



## Article

# A Method for Determining the Nitrogen Content of Wheat Leaves Using Multi-Source Spectral Data and a Convolution Neural Network

Jinyan Ju <sup>1,2,†</sup>, Zhenyang Lv <sup>2,3,†</sup>, Wuxiong Weng <sup>2,4,†</sup>, Zongfeng Zou <sup>5</sup>, Tenghui Lin <sup>2</sup>, Yingying Liu <sup>6</sup>, Zhentao Wang <sup>2,3,\*</sup> and Jinfeng Wang <sup>2,3,\*</sup>

<sup>1</sup> School of Mechanical Engineering, Heilongjiang University of Science and Technology, Harbin 150022, China; ju\_jinyan@163.com

<sup>2</sup> College of Engineering, Northeast Agricultural University, Harbin 150030, China; lvzhenyang\_mc@163.com (Z.L.); wuxiong\_w@126.com (W.W.)

<sup>3</sup> Heilongjiang Provincial Key Laboratory of Modern Agricultural Equipment Technology in Northern Cold Regions, Harbin 150030, China

<sup>4</sup> College of Mechanical Electronic Engineering, Fujian Agriculture and Forestry University, Fuzhou 350002, China

<sup>5</sup> Yantai Agricultural Technology Popularization Center, Yantai 261400, China

<sup>6</sup> Xinbin Manchu Autonomous County Shangjiahe Town Comprehensive Affairs Service Center, Fushun 113000, China

\* Correspondence: 15770085650@163.com (Z.W.); jinfeng\_w@126.com (J.W.); Tel.: +86-157-7008-5650 (Z.W.); +86-130-4515-6938 (J.W.)

† These authors contributed equally to this work.

**Abstract:** Accurate estimation of wheat leaf nitrogen concentration (LNC) is critical for characterizing ecosystem and plant physiological processes; it can further guide fertilization and other field management operations, and promote the sustainable development of agriculture. In this study, a wheat LNC test method based on multi-source spectral data and a convolutional neural network is proposed. First, interpolation reconstruction was performed on the wheat spectra data collected by different spectral instruments to ensure that the number of spectral channels and spectral range were consistent, and multi-source spectral data were constructed using interpolated, reconstructed imaging spectral data and non-imaging spectral data. Afterwards, the convolutional neural network DshNet and machine learning methods (PLSR, SVR, and RFR) were compared under various scenarios (non-imaging spectral data, imaging spectral data, and multi-source spectral data). Finally, the competitive adaptive reweighted sampling (CARS) and successive projections algorithm (SPA) were used to optimize the LNC detection model. The results show that the model based on DshNet has the highest test accuracy. The CARS method is more suitable for DshNet model optimization than SPA. In the modeling scenario with non-imaging spectral, imaging spectral, and multi-source spectral, the optimized  $R^2$  is 0.86, 0.82, and 0.82, and the RMSE is 0.29, 0.31, and 0.31, respectively. The LNC visualization results show that DshNet modeling using multi-source spectral data is conducive to the visualization expansion of non-imaging spectral data. Therefore, the method presented in this paper provides new considerations for spectral data from different sources and is helpful for related research on the chemometric task of multi-source spectral data.

**Keywords:** hyperspectral; convolution neural network; interpolation reconstruction; multi-source spectral data; leaf nitrogen content



**Citation:** Ju, J.; Lv, Z.; Weng, W.; Zou, Z.; Lin, T.; Liu, Y.; Wang, Z.; Wang, J. A Method for Determining the Nitrogen Content of Wheat Leaves Using Multi-Source Spectral Data and a Convolution Neural Network. *Agronomy* **2023**, *13*, 2387. <https://doi.org/10.3390/agronomy13092387>

Academic Editors: Chao Chen and Luis Manuel Navas Gracia

Received: 2 July 2023

Revised: 29 August 2023

Accepted: 12 September 2023

Published: 14 September 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

Wheat plays an important role in world agricultural production and strategic grain reserves [1]. However, when wheat is deficient in nitrogen, it not only affects crop yield but also reduces its quality. Conversely, excess nitrogen nutrition can cause environmental

pollution in water and the atmosphere [2]. Therefore, an accurate and scalable assessment of leaf nitrogen concentration (LNC, %) is increasingly important for improving N use efficiency. The hyperspectrum contains hundreds of continuous wavebands with rich reflectance spectral information about different chemical components and molecular structures of the sample to be measured. In recent years, it has been widely used in the fields of water environment [3], quality control [4], and biomedicine [5]. Similarly, hyperspectral technology provides new opportunities for real-time, non-destructive detection of nitrogen in wheat in smart agriculture.

Machine learning, including deep learning, has been widely used in different areas of plant science, such as plant breeding [6], in vitro culture [7], stress phenotyping [8], stress physiology [9], plant systems biology [10], plant identification [11], and pathogen identification [12]. Partial least squares regression (PLSR), artificial neural networks (ANN), k-nearest neighbor (k-NN), random forest regression (RFR), and support vector machines (SVM) are some classical machine learning methods [13,14]. However, the performance of the model depends on the applied preprocessing technology and requires the experience of experts [15]. In recent years, on the basis of ANN, researchers have added concepts such as local connections, weights, convolutional nuclei, and attention mechanisms and increased the depth of the network, thus developing various types of deep neural networks [16]. Due to the characteristic learning ability of deep learning, it has become an alternative method to seek better performance in classification and regression [17]. Recently, it has been reported that depth learning is applied to regression problems in spectral data analysis [18,19]. Zhang et al. [20] used near-infrared hyperspectral imaging technology combined with a CNN model to determine the total phenolic, total flavonoid, and total anthocyanin content in dried black wolfberry and compared it with partial least squares and least squares support vector machines. The results show that the deep learning method achieves good performance comparable to traditional methods. Ni et al. [21] used spectroscopy and an improved one-dimensional convolutional neural network (VWCNN) to predict the nitrogen content of the leaves of *Pinus massoniana* seedlings and found that the classical shallow CNN prediction model had better prediction accuracy and robustness compared to other machine learning models. These studies show that, compared with other machine learning methods, deep learning for regression can achieve the same or even better effects.

In addition, it is worth noting that most current studies use data collected using the same hyperspectral camera. For example, Yu et al. used a spectral range of 400–1000 nm with 256 spectral bands at a spectral resolution of 2.8 nm; the imaging spectrum of lettuce soluble solid (SSC) was studied, and the prediction accuracy  $R^2$  reached 0.90 [22]. Song et al. used non-imaging spectroscopy with a spectral range of 325–1075 nm and 750 spectral bands at a spectral resolution of 1 nm to study maize chlorophyll. The prediction accuracy  $R^2$  reached 0.78 [23]. Their research has achieved good results. But in recent years, as a result of the continuous development of spectroscopic devices and spectroscopy technology, portable spectroscopy products have become increasingly accessible. Accompanying this is multi-source data [24], such as non-imaging and imaging data. The resulting problem is that the models established for imaging spectral data and non-imaging spectral data cannot be universal, and the data cannot be used interchangeably without hindrance [25] because the spectral information collected by different instruments has different spectral ranges, spectral channels, and spectral resolutions. Similarly, as far as we know, the development of a wheat LNC detection model using spectral information from different sources is still an urgent problem to be solved, especially imaging spectral data and non-imaging spectral data.

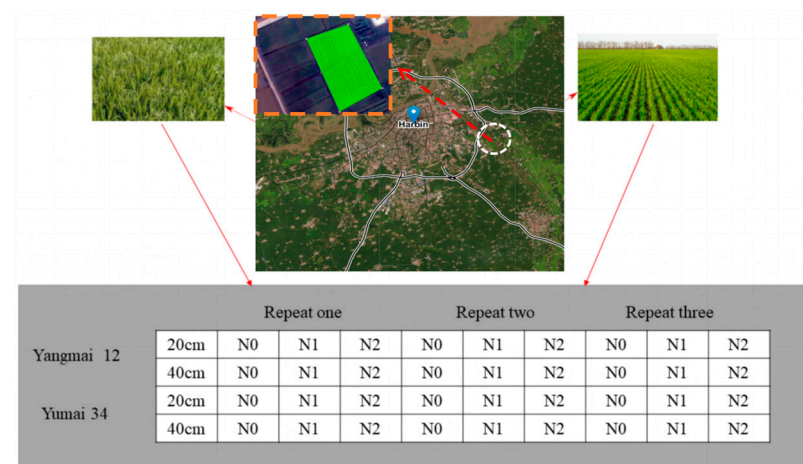
Given the aforementioned circumstances, the research aims are as follows: (1) Utilize interpolation reconstruction to make the spectral range and spectral channel numbers of imaging spectral data consistent with non-imaging spectral data and construct a multi-source dataset using the reconstructed imaging spectral data and non-imaging spectral data. (2) Convolutional neural networks DshNet for wheat LNC detection were established, and (3) the obtained detection results of the established model were systematically compared

to the respective outcomes from the traditional methods (PLSR, SVR, and RFR), thus establishing a method for estimating nitrogen content in wheat leaves using multi-source data.

## 2. Materials and Methods

### 2.1. Wheat Experiment Design

This study conducted experiments in two groups of fields, including different years, different nitrogen treatment levels, different wheat varieties, and different planting densities. The specific schemes are as follows: From 2019 to 2021, it was carried out in Harbin (44°04′–46°40′ N, 125°42′–130°10′ E), Heilongjiang Province. As shown in Figure 1, the experiment set two planting densities, the row spacing was 20 cm and 40 cm, the test varieties were ‘Yangmai 12’ and ‘Yumai 34’, and the three nitrogen levels were 0 kg/hm<sup>2</sup> (N0), 150 kg/hm<sup>2</sup> (N1) and 300 kg/hm<sup>2</sup> (N2) pure nitrogen. Nitrogen fertilizer is used at 50% as the base fertilizer, with 50% topdressing at the jointing stage. Phosphate fertilizer (P<sub>2</sub>O<sub>5</sub>) is applied per hectare at 120 kg, and potash fertilizer (K<sub>2</sub>O) applied per hectare is 135 kg. All phosphorus and potassium fertilizers are used as base fertilizers. The size of each area is 4 m × 5 m. The experiment was repeated three times. Field sampling occurred during the jointing and heading periods.



**Figure 1.** Study area and fertilization settings.

### 2.2. Data Collection

Both the non-imaging and imaging spectrometers require whiteboard correction prior to data acquisition, and the principle is essentially the same: the raw data were calibrated using a section of the whiteboard and were then transformed into reflectance data. Each pixel underwent the reflectance change using Equation (1) [26].

$$R_e = \frac{D_{Raw} - D_{Dark}}{D_{White} - D_{Dark}} \quad (1)$$

$R_e$  is a pixel's reflectance data,  $D_{Raw}$  is that pixel's raw data,  $D_{Dark}$  is the current value obtained from the dark frame, and  $D_{White}$  is the average value of the chosen white reference zone.

The collection method and specific parameters are as follows: The non-imaging spectrum used the ASD Fieldspec Pro FR2500 hyperspectral radiometer produced by the American Analytical Spectral Device (ASD) company. Because of its equipment structure and spectrum acquisition method, non-imaging spectrum is also called single-point spectrum. The band is 350–1000 nm with a spectral sampling interval of 1.4 nm, the spectral resolution is 3 nm, and the number of spectral channels is 651. Combined with related studies [27], the day when the weather is clear and cloudless is selected, and the canopy spectrum is measured from 10 am to 2 pm. When measured, the fiber optical probe is

vertically downward, the field of view of the spectrometer is  $25^\circ$ , and the measurement is performed at a height of 1 m from the wheat. Take 10 spectra as a set of sampled spectra and use their average value as the spectral reflectance value of the observation point. Correction of the whiteboard was performed before and after each acquisition target.

The imaging spectrometer used a Dualix Spectral Imaging Gaiafield Pro-V10E hyperspectral imager (Jiangsu Dualix Spectral Imaging Technology Co., Ltd., Beijing, China). The spectral range is 400–1000 nm, and the spectral resolution is 2.8 nm. The number of spectral channels is 468, and the number of pixels (space dimension  $\times$  spectral dimension) is  $1936 \times 1440$ . Weather conditions are clear and windless for collecting wheat hyperspectral canopy image data, just as they are for collecting non-imaging wheat canopy data. The time range was from 10 am to 2 pm, and the lens was positioned vertically downward during measurement. The vertical height from the top of the canopy is 1 m, and the whiteboard is placed at the starting position of the scan and collected by each plot. In order to get spectroscopic data that accurately represents the samples, a specific portion of the wheat plant was selected as the area of interest for nitrogen assessment. Then, the spectrum information was retrieved from this region. Subsequently, a ratio picture (grayscale image) was generated by dividing the wavelength images obtained at the highest and lowest spectral values of the sample. This division was performed to highlight the notable disparity between the sample and the background. The grayscale picture that was acquired then underwent a conversion process to transform it into a mask image (binary image). This mask image was then used to segregate the sample region of the original hyperspectral image from the surrounding backdrop. After that, the sample spectrum was obtained by calculating the spectral average of all pixels in the sample hyperspectral picture.

### 2.3. Determination of Nitrogen Content in Wheat Leafs

Synchronous with the spectrum measurement, 20 representative wheat plants were taken from each plot, separated by organs (leaf, stem, and ear), placed in an oven at  $105^\circ\text{C}$  for 30 min, dried at  $80^\circ\text{C}$ , finely ground and homogenized with an MM 400 mixer mill (Retsch, GmbH, Haan, Germany), and stored in airtight containers. The leaf layer nitrogen content was measured with an automatic element analyzer EA3000 from Italy Euro Vector company (Foggia, Italy).

### 2.4. Linear Interpolation Reconstruction

Construct a simple function as an approximation of a function that is inconvenient to process or calculate, and then obtain an approximation result of the function by processing the simple function. When an approximation function is required to take given discrete data, this processing method is called the interpolation method.

Let function  $y = f(x)$  give a series of function values on the interval  $[a, b]$ .

$$y_i = f(x_i), i = 0, 1, \dots, n \quad (2)$$

Here  $a \leq x_0 < x_1 < \dots < x_n \leq b$ , select a function  $\varphi(x)$  that satisfies.

$$\varphi(x_i) = y_i, i = 0, 1, \dots, n \quad (3)$$

As an approximate expression of function,  $y = f(x)$  is called an interpolation problem.  $\varphi(x)$  that satisfies the relationship is called the interpolation function of  $f(x)$ , and  $f(x)$  is called the interpolated function. The point  $x_0, x_1, \dots, x_n$  is called an interpolation node. The interval  $[a, b]$  is called an interpolation interval. In this way, the value of the interpolation function  $\varphi(x)$  is calculated as an approximation of the interpolated function  $f(x)$  given  $x$ . This process is called interpolation, and point  $x$  is called the interpolation point. Linear interpolation refers to the interpolation method in which the interpolation function  $\varphi(x)$  is a polynomial of degree one, and the interpolation error on the interpolation node is zero. Compared with other interpolation methods, such as nearest interpolation and trilinear interpolation, linear interpolation has the advantages of speed, strong convergence, good

numerical stability, and easy programming on the computer. Therefore, in this study, linear interpolation is used to reconstruct the data range and spectral channel of imaging spectrum and non-imaging spectrum data to keep the imaging spectrum and non-imaging spectrum data consistent, which is conducive to the construction of multi-source data and the visual expansion of a non-imaging spectrum.

## 2.5. Principal Component Analysis of LNC Spectral Diversity in Wheat

To explain the variation in wheat spectral reflectance with different sampling methods and the relationship between the spectral reflectance of wheat samples and LNC, the spectral reflectance principal component (PC) of wheat samples in different spectral regions can be extracted, and the LNC spectral diversity of varying sampling methods can be analyzed by PCA. The PC coefficients represent the variation in the spectral reflectance of wheat samples in different spectral regions, and they are related to the biochemical and structural characteristics of wheat and its interaction with N. Therefore, the variation trend of pc along the wavelength demonstrates how the wheat sample reflectance responds to sampling methods and LNC. In addition, PCA results can reflect differences in wheat spectra among interpolated reconstructed datasets and provide interpretable portability of multi-source data LNC detection models across different datasets.

## 2.6. Data Analysis

### 2.6.1. Partial Least Squares Regression (PLSR)

Partial least squares regression (PLSR) is a multivariate statistical technique. In order to address the issue of collinearity among predictors, a proposed approach involves decomposing the spectral matrix and the property matrix into latent variables (LVs) that are mutually orthogonal and uncorrelated. Subsequently, a regression model is constructed using these LVs. The PLSR model is established based on the obtained reflectance of the wheat canopy and leaf nitrogen content (LNC) as follows:

$$U = J * V^t + Q \quad (4)$$

$$I = J * Z^t + M \quad (5)$$

The reflectivity and leaf nitrogen content (LNC) of the wheat canopy are denoted by  $U$  ( $k \times y$ ) and  $I$  ( $k \times 1$ ) correspondingly. Here,  $k$  and  $y$  represent the LNC and reflectivity band numbers. The matrix  $J$  ( $k \times h$ ) represents the  $U$ -scores, whereas  $Q$  ( $k \times y$ ) and  $M$  ( $k \times 1$ ) denote the error terms. The loading matrixes of  $U$  and  $I$  are denoted as  $V$  ( $y \times h$ ) and  $Z$  ( $1 \times h$ ), respectively. The variable  $h$  represents the number of latent variables (LVs) in the partial least squares regression (PLSR) model, which is estimated using the projected residual sum of squares (PRESS) [28]. In order to construct the PLSR model for estimating LNC and mitigate the issues of over-fitting or under-fitting, a 10-fold cross-validation technique was used to train the PLSR model. The optimal number of latent variables (LVs) was found based on the criteria of minimizing the root mean square error (RMSE) and the Akaike Information Criterion (AIC) during the training process. Ultimately, the most favorable latent variables (LVs) were used to construct the correlation between  $U$  and  $I$ .

### 2.6.2. Support Vector Regression (SVR)

SVR is a machine learning method based on statistical learning theory. Structural risk minimization is adopted as the optimization criterion to seek a compromise between the accuracy of the data approximation and the complexity of the approximation function to obtain the best model generalization ability. To map the original data to a higher dimension, the kernel function is also exploited, which takes advantage of non-linear mapping. Thus, the solution to problems that cannot be handled in linear space can be facilitated. In addition, the SVR does not require an excessive number of samples for training, which is advantageous for scenarios with a small sample size [29–31].



### 2.6.3. Random Forest Regression (RFR)

Random forest regression is a decision-tree-based algorithm that integrates multiple trees using the concept of ensemble learning. The random forest algorithm is based on statistical learning theory, and it extracts multiple samples from the original sample using the bootstrap resampling method, constructs a decision tree for each bootstrap sample, and then uses the best results in the decision tree as the final test result [32,33].

### 2.6.4. Convolutional Neural Networks (DshNet)

CNN is a typical feed-forward neural network that essentially extracts the characteristics of input data by building multiple filters. These filters convolute and pool the input data layer by layer, extracting the topological features contained in the input data layer by layer. As the number of layers of the network increases, eventually robust features that are invariant to translation and rotation are extracted from the original input data. A 1D-CNN deep neural network framework, DshNet, was developed to train and test wheat LNC detection from a multi-source dataset based on hyperspectral characteristics. The convolutional neural network DshNet is shown in Figure 2. The specific model process is shown in Figure S2. The input to the DshNet network is spectral reflectance data, and the output is the LNC. DshNet has six modules, and the first five modules are convolutional layers. Specifically, in the convolution layer, the convolution kernel convolutes the feature vectors of the previous output layer and constructs the output feature vectors using the non-linear activation function ReLU. The output of each layer is the convolution result of multiple input features. Its mathematical model can be described as follows:

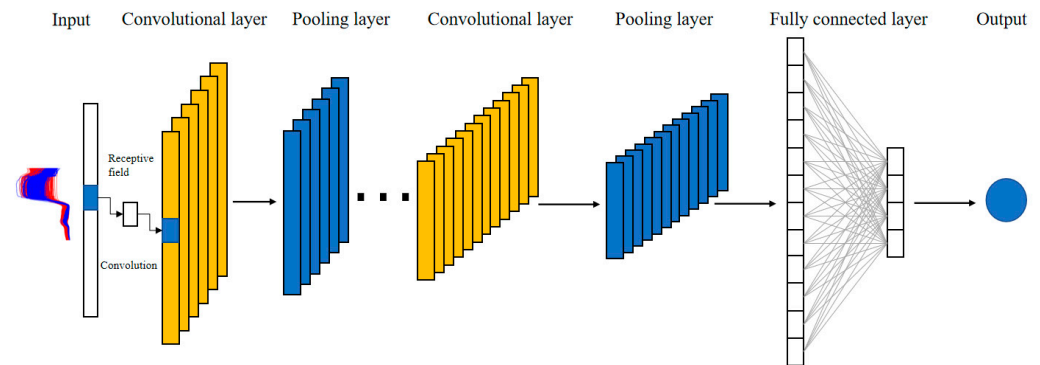
$$x_j^l = f \left( \sum_{i \in M_j} x_i^{l-1} \times k_{ij}^l + b_j^l \right) \quad (6)$$

where  $M_j$  is the input feature vectors,  $l$  is the layer  $l$  network,  $k$  is the convolution kernel,  $b$  is Network bias, and  $x_j^{l-1}$  is the layer  $l$  input. Layer-by-layer extraction of input data features is performed by alternately utilizing convolutional and pooling layers (max pooling), ending with a fully connected layer before the network output layer. The feature map is flattened into a one-dimensional vector before being sent to the full-connection layer. The dropout regularization method was used to prevent overfitting. The data are further abstracted and recombined by the features through the full-connection layer, and the model output can be obtained by activating the function. In this study, the convolution kernel size of the convolution layer was set to 3, and the step size was 1. An optimal learning rate of 0.0001 was utilized based on the forecast results. The backpropagation algorithm was used to randomly initialize and train the ownership values. We used minibatches of size 30 and trained the network with no data augmentation. The number of iterations was set to 500, and the callback function (ModelCheckpoint) was used to monitor the test loss value at each epoch. When the predicted loss value of the iteration is less than the value of the previous time, the model parameters are recorded, and finally, the 500-epoch optimal model parameters are retained. Among them, PLS, SVM, and RFR are all implemented in scikit-learn (0.24.2). The DshNet is based on Google's Tensorflow(2.4.0), using Tensorflow-GPU, and the graphics card is NVIDIA GEFORCE GTX 1650.

### 2.6.5. Train and Test Set

In this study, refer to Chu Xiaoli's method of dividing samples [34]. Specifically, it was divided according to the concentration gradient of the nitrogen content, and the train (calibration) and test (prediction) sets were divided according to 2:1. As shown in Table 1, the train set and test set were divided by a non-imaging spectrometer, an imaging spectrometer, and mixed data (multi-source). Among them, there are 429 train sets and 212 test sets of multi-source data; the minimum value of the train set is 1.38%, the maximum value is 4.84%, the minimum value of the test set is 1.51%, and the maximum value is

4.73%. The coefficient of variation (CV) between the train set and the test set differs only by 0.01, indicating that the sample partition is reasonable. In addition, in order to obtain reliable models, this study used 10-fold cross validation to evaluate model performance. The cross-validation process can be seen in Supplementary Materials, Figure S3.



**Figure 2.** Schematic diagram of the DshNet convolutional neural network.

**Table 1.** The statistical parameters of the train and test sets for the wheat LNC.

Spectral Type	Dataset	Number of Samples	Minimum (%)	Maximum (%)	Mean (%)	SD	CV
Non-imaging	Train	293	1.54	4.73	3.12	0.75	0.24
	Test	142	1.51	4.84	3.06	0.79	0.25
Imaging	Train	138	1.38	4.52	3.00	0.69	0.23
	Test	68	1.51	4.22	2.85	0.68	0.24
Multi-source	Train	429	1.38	4.84	3.05	0.75	0.25
	Test	212	1.51	4.73	3.05	0.74	0.24

Note: A multi-source spectral dataset refers to a dataset that combines interpolated reconstructed imaging spectral data with non-imaging spectral data.

## 2.7. Model Optimization

The model can be made simpler via wavelength selection by lowering the number of wavelength variables. Furthermore, by removing extraneous or non-linear variables, it is possible to derive a corrective model that exhibits a high degree of prediction accuracy and good robustness. Additionally, the process of model optimization by means of wavelength selection enables expedited detection in real-world scenarios. The CARS method, developed by Li et al. is a variable selection algorithm based on the simple yet efficient premise of ‘survival of the fittest’ [35]. The absolute values of the regression coefficients of a PLSR model are used as indexes in CARS to evaluate the relevance of each variable. It picks N subsets of variables by repeating N sample runs to produce the optimum subset variables with the lowest RMSECV value. Variables with tiny absolute values of regression coefficients are removed using exponentially declining functions (EDFs) and adaptive reweighted sampling (ARS) [36].

The successive projections algorithm (SPA) was initially proposed for constructing multivariable calibration models, aiming to select variables for use in the multiple linear regression model. In this case, the collinearity avoidance mechanism embedded in the SPA reduces the propagation of measurement noise during calibration. SPA uses the projection analysis of the vector, by projecting the wavelength to other wavelengths, compare the size of the projection vector, take the maximum wavelength of the projection vector as the selected wavelength, and then select the final characteristic wavelength based on the correction model. The initial iteration vector is  $x_{k(0)}$ , the variable to be extracted is N, and the spectral matrix is column J. The brief steps of the algorithm are as follows:

- (1) A column of the spectral matrix (column  $J_{th}$ ) is randomly selected, and column  $J_{th}$  of the modeling set is assigned to  $x_j$ , denoted as  $x_{k(0)}$ .

- (2) The set of unselected column vector positions is denoted as  $s$ :

$$s = \{j, 1 \leq j \leq J, j \notin \{k(0), k(1), \dots, k(n-1)\}\} \quad (7)$$

- (3) The projection of  $x_j$  onto the remaining column vectors is calculated separately.

$$p_{xj} = x_j - \left(x_j^T x_{k(n-1)}\right) x_{k(n-1)} \left(x_{k(n-1)}^T x_{k(n-1)}\right)^{-1} \quad (8)$$

- (4) The spectral wavelength of the maximum projection vector is extracted.

$$p_{xj} = x_j - \left(x_j^T x_{k(n-1)}\right) x_{k(n-1)} \left(x_{k(n-1)}^T x_{k(n-1)}\right)^{-1} \quad (9)$$

- (5) Let  $x_j = p_{xj}, j \in s$

- (6)  $n$  sum, if  $n < N$ , then it shall be calculated circularly according to Equation (7).

Finally, the extracted variable is  $\{x_{k(n)} = 0, 1, \dots, N-1\}$ . Corresponding to  $k(0)$  and  $N$  in each cycle, multiple linear regression analysis (MLR) models are established respectively, and the root mean square error (RMSECV) of the modeling set interactive verification is obtained, which corresponds to different phenological subsets, where  $k(0)$  and  $N$  corresponding to the minimum RMSECV value are the optimal values.

## 2.8. Model Assessment

The performance of all models was evaluated based on the following parameters: They are the determination coefficient ( $R^2$ ), the root mean square error of train (RMSEC), the root mean square error of test (RMSEP), and residual test deviation (RPD).

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (y_{i,actual} - y_{i,predicted})^2}{n-1}} \quad (10)$$

$y_{i,actual}$  is the measured  $LNC_i$ ,  $y_{i,predicted}$  is the predicted  $LNC_i$  and  $n$  is the number of samples. The smaller the RMSEC, the better the model established by the model train set. The smaller the RMSEP, the better the test of the model's ability.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{i,actual} - \bar{y}_{i,predicted})^2}{\sum_{i=1}^n (y_{i,actual} - \bar{y}_{i,actual})^2} \quad (11)$$

where  $y_{i,actual}$  is the measured  $LNC_i$ ,  $\bar{y}_{i,predicted}$  is the mean of predicted  $LNC_i$  and  $\bar{y}_{i,actual}$  is the mean of measured  $LNC_i$ . The range of  $R^2$  is 0~1. The closer to 1, the better the model's performance.

$$RPD = \frac{SD}{RMSEP} \quad (12)$$

$SD$  is the standard deviation. When  $RPD$  is greater than 2.0, it indicates that the model has good predictive ability. When  $1.4 < RPD < 2.0$ , it indicates that the model can roughly estimate the sample. When  $RPD < 1.4$ , it indicates that the model cannot predict the sample.

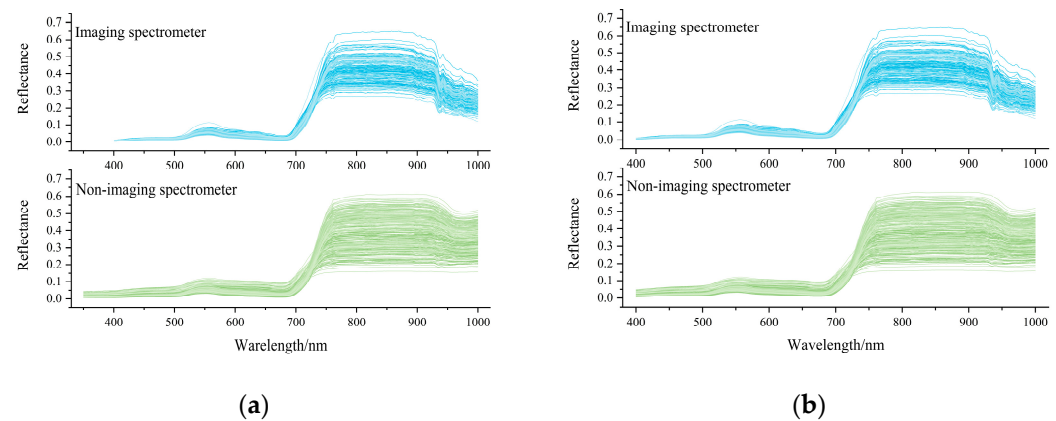
## 3. Results and Discussion

### 3.1. Spectral Properties of Wheat Samples and Spectral Interpolation

Depending on the method used to obtain spectral data, the reflectance curve of wheat varies. However, the general trend of the wheat spectrum of different LNCs is similar. At 350–680 nm of the visible light band, the reflectivity of wheat is low, as shown in Figure 3a, and the curve initially increases before decreasing. Following this, the spectral reflectance



swiftly increases in the 680–750 nm band, which is a distinctive spectral characteristic of green plants. The spectral reflectance of the near-infrared band (750–940 nm) was stable and then progressively decreased between 940 and 1000 nm. Moreover, despite the fact that spectral data from various sources exhibit similar trends, imaging spectral data exhibit more noise and baseline drift.



**Figure 3.** Reflectance curves of the sample. (a) Original spectrum; (b) Linear interpolation spectrum. Note: Original spectrum (non-imaging spectral dataset spectral range 350–1000 nm, spectral channel number 651, imaging spectral dataset spectral range 400–1000 nm, spectral channel number 468). Linear interpolation spectrum (both imaging and non-imaging spectral ranges are 400–1000 nm, with 601 channels).

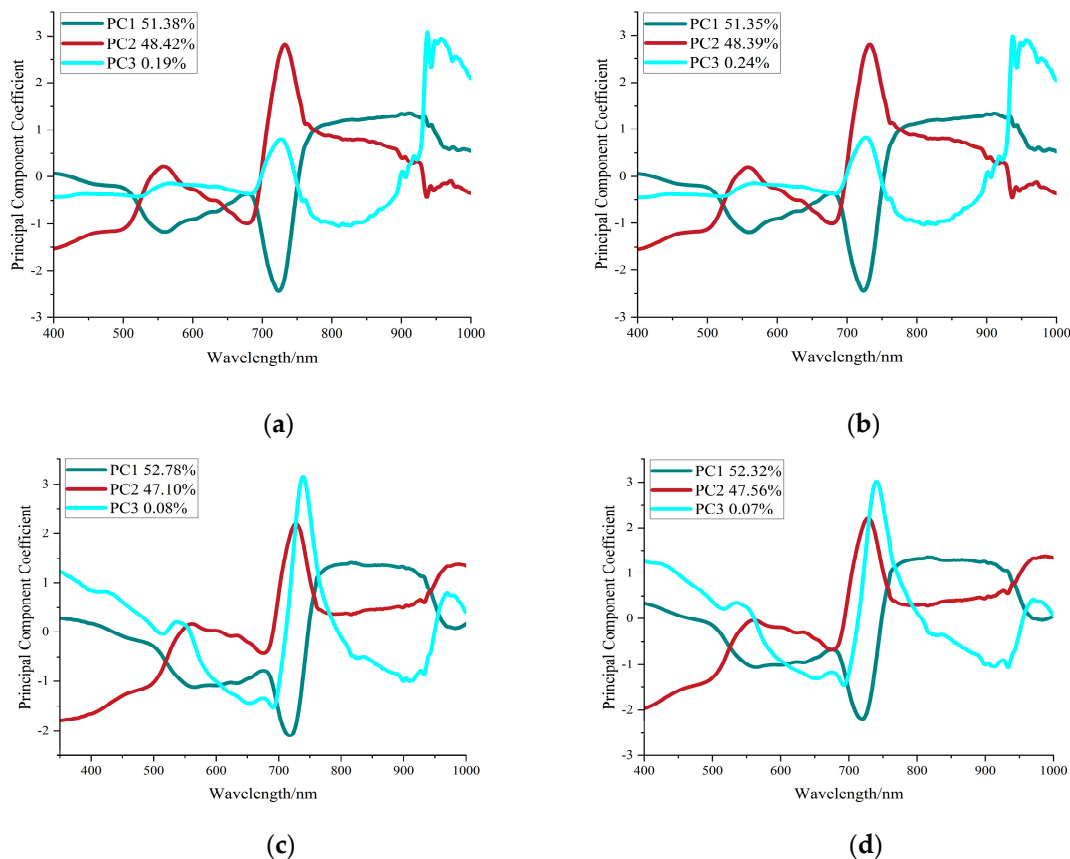
Specifically, several reflection peaks can also be observed, reflecting the prominent spectral characteristics shown in Figure 3a. The band around 550 nm is related to xanthophylls [37]. The 680–750 nm (red-edge) and 700 nm bands are related to chlorophyll. Many weak peaks were recorded at 760–980 nm, mainly reflecting the tensile vibration of the third overtone of the O-H bond and C-H bond [38]. From the results obtained, it can be concluded that these observed differences can further serve as the basis for determining LNCs in wheat.

However, different instrument manufacturers have different spectral ranges and spectral channels, which affect the generalization and robustness of LNC detection models. Therefore, when the two sets of data (non-imaging spectrum data and imaging spectrum data) are transformed into a 400–1000 nm range using the linear interpolation technique, the number of spectral channels is 601. The effect of the interpolation is shown in Figure 3b, and it can be seen that the wheat spectral range and number of spectral channels have undergone significant changes after transformation; however, there was no significant change in the spectral morphological characteristics, which are still retaining valuable characteristic information.

### 3.2. Preliminary Investigation by PCA

Leaf spectral diversity reflects the variation trend of wheat leaf reflectance after different sampling methods (imaging or non-imaging) and interpolation reconstructions and reflects whether spectral information is lost after interpolation reconstruction [39]. Figure 4a,c reflects the variation trend of wheat leaf reflectance under different sampling methods. In general, different sampling methods were relatively consistent in PC1 and PC2, but PC3 showed different patterns. Since PC1 and PC2 account for more than 99% of the variation in leaf reflectance, it can be seen that spectral information specific to LNC is acquired with different sampling methods and that PC3 may reflect differences in different sampling methods. In the PCA diagram, after interpolation reconstruction of the original spectrum (Figure 4b,d), it can be found that the spectral information did not change significantly and that the interpolation reconstruction retained the characteristic information of

the original spectrum. Both the non-imaging and the image spectral data show the same trend change as the original spectral PCA.



**Figure 4.** PCA analysis results. (a) Original imaging spectra data; (b) Image spectral data after linear interpolation; (c) Original non-imaging spectra data; (d) Non-imaging after linear interpolation.

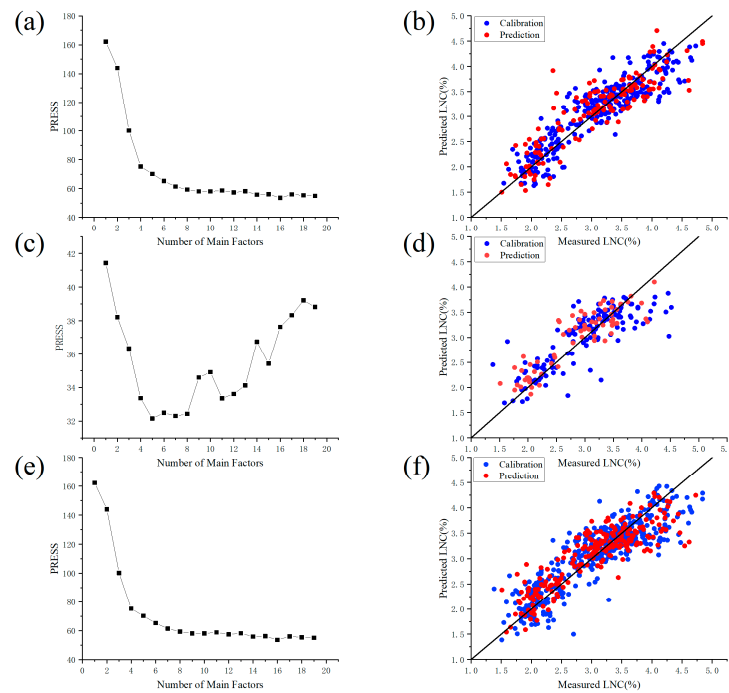
### 3.3. Establishment of Models for Predicting LNC and Comparison

#### 3.3.1. PLSR

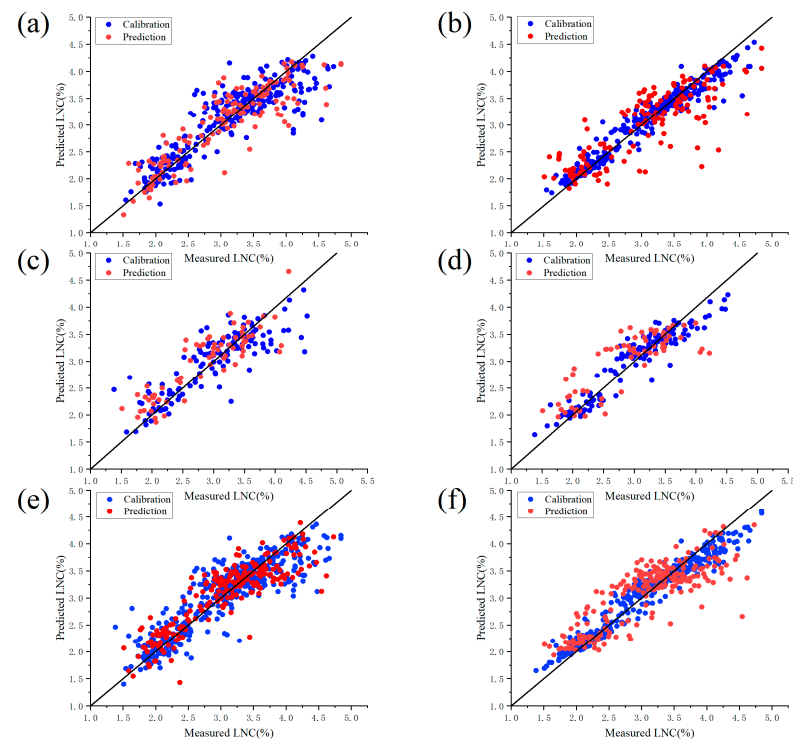
When performing PLSR, it is particularly important to determine the number of LVs in the regression. If the number of LVs is too small, more useful information about the original spectrum will be lost. When the number of LVs is too large, an excessive amount of noise information will be included, and the model's predictive ability will suffer. Shown in Figure 5a,c,e are 10-fold cross-validation PRESS graphs, and the number of LVs was determined by the AIC during training. The number of LVs in non-imaging spectral data, imaging spectral data, and multi-source spectral data are 16, 5, and 16, respectively. On this basis, PLSR modeling is conducted, and the results are shown in Figure 5b,d,f.

#### 3.3.2. SVR and RFR

SVR and RFR need to select hyperparameters when extracting features to build a model. This article uses grid search hyperparameters, the GridSearchCV function in scikit-learn, and 10-fold cross validation and selects  $R^2$  as the evaluation function. Shown in Figure 6a,c,e are the scatter plots of the train set and test set of the SVR model. Figure 6b,d,f are scattered plots of the RFR model train and test sets.



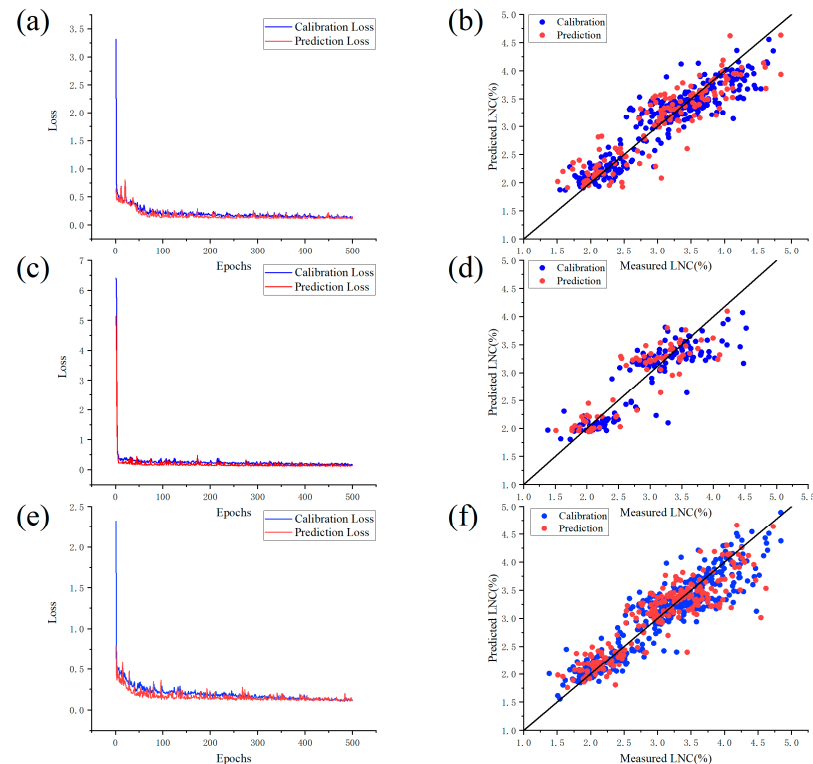
**Figure 5.** PLSR test results and the 10-fold cross-validation approach of the PRESS graph. (a) PRESS graph of non-imaging spectra data; (b) Predicted LNC scatter plot of non-imaging spectra data; (c) PRESS graph of imaging spectra data; (d) Predicted LNC scatter plot of imaging spectra data; (e) PRESS graph of multi-source spectra data; (f) Predicted LNC scatter plot of multi-source spectra data.



**Figure 6.** SVR and RFR model-predicted LNC scatter plot. (a) Non-imaging spectra data SVR model; (b) Non-imaging spectra data RFR model; (c) Imaging spectra data SVR model; (d) Imaging spectra data RFR model; (e) Multi-source spectra data SVR model; (f) Multi-source spectra data RFR model.

### 3.3.3. DshNet

DshNet feature extraction and model building also need to set hyperparameters. Adam is considered the most suitable optimizer for spectral analysis [40]. Shown in Figure 7a,c,e are the loss function curves of the train and test sets of DshNet. From 300 to 500 epochs, the loss function of the train and test sets tends to be stable. Figure 7b,d,f shows the scatter plots of the train and test sets of the DshNet model.



**Figure 7.** DshNet model test results. (a,c,e), respectively, refer to the loss function graph in the case of DshNet modeling with non-imaging spectrum data, imaging spectrum data, and multi-source spectrum data; (b,d,f) are the predicted LNC scatter plots of DshNet using non-imaging spectra data, imaging spectra data, and multi-source spectra data, respectively.

### 3.3.4. Comparison of PLSR, SVR, RFR and DshNet Establish LNC Model

For PLSR, SVR, RFR, and DshNet, this study used non-imaging spectra, imaging spectra, and mixed spectra as input and used the same train set to build models to compare these four algorithms (Table 2). For non-imaging spectra, the LNC model established by its DshNet is the best; the  $R^2$  of its train set reaches 0.85, the predicted  $R^2$  reaches 0.78, and the RPD is 2.30. The second is the PLSR-established LNC model. The predictive ability of SVR is slightly lower than that of PLSR, and the overall accuracy is not significantly different. There is still a certain overfitting phenomenon after gridding and screening of hyperparameters, consistent with Yao et al.'s [27] research on the LNC detection method, regardless of whether non-imaging, imaging, or multi-source were used to build models. For imaging spectra, the  $R^2$  of the LNC model test set established by DshNet is only 0.01 higher than that of the model established by PLSR, and its RPD is only increased by 0.04, which is not a significant improvement. Because deep learning requires a large number of samples to perform learning, while there are only 138 imaging spectra, there is no advantage over traditional machine learning methods. Yang et al. [40] reported the same conclusion in deep learning for vibration spectrum analysis. Using mixed spectra for modeling, the LNC model established by SVR has a higher predictive ability than the model established by PLSR because the resolution of non-imaging spectra and imaging spectra is inconsistent, the nonlinearity of the spectrum increases, and PLSR is a linear

model, so the LNC model established by SVR has a better predictive ability than the model established by PLSR. Even if the nonlinearity of the spectrum increases, the convolutional layer and pooling layer of the DshNet are used to mine the depth characteristics of the spectrum. By 0.1, the RPD of the LNC model established by DshNet was higher than that of the model established by SVR. On the whole, when the sample size is relatively small, the predictive ability of the LNC model established by DshNet and the traditional algorithm is not substantially different. When the sample size increases, the predictive ability of DshNet's LNC model outperforms that of the traditional algorithm. Yang et al. used a convolutional neural network model (CNN) based on the PyTorch framework to obtain depth features. Spectral characteristics were obtained by using position features (PFs) and a vegetation index (VIs). It is found that the accuracy of the LNC estimation model based on the gradient boosting decision tree GBDT model is the best (calibration set  $R^2 = 0.975$ , verification set  $R^2 = 0.861$ ) after feature extraction by CNN, indicating the advantage of CNN in feature extraction [41].

**Table 2.** The LNC models based on PLSR, SVR, RFR, and DshNet.

	Model	Train Set		Test Set		
		$R^2$	RMSEC	$R^2$	RMSEP	RPD
Non-imaging	PLSR	0.83	0.31	0.78	0.37	2.15
	SVR	0.78	0.35	0.78	0.37	2.12
	RFR	0.93	0.19	0.66	0.46	1.70
	DshNet	0.85	0.29	0.81	0.34	2.30
Imaging	PLSR	0.68	0.39	0.75	0.34	1.99
	SVR	0.73	0.36	0.73	0.35	1.91
	RFR	0.89	0.23	0.63	0.42	1.63
	DshNet	0.73	0.36	0.76	0.33	2.03
Multi-source	PLSR	0.78	0.50	0.75	0.37	2.01
	SVR	0.79	0.34	0.78	0.35	2.13
	RFR	0.94	0.19	0.70	0.41	1.83
	DshNet	0.85	0.29	0.80	0.33	2.23

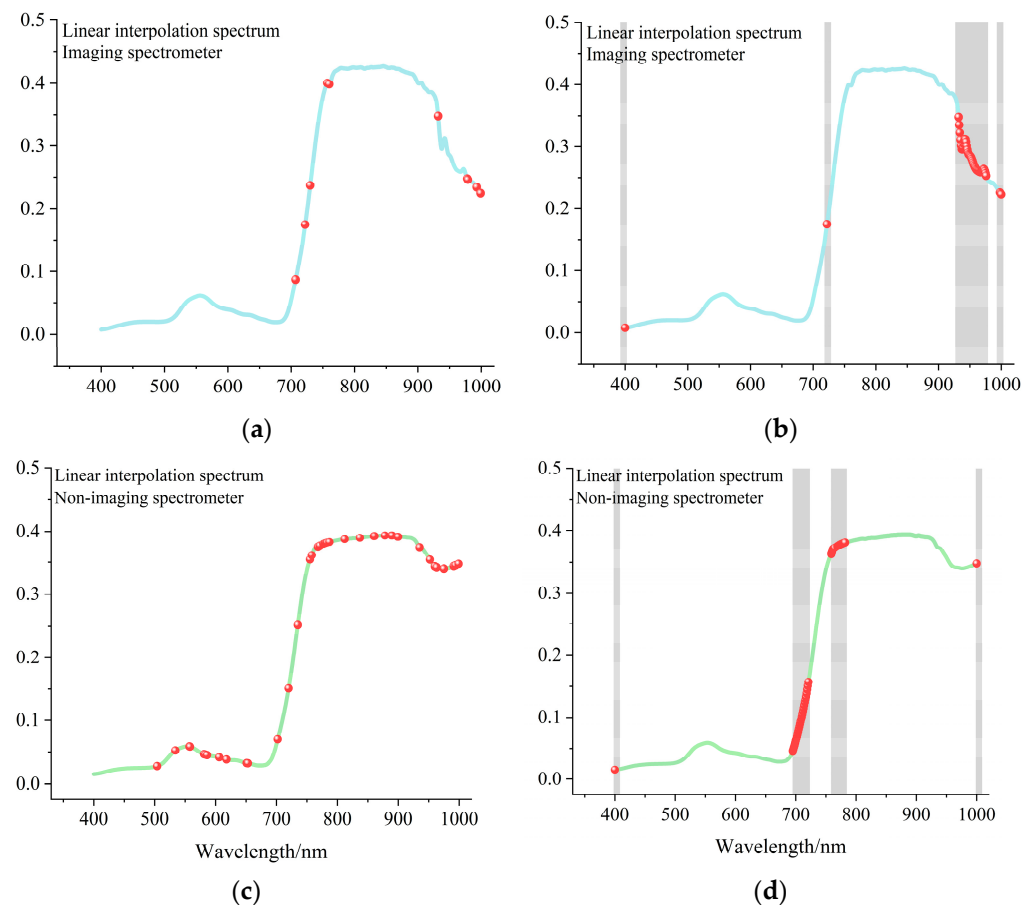
Note: A multi-source spectral dataset refers to a dataset that combines interpolated reconstructed imaging spectral data with non-imaging spectral data.

### 3.3.5. Comparison of Non-Imaging, Imaging and Multi-Source Spectra Data to Establish the LNC Model

As shown in Table 2, the model was built using four methods (PLSR, SVR, RFR, and DshNet) on different datasets (non-imaging spectral data, imaging spectral data, and multi-source spectral dataset). In a comprehensive comparison of  $R^2$ , RMSEC, RMSEP, and RPD, the LNC model established by DshNet provides the best performance. Therefore, we compared the DshNet models of the non-imaging, imaging, and multi-source spectra. In the training phase, non-imaging spectra and the LNC model were derived from mixed spectra, both of which are 0.85. In the test phase, the  $R^2$  of the LNC model established by the non-imaging spectra is 0.01 higher than that of the LNC model established by the mixed spectra, and its RPD is 0.07 higher. Taken together, we can find that when multi-source spectra are used for modeling, the predictive ability of its LNC model is only reduced by 0.07 compared to the best model established by non-imaging spectra, and its RPD reaches 2.23, which still has a good predictive ability and can predict the data from two data sources. Jia et al. used the hyperspectral data measured in the field to design a variety of spectral indexes for predicting wheat LNC, and the maximum  $R^2$  obtained was 0.66 [42]. Based on UAV hyperspectral images, Zhang et al. used PLSR, GA-PLSR, RF, XGBoost, and other methods to predict LNC, and the maximum  $R^2$  of the four methods was 0.55 [43]. Compared with previous studies, the results of our methods (DshNet, PLSR, SVR, and RFR) are all in a reasonable range, and the combination of multi-source spectral data with convolutional neural networks increases the degree of data utilization.

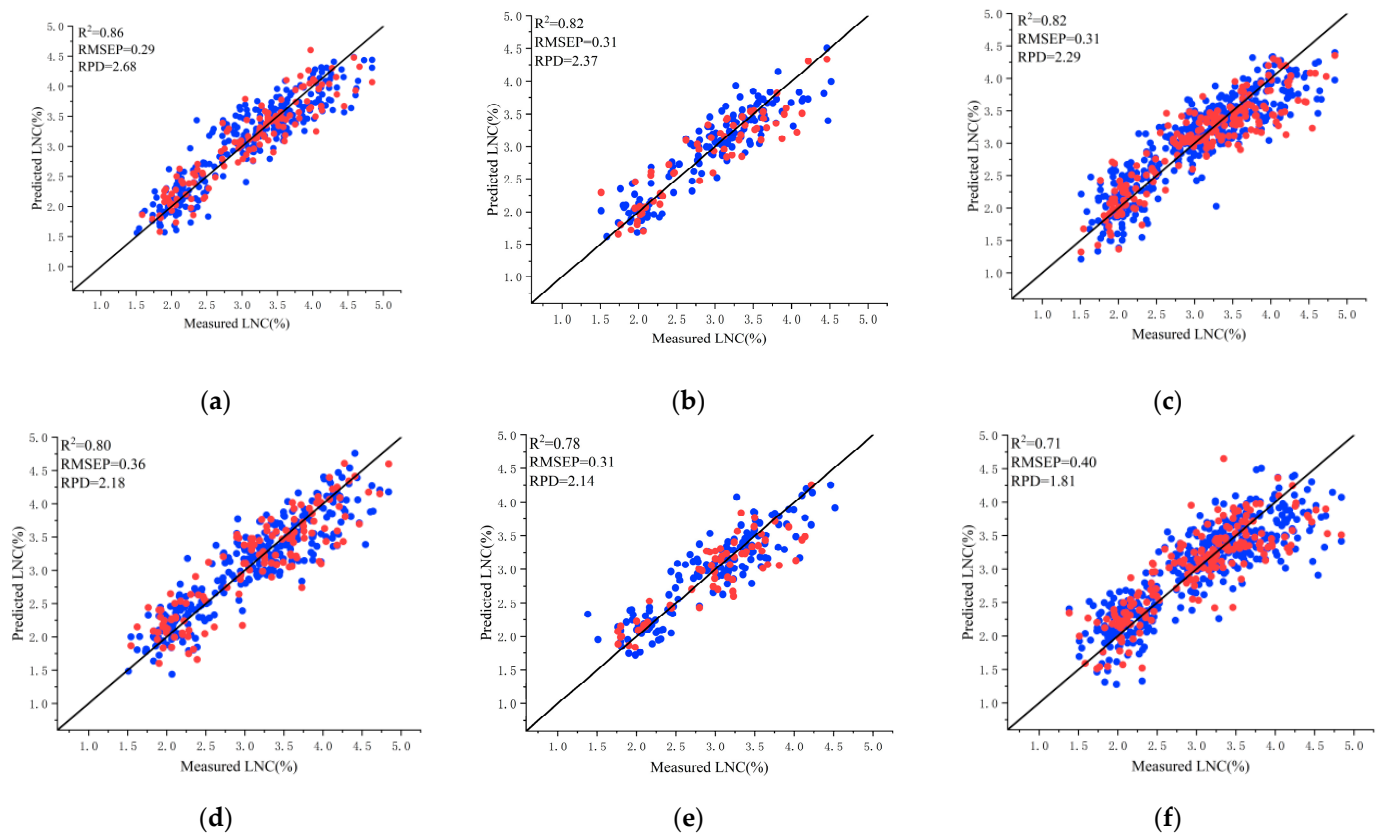
### 3.4. Optimization of the LNC Test Model Based on the CARS and SPA Extraction of Characteristic Wavelengths

The LNC test model was optimized using the characteristic wavelength obtained using CARS and SPA techniques. As seen in Figure 8, there is a difference in the characteristic wavelength between the non-imaging and imaging spectra, as well as between the original spectrum and linear interpolation spectrum. In this study, the typical wavelengths of the spectral data were used for the construction of an LNC test model. Figure 9 presents a summary of the anticipated outcomes derived from the optimized DshNet model. The results indicate that the models using CARS for extracting distinctive wavelengths outperform the models utilizing the whole spectrum range (full wavelength). However, it is important to note that the utilization of CARS leads to a reduction of over 85% in the number of spectral channels. The coefficient of determination ( $R^2$ ) for the optimized non-imaging spectra, imaging spectra, and multi-source spectra are 0.86, 0.82, and 0.82, respectively. The root mean square errors (RMSE) for the three cases are 0.29, 0.31, and 0.31, respectively. In comparison to full-wavelength modeling, the values of  $R^2$  show an increase of 6.17%, 7.89%, and 2.25%, correspondingly. The root mean square error (RMSE) shows a reduction of 14.70%, 6.06%, and 6.06%. The performance of the feature wavelength derived by SPA is comparable to that of the model using the entire spectrum. However, both of these performances are inferior to that of the model generated by CARS. Obviously, DshNet model performs better after the characteristic wavelength selection of CARS.



**Figure 8.** Characteristic wavelength extracted by the CARS and SPA algorithm. (a) CARS + imaging spectrometer; (b) SPA + imaging spectrometer; (c) CARS + non-imaging spectrometer; (d) SPA + non-imaging spectrometer.





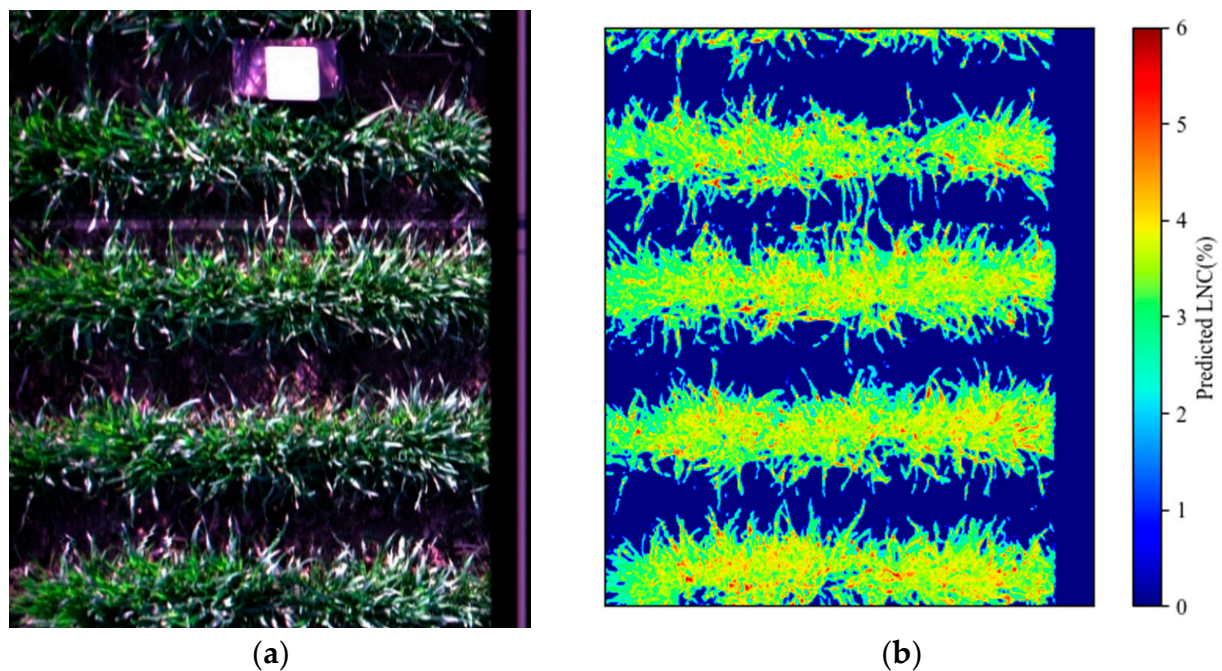
**Figure 9.** LNC test results of DshNet model. (a) Non-imaging spectrum + CARS; (b) Imaging spectrum + CARS; (c) Multi-source spectrum + CARS; (d) Non-imaging spectrum + SPA; (e) Imaging spectrum + SPA; (f) Multi-source spectrum + SPA.

Compared with previous studies, Abdel-Rahman used EO-1 Hyperion hyperspectral data to predict sugarcane LNC by RF model, and the prediction accuracy  $R^2$  was 0.67 [44]. Zhai et al. estimated nitrogen content in leaves of different plants by partial least squares regression and support vector machine regression using visible light and near-infrared reflectance spectra in the laboratory and found that SVR ( $R^2 = 0.706$ ) had a better prediction effect than PLSR ( $R^2 = 0.663$ ) [45]. Min and Lee used the diffuse reflection data of a spectrophotometer at 400–2500 nm in the laboratory environment and used stepwise multiple regression (SMLR) and partial least squares regression (PLS) to predict the nitrogen concentration of leaves, and the prediction accuracy  $R^2$  was 0.839 and 0.828, respectively [46]. These studies have all achieved good results; however, these methods show different prediction accuracy with different pretreatments [15]. In our study, convolutional neural networks DshNet and CARS were used to predict LNC using multi-source data without any preprocessing algorithm, and the accuracy  $R^2$  was between 0.82 and 0.86, indicating that CNN has great potential for wheat LNC detection.

### 3.5. Spatial Distributions of Predicted LNC

In this work, the distribution map and potential distribution pattern of the anticipated LNC were generated using a DshNet model using multi-source spectra data. In ENVI 5.4 software (Research Systems Inc., Co., Melbourne, FL, USA), the background data and specular reflection were removed using a binary mask before creating a distribution map. Then, the wheat spectrum of each pixel was interpolated and reconstructed. A wheat sample hypercube was subjected to the DshNet model, and identically predicted values were generated for pixels with similar spectral features. Python 3.7 (Python Software Foundation, Wilmington, DE, USA) is used to carry out this operation. This made it easier to collect concentration data from each and every area of the wheat sample and gave

pertinent information on its heterogeneity. Figure 10 depicts the LNC visualization of the wheat sample at the jointing stage. Different colors correspond to different LNCs; dark red areas denote higher LNCs, and dark blue areas denote lower LNCs. The color progressively shifts from dark blue to dark red as the LNC rises. Hyperspectral imaging makes it easier to implement visualization of arbitrary local regions in wheat samples than non-imaging detection methods do, while multi-source spectral data modeling makes it possible to obtain larger visual areas of non-imaging spectral data. The LNC across wheat samples and between distinct regions within the same sample is shown by these distribution maps. This made it easier to see how concentrations changed between samples and even within the same sample. As long as a precise DshNet model is established, it is possible to obtain such chemical images online.



**Figure 10.** The predicted LNC distribution map obtained by the DshNet model. (a) Original wheat hyperspectral data map; (b) LNC distribution map.

#### 4. Conclusions

In this study, a non-imaging spectrograph and an imaging spectrograph were used to collect wheat hyperspectral information in different periods, and the linear interpolation method was used to build multi-source spectral data. The convolutional neural network DshNet and classical machine learning models (PLSR, SVR, and RFR) were established to predict wheat LNC, and the modeling performances of DshNet, PLSR, SVR, and RFR in different source spectral data were compared. It was found that the test results using DshNet are better than PLSR, SVR, and RFR under any spectral data modeling condition. The CARS method is more suitable for DshNet model optimization than the SPA. In the modeling scene of the non-imaging spectrum, imaging spectrum, and multi-source spectrum, the optimized  $R^2$  is 0.86, 0.82, and 0.82, respectively. The RMSE is 0.29, 0.31, and 0.31, respectively. The LNC visualization map shows that the method we proposed for detecting nitrogen content in wheat leaves using multi-source spectral data is helpful for the visualization expansion of non-imaging spectral data. The above results show that our DshNet outperforms traditional modeling methods. However, whether this method can be applied to low-altitude remote sensing of UAVs and satellite remote sensing levels needs further research. In addition, this method can provide an additional reference for research related to the chemometrics task of multi-source spectral data.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agronomy13092387/s1>, Figure S1: Schematic diagram of non-imaging spectrometer and imaging spectrometer collecting wheat canopy spectra; Figure S2: Flow chart of wheat LNC test with DshNet; Figure S3: Cross-validation diagram.

**Author Contributions:** J.J.: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Supervision; Writing—original draft; Test; Funding acquisition. Z.L.: Conceptualization; Data curation; Project administration; Writing—review and editing. W.W.: Data curation; Investigation; Test. Z.Z.: Formal analysis; Investigation; Test; Visualization. T.L.: Investigation; Software; Visualization. Y.L.: Data curation; Investigation. Z.W.: Software; Supervision; Writing—original draft; Test; Funding acquisition. J.W.: Investigation; Visualization; Conceptualization; Methodology; Supervision. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by China's National Key R & D Plan, grant number (2021YFD200060502; 2018YFD0300105); the Opening Project of Heilongjiang Provincial Key Laboratory of Modern Agricultural Equipment Technology in Northern Cold Regions, grant number (55200512); Basic Research Operating Expenses of Provincial Undergraduate Colleges and Universities in Heilongjiang Province.

**Data Availability Statement:** Data will be made available on request.

**Acknowledgments:** This work is supported by China's National Key R & D Plan—(2016YFD0300909).

**Conflicts of Interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. Zhu, Y.; Tian, Y.C.; Yao, X.; Liu, X.J.; Cao, W.X. Analysis of common canopy reflectance spectra for indicating leaf nitrogen concentrations in wheat and rice. *Plant Prod. Sci.* **2007**, *10*, 400–411. [\[CrossRef\]](#)
2. Xia, L.L.; Yan, X.Y. How to feed the world while reducing nitrogen pollution. *Nature* **2023**, *613*, 34–35. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Sun, X.; Zhang, Y.L.; Shi, K.; Zhang, Y.B.; Li, N.; Wang, W.J.; Huang, X.; Qin, B.Q. Monitoring water quality using proximal remote sensing technology. *Sci. Total Environ.* **2022**, *803*, 149805. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Abel, B.; Cruz-TiradoRaúl, J.P.; Roberto, S.; Quevedo, R. Determination of starch content in adulterated fresh cheese using hyperspectral imaging. *Food Biosci.* **2018**, *21*, 14–19. [\[CrossRef\]](#)
5. Liu, C.C.; Chu, Z.J.; Weng, S.Z.; Zhu, G.Q.; Han, K.X.; Zhang, Z.X.; Huang, L.S.; Zhu, Z.D.; Zheng, S.G. Fusion of electronic nose and hyperspectral imaging for mutton freshness detection using input-modified convolution neural network. *Food Chem.* **2022**, *385*, 132651. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Van, D.A.D.J.; Kootstra, G.; Kruijer, W.; Ridder, D. Machine learning in plant science and plant breeding. *iScience* **2021**, *24*, 101890. [\[CrossRef\]](#)
7. Hesami, M.; Jones, A.M.P. Application of artificial intelligence models and optimization algorithms in plant cell and tissue culture. *Appl. Microbiol. Biotechnol.* **2022**, *104*, 9449–9485. [\[CrossRef\]](#)
8. Singh, A.; Ganapathysubramanian, B.; Singh, A.K.; Sarkar, S. Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.* **2016**, *21*, 110–124. [\[CrossRef\]](#)
9. Marziyeh, J.; Alireza, S. The application of artificial neural networks in modeling and predicting the effects of melatonin on morphological responses of citrus to drought stress. *PLoS ONE* **2020**, *15*, e0240427. [\[CrossRef\]](#)
10. Mohsen, H.; Milad, A.; Andrew, M.P.J.; Davoud, T. Machine learning: Its challenges and opportunities in plant system biology. *Appl. Microbiol. Biotechnol.* **2022**, *106*, 3507–3530. [\[CrossRef\]](#)
11. Grinblat, G.L.; Uzal, L.C.; Lares, M.G.; Granitto, P.M. Deep learning for plant identification using vein morphological patterns. *Comput. Electron. Agric.* **2016**, *127*, 418–424. [\[CrossRef\]](#)
12. Mishra, B.; Kumar, N.; Mukhtar, M.S. Systems biology and machine learning in plant-pathogen interactions. *Mol. Plant-Microbe Interact.* **2018**, *32*, 45–55. [\[CrossRef\]](#)
13. Chen, Y.S.; Lin, Z.H.; Zhao, X.; Wang, G.; Gu, Y.F. Deep learning-based classification of hyperspectral data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote* **2014**, *7*, 2094–2107. [\[CrossRef\]](#)
14. Xu, M.; Wang, Y.; Wang, X.B.; Ding, W.W.; Jia, P.F.; Che, Z.M.; Liu, P. Fermentation process monitoring of broad bean paste quality by NIR combined with chemometrics. *J. Food Meas. Charact.* **2022**, *16*, 2929–2938. [\[CrossRef\]](#)
15. Zhang, X.L.; Lin, T.; Xu, J.F.; Luo, X.; Ying, Y.B. DeepSpectra: An end-to-end deep learning approach for quantitative spectral analysis. *Anal. Chim. Acta* **2019**, *1058*, 48–57. [\[CrossRef\]](#)
16. Jürgen, S. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117.
17. Liu, L.; Li, M.; Zhao, Z.; Qu, J. Recent Advances of Hyperspectral Imaging Application in Biomedicine(Review). *Chin. J. Lasers* **2018**, *45*, 0207017. [\[CrossRef\]](#)



18. Kong, D.D.; Shi, Y.Q.; Sun, D.W.; Zhou, L.; Zhang, W.K.; Qiu, R.C.; He, Y. Hyperspectral imaging coupled with CNN: A powerful approach for quantitative identification of feather meal and fish by-product meal adulterated in marine fishmeal. *Microchem. J.* **2022**, *180*, 107517. [[CrossRef](#)]
19. Yang, X.F.; Ye, Y.M.; Li, X.T.; Lau, R.; Zhang, X.F.; Huang, X.H. Hyperspectral image classification with deep learning models. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5408–5423. [[CrossRef](#)]
20. Zhang, C.; Wu, W.Y.; Zhou, L.; Cheng, H.; Ye, X.Q.; He, Y. Developing deep learning based regression approaches for determination of chemical compositions in dry black goji berries (*Lycium ruthenicum* Murr.) using near-infrared hyperspectral imaging. *Food Chem.* **2020**, *319*, 126536. [[CrossRef](#)]
21. Ni, C.; Wang, D.; Tao, Y. Variable weighted convolutional neural network for the nitrogen content quantization of masson pine seedling leaves with near-infrared spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2018**, *209*, 32–39. [[CrossRef](#)]
22. Yu, S.; Fan, J.C.; Lu, X.J.; Wen, W.L.; Shao, S.; Guo, X.Y.; Zhao, C.J. Hyperspectral technique combined with deep learning algorithm for prediction of phenotyping traits in lettuce. *Front. Plant Sci.* **2022**, *13*, 927832. [[CrossRef](#)]
23. Song, D.; Gao, D.H.; Sun, H.; Qiao, L.; Zhang, R.M.; Tang, W.J.; Li, M.Z. Chlorophyll content estimation based on cascade spectral optimizations of interval and wavelength characteristics. *Comput. Electron. Agric.* **2021**, *189*, 106413. [[CrossRef](#)]
24. Zhang, K.; Kimball, J.S.; Running, S.W. A review of remote sensing based actual evapotranspiration estimation. *Wiley Interdiscip. Rev. Water* **2016**, *3*, 834–853. [[CrossRef](#)]
25. Cao, J.; Zhang, Z.; Tao, F.L.; Zhang, L.L.; Luo, Y.C.; Zhang, J.; Han, J.C.; Xie, J. Integrating multi-source data for rice yield test across China using machine learning and deep learning approaches. *Agric. For. Meteorol.* **2021**, *297*, 108275. [[CrossRef](#)]
26. Wang, J.F.; Lin, T.H.; Ma, S.Y.; Ju, J.Y.; Wang, R.D.; Chen, G.Q.; Jiang, R.; Wang, Z.T. The qualitative and quantitative analysis of industrial paraffin contamination levels in rice using spectral pretreatment combined with machine learning models. *J. Food Compos. Anal.* **2023**, *121*, 105430. [[CrossRef](#)]
27. Yao, X.; Huang, Y.; Shang, G.Y.; Zhou, C.; Cheng, T.; Tian, Y.C.; Cao, W.X.; Zhu, Y. Evaluation of six algorithms to monitor wheat leaf nitrogen concentration. *Remote Sens.* **2015**, *7*, 14939–14966. [[CrossRef](#)]
28. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [[CrossRef](#)]
29. Momeni-Esfahani, M.; Amini, A.S. Assessing the impact of SVR algorithm on determining optimal Landsat satellite bands for measuring lake CDOM. *Iran. J. Remote Sens. GIS* **2021**, 54–65.
30. Soyeurt, H.; Grelet, C.; McParland, S.; Calmels, M.; Coffey, M.; Tedde, A.; Delhez, P.; Dehareng, F.; Gengler, N. A comparison of 4 different machine learning algorithms to predict lactoferrin content in bovine milk from mid-infrared spectra. *J. Dairy Sci.* **2020**, *103*, 11585–11596. [[CrossRef](#)]
31. Sun, Y.W.; Huang, J.L.; Shan, L.X.; Fan, S.T.; Zhu, Z.X.; Liu, X.D. Quantitative analysis of bisphenol analogue mixtures by terahertz spectroscopy using machine learning method. *Food Chem.* **2021**, *352*, 129313. [[CrossRef](#)] [[PubMed](#)]
32. Pyo, J.; Hong, S.M.; Kwon, Y.S.; Kim, M.S.; Cho, K.H. Estimation of heavy metals using deep neural network with visible and infrared spectroscopy of soil. *Sci. Total Environ.* **2020**, *741*, 140162. [[CrossRef](#)]
33. Schonlau, M.; Zou, R.Y. The random forest algorithm for statistical learning. *Stata J.* **2020**, *20*, 3–29. [[CrossRef](#)]
34. Chu, X.L. Chemometrics. In *Stoichiometric Methods and Molecular Spectral Analysis Techniques*, 2nd ed.; Chemical Industry Press: Beijing, China, 2011; pp. 90–95.
35. Li, H.; Liang, Y.; Xu, Q.; Cao, D. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate Train. *Anal. Chim. Acta* **2009**, *648*, 77–84. [[CrossRef](#)] [[PubMed](#)]
36. Li, H.; Hassan, M.M.; Wang, J.J.; Wei, W.Y.; Zou, M.; Ouyang, Q.; Chen, Q.S. Investigation of nonlinear relationship of surface enhanced Raman scattering signal for robust test of thiabendazole in apple. *Food Chem.* **2020**, *339*, 127843. [[CrossRef](#)]
37. Viscarra-Rossel, R.A.; Walvoort, D.J.J.; McBratney, A.B.; Janik, L.J.; Skjemstad, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2006**, *131*, 59–75. [[CrossRef](#)]
38. Wang, J.F.; Chen, G.Q.; Ju, J.Y.; Lin, T.H.; Wang, R.D.; Wang, Z.T. Characterization and classification of urban weed species in northeast China using terrestrial hyperspectral images. *Weed Sci.* **2023**, *71*, 353–368. [[CrossRef](#)]
39. Wang, Z.T.; Tang, Y.R.; Liu, Y.; Zhang, H.; Zhang, Y.C.; Lan, H.P. Inhibitory effect of CaCl<sub>2</sub> and carboxymethyl chitosan coating on the after-ripening of korla fragrant pears in cold storage. *Int. J. Food Sci. Technol.* **2021**, *56*, 6777–6790. [[CrossRef](#)]
40. Yang, J.; Xu, J.F.; Zhang, X.L.; Wu, C.Y.; Lin, T.; Ying, Y.B. Deep learning for vibrational spectral analysis: Recent progress and a practical guide. *Anal. Chim. Acta* **2019**, *1081*, 6–17. [[CrossRef](#)]
41. Yang, B.H.; Ma, J.F.; Yao, X.; Cao, W.X.; Zhu, Y. Estimation of leaf nitrogen content in wheat based on fusion of spectral features and deep features from near infrared hyperspectral imagery. *Sensors* **2021**, *21*, 613. [[CrossRef](#)]
42. Jia, M.; Colombo, R.; Rossini, M.; Celesti, M.; Zhu, J.; Cogliati, S.; Cheng, T.; Tian, Y.; Zhu, Y.; Cao, W. Estimation of leaf nitrogen content and photosynthetic nitrogen use efficiency in wheat using sun-induced chlorophyll fluorescence at the leaf and canopy scales. *Eur. J. Agron.* **2021**, *122*, 126192. [[CrossRef](#)]
43. Zhang, Y.; Wang, T.; Guo, Y.; Skidmore, A.; Zhang, Z.; Tang, R.; Song, S.; Tang, Z. Estimating community-level plant functional traits in a species-rich alpine meadow using UAV image spectroscopy. *Remote Sens.* **2022**, *14*, 3399. [[CrossRef](#)]
44. Abdel-Rahman, E.M.; Ahmed, F.B.; Ismail, R. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *Int. J. Remote Sens.* **2013**, *34*, 712–728. [[CrossRef](#)]

45. Zhai, Y.; Cui, L.; Zhou, X.; Gao, Y.; Fei, T.; Gao, W. Estimation of nitrogen, phosphorus, and potassium contents in the leaves of different plants using laboratory-based visible and near-infrared reflectance spectroscopy: Comparison of partial least-square regression and support vector machine regression methods. *Int. J. Remote Sens.* **2013**, *34*, 2502–2518.
46. Min, M.; Lee, W. Determination of significant wavelengths and prediction of nitrogen content for citrus. *Trans. ASAE* **2005**, *48*, 455–461. [[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.