved the accuracy and generalization ability by taking into account the spectral information and texture features and has improved the phenomenon of "the same objects with the different spectrum, and the same spectrum with different objects" in previous studies only for single spectral information features.

**Table 5.** Predicting results of various characteristic models for leaf nitrogen content in crop.

| Model | Original Spectrum | | Spectral Features | | Texture Features | | Fusion Features | |
|---|---|---|---|---|---|---|---|---|
| | SVR | PLSR | SVR | PLSR | SVR | PLSR | SVR | PLSR |
| $R_P^2$ | 0.704 | 0.876 | 0.807 | 0.877 | 0.805 | 0.852 | 0.927 | 0.987 |
| RMSEP | 0.248 | 0.143 | 0.192 | 0.142 | 0.190 | 0.268 | 0.120 | 0.047 |

Meanwhile, in terms of the prediction effect of each feature, the PLSR-based model outperforms the SVR model. The PLSR algorithm decomposed and filtered the variable information in a specific way, extracted the most explanatory composite variable in the dependent variable, and identified the information and noise in the model. Thus, it is able to better overcome the adverse effects of multiple correlations of variables in the model, have a specific anti-interference ability, and improve the generalization and robustness of the prediction model.

## 4. Conclusions

In this paper, we propose a new hyperspectral imaging technique for predicting the LNC of corn. Before building the prediction model, we introduce the GLCM method for extracting texture features, a matrix function of the distance and angle of different pixel points, using the joint probability distribution of the simultaneous occurrence of different grayscale pixel points to reflect texture information. It has good results in identifying texture features of corn leaves, improving the prediction accuracy of the spectral image model, and reducing the complexity of the model. In this paper, a fusion of texture features and spectral features is used for modeling. Comparing the eight models, the PLSR model based on fused features has the best prediction of the nitrogen content of corn leaves ($R_P^2$ = 0.987, RMSEP = 0.047). The PLSR model is an extension of the multiple linear regression model, which can effectively overcome the covariance problem of hyperspectral data and retain the original spectral information. It improves the accuracy and generalization of nitrogen content prediction in corn leaves. Compared with articles in the literature, the advantage of this study is that the combination of spectra and images using the expandability of hyperspectral techniques will be a promising method for predicting plant nutrient elements. Therefore, the following work will expand the study to a broader range of nutrient elements, which has far-reaching implications for modeling the growth dynamics of corn.

**Author Contributions:** Conceptualization, Z.W., Y.G., W.W.; methodology, L.W.; software, L.W.; validation, L.W.; formal analysis, L.W.; investigation, L.W., X.B.; resources, X.B.; data curation, L.W.; writing-original draft preparation, L.W.; writing-review and editing, Y.G., Z.W.; visualization, L.W.; supervision, Z.W.; project administration, Z.W.; funding acquisition, Z.W., W.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was financially supported by the "Revitalizing Liaoning Talents Plan" of Liaoning, funding number XLYC1902006.

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

**Informed Consent Statement:** No applicable.

**Data Availability Statement:** The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to their use in subsequent studies.

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

## References

1. Liu, Z.; Zhang, F.; Ma, Q.; An, D.; Li, L.; Zhang, X.; Zhu, D.; Li, S. Advances in crop phenotyping and multi-environment trials. *Front. Agric. Sci. Eng.* **2015**, *2*, 28–37. [CrossRef]
2. Tester, M.; Langridge, P. Breeding technologies to increase crop production in a changing word. *Science* **2010**, *327*, 818–822. [CrossRef] [PubMed]
3. Zhao, M.; Sun, X.; Wang, D. A review of rice hyperspectral remote sensing monitoring research. *Acta Agric. Univ. Jiangxi* **2019**, *41*, 1–12.
4. Song, L.; Ye, J.; Zheng, Y. Research Progress on Nondestructive Rapid Nutrition Diagnosis of Crop Nitrogen. *China Rice* **2017**, *23*, 19–22.
5. Jia, L.; Chen, P.; Zhang, S. Non-destructive testing techniques for crop nitrogen nutrient diagnosis. *World Agric.* **2010**, *06*, 36–37.
6. Zhu, C.; Liu, Y. UAV-based hyperspectral analysis and spectral indices constructing for quantitatively monitoring leaf nitrogen content of winter wheat. *Appl. Opt.* **2018**, *57*, 7722–7732. [CrossRef]
7. Sun, J.; Yang, J.; Shi, S. Estimating Rice Leaf Nitrogen Concentration: Influence of Regression Algorithms Based on Passive and Active Leaf Reflectance. *Remote Sens.* **2017**, *9*, 951. [CrossRef]
8. Raj, R.; Walker, J.P.; Pingale, R. Leaf nitrogen content estimation using top-of-canopy airborne hyperspectral data. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *104*, 102584. [CrossRef]
9. Li, T.; Zhu, Z.; Cui, J. Monitoring of leaf nitrogen content of winter wheat using multi-angle hyperspectral data. *Int. J. Remote Sens.* **2021**, *42*, 4672–4692. [CrossRef]
10. Shun, J.; Jin, M.; Mao, P. Detection of nitrogen content in lettuce leaves based on spectroscopy and texture using hyperspectral imaging technology. *Trans. Chin. Soc. Agric. Eng.* **2014**, *30*, 167–173.
11. Mishra, P.; Cordella, C.B.Y.; Rutledge, D.N. Application of independent components analysis with the JADE algorithm and NIR hyperspectral imaging for revealing food adulteration. *J. Food Eng.* **2016**, *168*, 7–15. [CrossRef]
12. Zhang, C.; Guo, C.; Liu, F. Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine. *J. Food Eng.* **2016**, *179*, 11–18. [CrossRef]
13. Cheng, W.; Sun, W.; Pu, H. Integration of spectral and textural data for enhancing hyperspectral prediction of K value in pork meat. *LWT-Food Sci. Technol.* **2016**, *72*, 322–329. [CrossRef]
14. Yan, Z.; Wang, Y.; Wu, Y. Detection of chlorophyll content of Epipremnum aureum based on fusion of spectrum and texture features. *J. Nanjing Agric. Univ.* **2021**, *44*, 568–575.
15. Zu, Q.; Zhang, F.; Cao, Y. Research on identification of cabbages and weeds combing spectral imaging technology and SAM taxonomy. *Spectrosc. Spectr. Anal.* **2015**, *35*, 479–485.
16. Tan, K.; Wang, S.; Song, Y. Estimating nitrogen status of rice canopy using hyperspectral reflectance combined with BPSO-SVR in cold region. *Chemom. Intell. Lab. Syst.* **2018**, *172*, 68–79. [CrossRef]
17. Bai, L.; Li, F.; Chang, Q. Increasing accuracy of hyper· spectral remote sensing for total nitrogen of winter wheat canopy by use of SPA and PLS methods. *J. Plant Nutr. Fertil.* **2018**, *24*, 52–58.
18. Wang, S.; Guan, K.; Wang, Z. Airborne hyperspectral imaging of nitrogen deficiency on crop traits and yield of corn by machine learning and radiative transfer modeling. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *105*, 102617.
19. Chen, L.; Liu, Q.; Yi, Q. Visualization of petroleum hyperspectral content in latosol based on hyperspectral imaging technology. *Spectrosc. Spectr. Anal.* **2018**, *38*, 2916–2922.
20. Li, L. Open Access Publishing of Scientific Scholarly Journals in China. Ph.D. Dissertation, Northwest A&F University, Shanxi, China, 2016.
21. Qin, J.; Chao, K.; Kim, K. Hyperspectral and multispectral imaging for evaluating food safety and quality. *J. Food Eng.* **2013**, *118*, 157–171. [CrossRef]
22. Press, W.H.; Teukolsky, S.A. Savitzky-Golay Smoothing Filters. *Comput. Phys.* **1990**, *4*, 669. [CrossRef]
23. Mishra, P.; Nordon, A.; Roger, J.-M. Improved prediction of tablet properties with near-infrared spectroscopy by a fusion of scatter correction techniques. *J. Pharm. Biomed. Anal.* **2021**, *192*, 113684. [CrossRef] [PubMed]
24. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *SMC-3*, 610–621. [CrossRef]
25. Li, H.; Liang, Y.; Xu, Q. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Anal. Chim. Acta* **2009**, *648*, 77–84. [CrossRef]

26. Araújo, M.C.U.; Saldanha, T.C.B.; Galvão, R.K.H. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemom. Intell. Lab. Syst.* **2001**, *57*, 65–73. [CrossRef]
27. Kamruzzaman, M.; ElMasry, G.; Sun, D.-W. Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis. *Anal. Chim. Acta* **2021**, *714*, 57–67. [CrossRef] [PubMed]
28. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [CrossRef]
29. Summers, D.; Lewis, M.; Ostendorf, B. Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. *Ecol. Indic.* **2011**, *11*, 123–131. [CrossRef]
30. Li, X. The correlation of total nitrogen content with leaf spectral reflectance and SPAD values in different corn varieties. *Soils Fertil. Sci. China* **2015**, *3*, 34–39+119.
31. Zhang, H.; Lan, J.; Guo, Y. A dense spatial–spectral attention network for hyperspectral image band selection. *Remote Sens. Lett.* **2021**, *12*, 1025–1037. [CrossRef]
32. Vane, G.; Goetz, A.F.H. Terrestrial imaging spectrometry: Current status, future trends. *Remote Sens. Environ.* **1993**, *44*, 117–126. [CrossRef]
33. Wang, C.; Yang, W.; Cui, L. Prediction of Soil Total Nitrogen Based on CatBoost Algorithm and Fusion of Image Spectral Features. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52* (Suppl. S1), 316–322.


**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.