Article

Estimation of Leaf Nitrogen Content in Wheat Based on Fusion of Spectral Features and Deep Features from Near Infrared Hyperspectral Imagery

Baohua Yang [1,2]**, Jifeng Ma** [1]**, Xia Yao** [1]**, Weixing Cao** [1] **and Yan Zhu** [1,*]

[1] National Engineering and Technology Center for Information Agriculture/Collaborative Innovation Center for Modern Crop Production/Jiangsu Collaborative Innovation Center for the Technology and Application of Internet of Things, Nanjing Agricultural University, Nanjing 210095, China; ybh@ahau.edu.cn (B.Y.); majifeng@njau.edu.cn (J.M.); yaoxia@njau.edu.cn (X.Y.); caow@njau.edu.cn (W.C.)

[2] School of Information and Computer, Anhui Agricultural University, Hefei 230036, China

[*] Correspondence: yanzhu@njau.edu.cn

Abstract: Nitrogen is an important indicator for monitoring wheat growth. The rapid development and wide application of non-destructive detection provide many approaches for estimating leaf nitrogen content (LNC) in wheat. Previous studies have shown that better results have been obtained in the estimation of LNC in wheat based on spectral features. However, the lack of automatically extracted features leads to poor universality of the estimation model. Therefore, a feature fusion method for estimating LNC in wheat by combining spectral features with deep features (spatial features) was proposed. The deep features were automatically obtained with a convolutional neural network model based on the PyTorch framework. The spectral features were obtained using spectral information including position features (PFs) and vegetation indices (VIs). Different models based on feature combination for evaluating LNC in wheat were constructed: partial least squares regression (PLS), gradient boosting decision tree (GBDT), and support vector regression (SVR). The results indicate that the model based on the fusion feature from near-ground hyperspectral imagery has good estimation effect. In particular, the estimation accuracy of the GBDT model is the best (R^2 = 0.975 for calibration set, R^2 = 0.861 for validation set). These findings demonstrate that the approach proposed in this study improved the estimation performance of LNC in wheat, which could provide technical support in wheat growth monitoring.

Keywords: convolutional neural network; leaf nitrogen content; deep features; wheat; spectral features

1. Introduction

Wheat occupies an important position in agricultural production and strategic food reserves. Nitrogen is one of the main nutrients that affect wheat growth, yield and quality [1]. Therefore, rapid and accurate detection of the wheat nitrogen nutrition status is of great significance for guiding farmland management and improving wheat production efficiency, yield and quality [2].

At present, the implementation of remote sensing technology in precision agriculture provides new opportunities for the non-destructive real-time diagnosis of wheat nitrogen status and precise nitrogen control [3]. Among them, feature extraction has gradually become a key technology for non-destructive monitoring and diagnosis of crop nutrition, which has greatly expanded the feature expression ability of the crop canopy [4–8]. Previous studies have utilized many feature extraction methods, including principal component analysis (PCA) [9,10], neighborhood preserving embedding (NPE) [11], and linear discriminant analysis (LDA) [12]. In addition, full bands or characteristic bands extracted from the hyperspectral image have also been used to monitor crop growth, which obtained good

estimation results. Leemans et al. successfully used the spectral features of the hyperspectral image to estimate the nitrogen content of wheat [13]. Mutanga et al. used the spectral features extracted by the depth analysis of the band to estimate the physiological and biochemical parameters of a variety of crops [14]. Although various methods of extracting spectral features have been proposed, the estimation model based only on spectral features has certain limitations. In particular, the inversion model of crop biochemical parameters based on spectral information is prone to saturation when the vegetation coverage is large [1]. Therefore, improving the accuracy of the estimation model still faces many difficulties.

Hyperspectral images (HSI) could provide not only spectral information, but also spatial information. Previous studies have shown that spatial feature extraction methods, such as discrete wavelet transform (DWT) [15,16], Gabor filtering [17,18], and local binary patterns (LBP) [19] have been successfully applied. However, most of the methods mentioned above closely rely on expert knowledge, which limits the predictive potential of the model [20]. Zheng et al. combined spectral features and spatial features to improve the effect of monitoring nitrogen nutrition in wheat [21]. However, facing the rapid growth of the information in hyperspectral images, it is difficult to achieve an optimal balance between typicality and robustness based on methods of feature extraction with prior knowledge. Therefore, it was still a big challenge to extract spatial features from HIS to monitor the nitrogen nutrition of wheat.

In recent years, as an important branch of artificial intelligence, deep learning could solve complex problems with deep neural network models [22]. Moreover, the deep features extracted by deep learning methods have greatly improved the cognitive ability of the network [23,24]. Although the deep features are very abstract and most of them cannot be easily displayed, they are still being valued by more and more scholars. Pan et al. proposed a multi-grained scanning strategy to construct a multi-grained network (MugNet), which realized the extraction of deep features of hyperspectral images to improve the classification accuracy [25]. Chen et al. proposed a deep learning framework based on the fusion of spectral information and spatial information to extract the deep features of hyperspectral images [26]. Xu extracted the deep features of the spectrum to improve the detection of different types of Pu'er tea [27]. Yang et al. designed different structures of stacked autoencoder (SAE) to extract the deep features of hyperspectral images, which improved the accuracy of the estimation of soluble solid content in peach [28]. The research mentioned above showed that deep features could effectively improve the classification, recognition, and prediction of target objects.

In addition, as a representative algorithm of deep learning, CNN has received more and more attention due to its good results in the field of image detection. In comparison with traditional features, deep features are automatically learned layer by layer from spatial features through CNN, which is widely used in crop qualitative analysis. In particular, deep features were obtained based on CNN to detect wheat spikes [29] and crop nitrogen deficiency [30], as well as to obtain crop high-throughput phenotypic features [31,32]. The above research results show that CNN overcomes the limitations of traditional machine learning methods, which provides a new idea for HSI to estimate the leaf nitrogen content (LNC). As far as we know, the quantitative analysis of physiological and biochemical parameters in wheat using deep features extracted by CNN has not been reported in the literature.

Therefore, we used CNN to extract deep features from hyperspectral images of the wheat canopy, and constructed PLSR, SVR and GBDT models using deep features, spectral features, and fusion features to verify the estimation effects of different features. The purpose of this research was to (1) extract deep features from hyperspectral images to overcome the limitations of feature expression; (2) fuse spectral features and deep features to improve the estimation accuracy of the model; (3) evaluate the performance of different models to test the validity of the model

2. Data and Methods

2.1. Study Site and Experimental Design

The experiments were carried out in Rugao Experimental Demonstration Base of the National Information Agriculture Engineering Technology Center in 2013 and 2014 (120°20′ E, 32°14′ N, Rugao City, Jiangsu Province), as shown in Figure 1. Rugao belongs to the subtropical monsoon climate zone. The annual average temperature and annual rainfall are 15.11 °C and above 1000 mm, respectively, which is very beneficial to the growth of wheat. The experiments were implemented in a total of 24 plots (7 m × 5 m for each plot), and the implementation of nitrogen fertilizer was based on three levels, including 0 (N0), 150 (N1) and 300 (N2) kg/ha, with two planting densities (300 plants·m^{-2} and 450 plants·m^{-2}), as shown in Figure 1. The experimental varieties were 'Yangmai 18' and 'Shengxuan 6'. Nitrogen fertilizer was applied according to 40% base fertilizer, 10% tiller fertilizer, 20% flower promoting fertilizer, and 30% flower retaining fertilizer. The basal fertilizer was combined with 135 kg/ha P_2O_5 phosphate fertilizer and 190 kg/ha K_2O potassium fertilizer. The near-ground hyperspectral image and wheat plant samples were acquired simultaneously. The key periods of acquisition include the jointing, heading, flowering, and filling period.
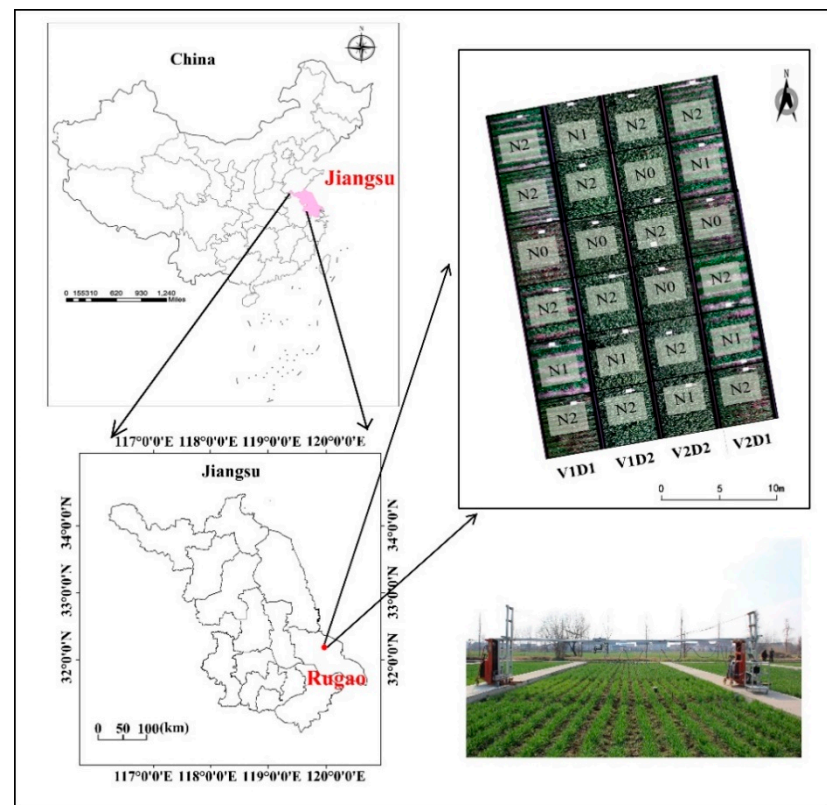
Figure 1. Location of the study area and hyperspectral imaging system. Note: D1 = 300 plants·m^{-2}, D2 = 450 plants·m^{-2}, N0 = 0 kg·N·ha^{-1}, N1 = 150 kg·N·ha^{-1}, N2 = 300 kg·N·ha^{-1}, V1 = 'Yangmai 18', V2 = 'Shengxuan 6'.

2.2. Data Collection

All near-infrared hyperspectral images were collected with a pushbroom scanning sensor (V10E-PS, SpecIm, Oulu, Finland) mounted on a motorized rail. The sensor was about 1.0 m above the wheat canopy. The data obtained through the hyperspectral imaging system include hyperspectral images with 1392 × 1040 pixels and a total of 520 bands which range from 360 to 1025 nm with a spectral resolution of 2.8 nm. The spatial resolution and field of view for near-nadir observation were 1.3 mm and 42.8°. The original images were processed by the software specVIEW (SpecIm, Oulu, Finland) [33].

On the same day when the wheat canopy hyperspectral image was obtained in the field experiment, 20 wheat plants were randomly selected from each sampling area of the experimental base and brought back to the laboratory. First, the wheat was separated according to different organs (leaf, stem, and ear) as experimental samples. Secondly, all samples were placed in an oven at 105 °C for 30 min and dried at 80 °C for more than 20 h. Finally, they were weighed to obtain the dry weight of each sample. The samples were crushed and the N content in the leaf, stem and ear was separately determined by the Kjeldahl method [34].

2.3. Spectral Features and Deep Features

2.3.1. Vegetation Indices

Vegetation indices are the combination of linear and non-linear features in the visible–near-infrared band, which is one of the most widely used indicators in crop growth monitoring, and reflects the growth of crops under certain conditions. In this study, 26 VIs related to LNC in wheat were selected, for which calculation formulas are shown in Table 1.

**Table 1.** The calculation formulas for vegetation indices.

Index	Formula	Reference
$NDVI_{g-b}$[#]	$(R_{573} - R_{440})/(R_{573} + R_{440})$	[35]
$DCNI$[#]	$(R_{720} - R_{700})/(R_{700} - R_{670})/(R_{720} - R_{670} + 0.03)$	[36]
$NDVI$ I	$(R_{800} - R_{670})/(R_{800} + R_{670})$	[37]
RVI I	R_{800}/R_{670}	[38]
DVI I	$R_{800} - R_{670}$	[39]
$SAVI$ I	$1.5 \times (R_{800} - R_{670})/(R_{800} + R_{670} + 0.5)$	[40]
$NDRE$	$(R_{790} - R_{720})/(R_{790} + R_{720})$	[41]
DVI II	$R_{NIR} - R_R$	[42]
EVI	$\frac{2.5 \times (R_{NIR} - R_R)}{(R_{NIR} + 6R_R - 7.5R_B + 1)}$	[43]
$NDVI$ II	$(R_{NIR} - R_R)/(R_{NIR} + R_R)$	[44]
$MSAVI2$	$\left(2R_{NIR} + 1 - sqrt\left((2R_{NIR} + 1)^2 - 8(R_{NIR} - R_R)\right)\right)/2$	[45]
$OSAVI$	$\frac{(1+0.16)(R_{NIR} - R_R)}{R_{NIR} + R_R + 0.16}$	[46]
RVI II	R_{NIR}/R_R	[47]
$SAVI$ II	$\frac{1.5 \times (R_{NIR} - R_R)}{R_{NIR} + R_R + 0.5}$	[48]
TVI	$60 \times (R_{NIR} - R_G) - 100 \times (R_R - R_G)$	[49]
$MTVI2$	$\frac{1.5*(1.2*(R_{NIR} - R_G) - 2.5*(R_R - R_G))}{\sqrt{\left((2*R_{NIR}+1)^2 - \left(6*R_{NIR} - 5*\sqrt{(R_R)} - 0.5\right)\right)}}$	[50]
$GNDVI$	$(R_{NIR} - R_R)/(R_{NIR} + R_R)$	[51]
MSR	$(R_{NIR}/R_R - 1)/(R_{NIR}/R_R + 1)$	[52]
$ARVI$	$\frac{R_{NIR} - R*B}{R_{NIR} + R - (B-R)}$	[53]
$VOG1$	R_{740}/R_{720}	[54]
$VOG2$	$(R_{734} - R_{747})/(R_{715} + R_{726})$	[54]
$VOG3$	$(R_{734} - R_{747})/(R_{715} + R_{720})$	[54]
PRI	$(R_{531} - R_{570})/(R_{530} + R_{570})$	[55,56]
$NPCI$	$(R_{680} - R_{430})/(R_{680} + R_{430})$	[57]
$SIPI$	$(R_{800} - R_{445})/(R_{800} - R_{680})$	[58]
$PSRI$	$(R_{680} - R_{500})/R_{750}$	[59]

Where R is the reflectance. I, II, III, IV, and V are only used to distinguish the same vegetation indices of different wave bands.

2.3.2. Position Features

The crop canopy has strong absorption and reflection characteristics in the visible and near-infrared bands, which are related to the physiological and biochemical components of the crop [60]. To enhance the absorption characteristics of LNC and remove the influence of soil and other background spectrum absorption, the continuum removal method was used to process the canopy reflectance spectrum in wheat.

In this study, the ENVI software was used to extract the spectral information (400–1000 nm) of the hyperspectral image, and then the continuum removal method was used to further extract the spectral position features, which generally include absorption characteristic and reflection characteristic [61]. The spectral reflectance curve and the continuum removal curve are shown in Figure 2. It can be seen from Figure 2 that the reflection characteristic bands are distributed at 500–721 and 753–959 nm, and the spectral absorption characteristic bands are distributed at 557–754 and 900–1030 nm. Within the green dotted line and the black dotted line are the reflection and absorption positions of the spectral features, respectively. The calculation formulas of the six characteristic parameters are shown in Table 2 [61,62].
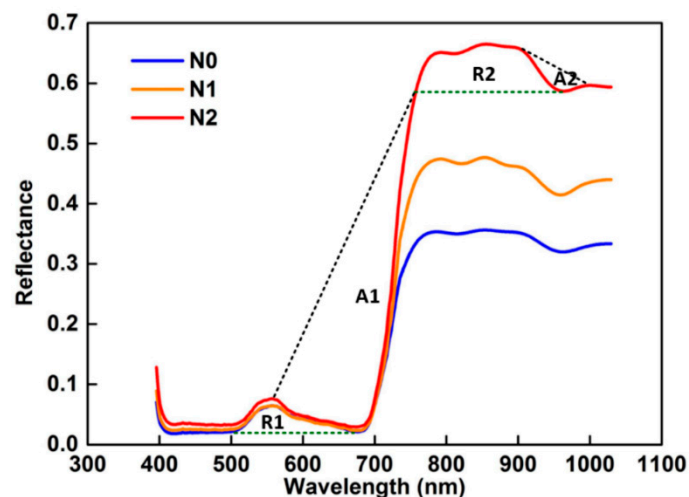
Figure 2. Characteristic absorption and reflection positions of the winter wheat for the three nitrogen treatments. (A) Absorption position, and (R) reflection position.

Table 2. Characteristic parameters for absorption and reflection positions.

Variables	Calculation Formula
$A\text{-}Depth_i$	$1 - R_i(\lambda_{\min})/R_{ci}(\lambda_{\min})$
$A\text{-}Area_i$	$\int_{\lambda_j}^{\lambda_k}(R_{ci}(\lambda) - R_i(\lambda))d\lambda$
$A\text{-}ND_i$	$A\text{-}Depth_i/A\text{-}Area_i$
$R\text{-}Depth_i$	$1 - R_{ci}(\lambda_{\max})/R_i(\lambda_{\max})$
$R\text{-}Area_i$	$\int_{\lambda_j}^{\lambda_k}(R_i(\lambda) - R_{ci}(\lambda))d\lambda$
$R\text{-}ND_i$	$R\text{-}Depth_i/R\text{-}Area_i$

R_{ci} is the continuum removal curve, R_i is the original reflectance curve, λ is the wavelength, λ_j and λ_k are the initial and final wavelengths of each absorption and reflection region, respectively, and the index i is the number of the corresponding band.

In addition, position features also include band position features, which include the red edge position, yellow edge position, and blue edge position, etc. The waveband position parameters used in this study are shown in Table 3, with a total of 13 features [61].

Table 3. Definition and description of waveband position parameters.

Variables	Names	Definition and Description
Db	Blue edge amplitude	Maximum value of the 1st derivative of a blue edge (490–530 nm)
λb	Blue edge position	Wavelength at Db
Dy	Yellow edge amplitude	Maximum value of the 1st derivative of a yellow edge (560–640 nm)
λy	Yellow edge position	Wavelength at Dy
Dr	Red edge amplitude	Maximum value of the 1st derivative with a red edge (680–760 nm)
λr	Red edge position	Wavelength at Dr
Rg	Green peak amplitude	Maximum reflectance of a green peak (510–560 nm)
λg	Location of green peak	Wavelength at Rg
Ro	Red valley amplitude	Lowest reflectance of a red well (650–690 nm)
λo	Red valley position	Wavelength at Ro
SDb	Blue-edge integral areas	Sum of the 1st derivative values within the blue edge
SDy	Yellow-edge integral areas	Sum of the 1st derivative values within the yellow edge
SDr	Red-edge integral areas	Sum of the 1st derivative values within the red well

2.3.3. Deep Features

In this study, deep features were extracted using a convolutional neural network (CNN), which is a deep feedforward artificial neural network including a convolutional layer, pooling layer, and fully connected layer [30]. Due to the large number of features extracted by CNN and the relatively small dataset used in this study, the model is prone to overfitting if the data are directly input to the network for training. Therefore, a transfer learning method (a machine learning method) was used to extract deep features from hyperspectral images using the trained CNN named AlexNet, which is a convolutional neural network proposed by Krizhevsky et al. in 2012, who designed and applied the network, and won the championship in the ILSVRC competition [63]. Moreover, AlexNet was trained on a subset of the large ImageNet database.

The convolutional neural network model is shown in Figure 3. It can be seen from Figure 3 that the number of feature maps and kernels in each layer is different, and the structure of CNN (AlexNet) includes five convolutional layers, three pooling layers and two fully connected layers. A total of 256 dimension feature vectors are extracted through FC2 as deep features. This experiment was performed based on a Pytorch framework with Windows system. The specific parameters are as follows: an Intel Core i7-8700 @ 3.20 GHz × 6, a memory of 16 GB, and a GPU of Nvidia GeForce RTX 2080.
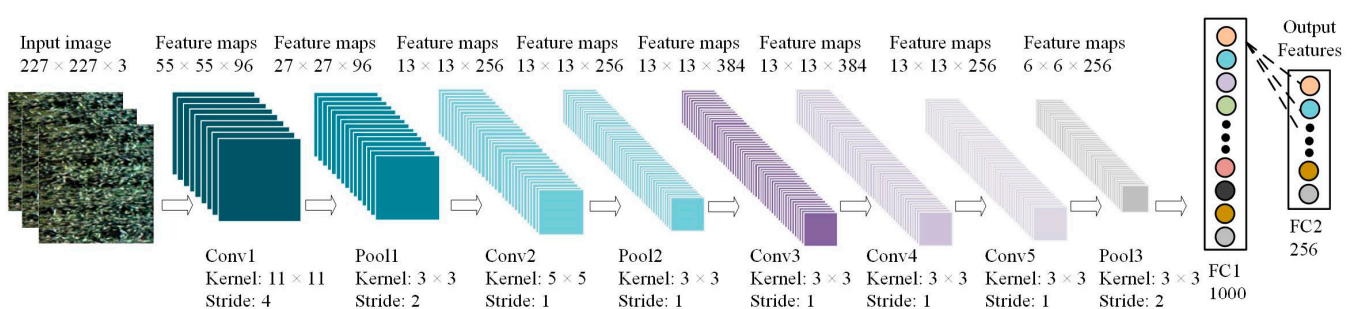
Figure 3. Convolutional neural network structure. Kernel represents the size of the convolution kernel (pixels), Stride represents the sliding step size, Conv1, Conv2, Conv3, Conv4, Conv5 represent the first, second, third, fourth, and fifth convolutional layers, and Pool1, Pool2, and Pool3 represent the first, second, and third pooling layer, FC1 and FC2 represent fully connected layer.

2.4. Feature Optimization Method

2.4.1. Random Forest Algorithm

To obtain features which have a higher contribution to the estimation model, random forest (RF) was used to optimize the features. The random forest algorithm is a multi-

classifier algorithm based on ensemble learning, which consists of a decision tree and a bagging algorithm [64]. Bagging is a process in which sub-sampling (collected samples) is put back into random sampling again [65]. Then, each round of different random sampling is used for model training, and other unsampled data are used as out-of-bag (OOB) to verify the model. Generally, random sampling twice can ensure that the model has strong generalization [64]. RF-based feature selection is mainly achieved by directly measuring the influence of each feature on the accuracy of the model by reducing the average accuracy.

2.4.2. Pearson Correlation Coefficient Method

Pearson correlation analysis is used to obtain the correlation coefficient (r) between variables, which reflects the degree of correlation. For two vectors X and Y of the same dimension, the Pearson correlation coefficient calculation formula is as follows:

$$r = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum (X - \overline{X})^2 (Y - \overline{Y})^2}} \tag{1}$$

where, $\overline{X}$ is the mean of X, and $\overline{Y}$ is the mean of Y.

The value of r ranges from -1 to 1. When r is less than zero, it means that the two vectors are linearly negatively correlated. However, in terms of evaluating the importance of features, the correlation between two vectors has been paid more attention, and its absolute value ($|r|$) shows the strength of the correlation.

2.5. Regression Method

2.5.1. Partial Least Squares Regression

Partial least squares (PLS) is a multivariate statistical data analysis method. Partial least squares regression (PLSR) is an extended form of multiple linear regression model, which combines multiple linear regression, canonical correlation analysis, and principal component analysis, and extracts principal components through variable mapping methods. PLSR can avoid multiple linear relationships between independent variables and dependent variables, which can improve the accuracy, robustness, and practicability of the model [66].

2.5.2. Support Vector Regression

Support vector regression (SVR) is based on the principle of structural risk minimization, which uses a small number of support vectors to represent the entire sample set [67]. When the SVR model is established, different parameters need to be set including the penalty coefficient C and the kernel parameter g. It is easy to cause over-fitting and under-fitting of the model if the value of C is too large or too small, and g determines the distribution of the data after mapping to the new feature space. The value of g indicates the number of support vectors.

2.5.3. Gradient Boosting Decision Tree

Gradient boosting decision tree (GBDT) is an important algorithm in ensemble learning, which combines multiple decision trees to build a more powerful model [68]. GBDT can effectively perform regression estimation by constructing and combining multiple learners to complete the learning task, which is usually more sensitive to parameter settings. Among them, a more important parameter is learning_rate, which is used to control the strength of each tree to correct the error of the previous tree. A higher learning_rate means that each tree can make stronger corrections, which makes the model more complicated.

In this study, the data obtained in 2013 were used as the calibration set of the model, and the data obtained in 2014 were used as the validation set. The coefficient of determination (R^2) and root mean square error (RMSE) were used as evaluation indicators to evaluate all established regression models, which include SVR, GBDT and PLSR. Generally, universal models have relatively higher R^2 and smaller RMSE [69].

3. Results and Analysis

3.1. Optimization of Vegetation Indices

To reduce the information redundancy and improve the accuracy of the model, the random forest algorithm was used to analyze the relative importance of 26 vegetation indices, as shown in Figure 4. The relative importance of the vegetation indices was ranked in descending order, and the top 30% are selected as the preferred VIs, which are NDVI $_{g\text{-}b}$[#], SIPI, NPCI, VOG3, VOG2, RVI I, SAVI II and MTVI2.
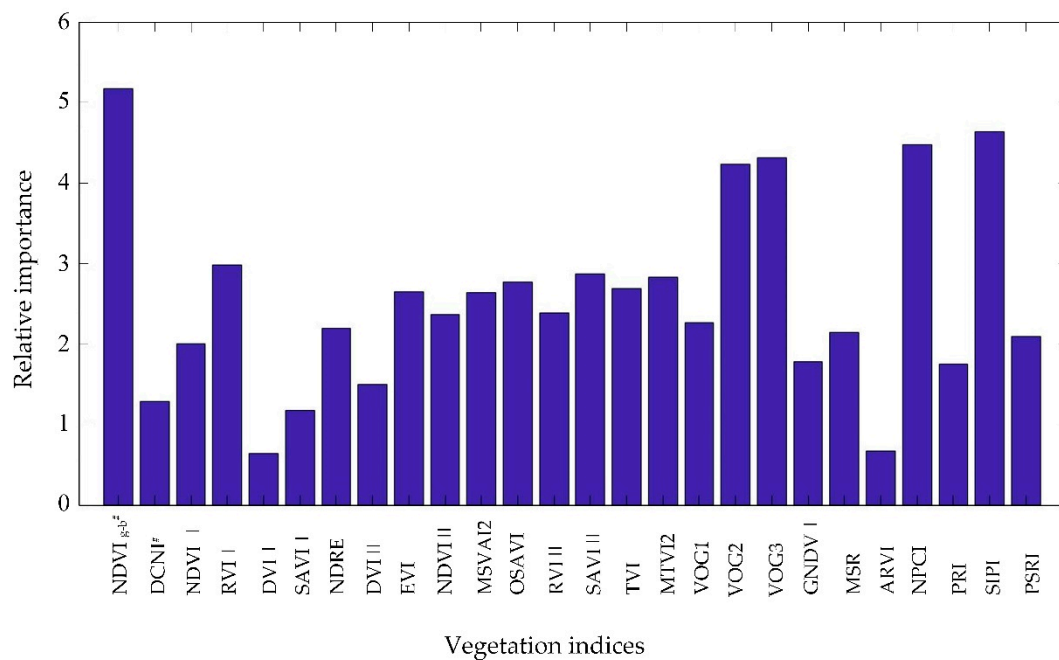

**Figure 4.** Feature importance of vegetation indices based on the random forest algorithm.

3.2. Optimization of Position Features

To extract sensitive position features, the correlation coefficients of 25 position features with LNC were calculated using Pearson correlation coefficient method. The results are shown in Figure 5. The first row in the triangular matrix represents the correlation between position features and LNC, and the rest represent the correlation between position features. The first 30% of the larger absolute value of the correlation coefficient are selected as the preferred position features, including Rg, R_Depth1, R_Aear1, R_ND1, A_Depth1, A_Aear1, A_ND1, and the corresponding correlation coefficients are −0.79, 0.73, 0.85, −0.81, 0.73, 0.87, and −0.89, respectively.
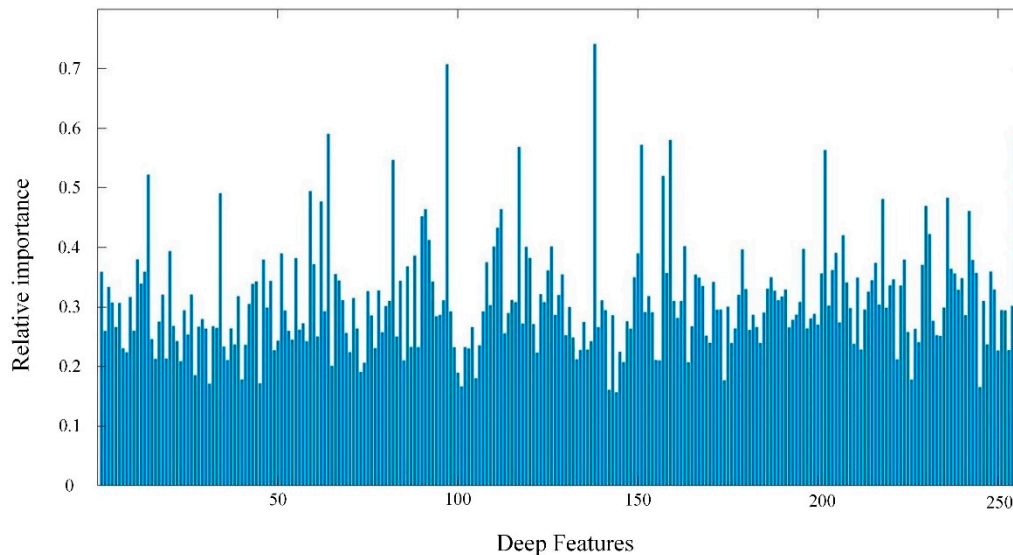
3.3. Optimization of Deep Features

The hyperspectral image of wheat canopy with 256 dimensional deep features is extracted by convolutional neural network. To more effectively estimate LNC of wheat, random forest algorithm was used to optimize deep features. As shown in Figure 6, the relative importance of these deep features is different. The first 20 features with relative importance greater than 0.45 were selected and merged with the spectral features for subsequent modeling research.
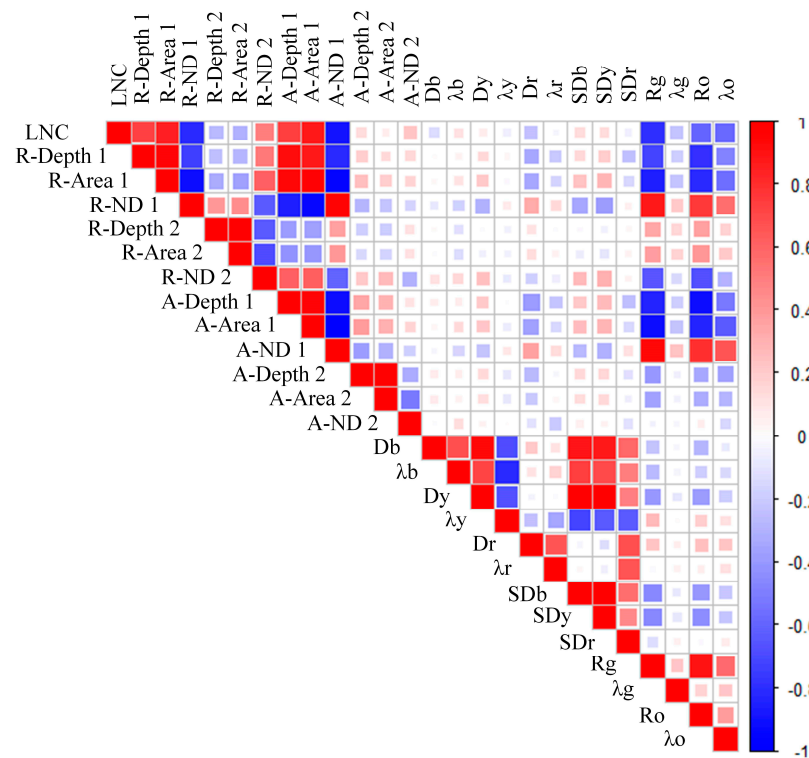
Figure 5. Matrix of correlation coefficient between leaf nitrogen content (LNC) and position features.

Figure 6. Optimal selection of deep features based on random forest algorithm.

3.4. Comparison of Models for Estimating LNC in Winter

Different features were obtained from the near-ground imaging spectroscopy of wheat canopy based on different methods, including spectral features and deep features, to construct different models for estimation LNC, which includes PLS, GBDT, and SVR models. The results of the model comparison are shown in Table 4. For the PLS model, R^2 ranges from 0.791 to 0.895 for the calibration set and from 0.708 to 0.814 for the validation set. For the SVR model, R^2 is 0.791–0.954 and 0.659–0.842. For the GBDT model, R^2 is 0.848–0.975 for the calibration set and 0.717–0.861 for the validation set. The results show that the estimations based on the PLS, GBDT and SVR models all perform well.

Table 4. Estimating model for wheat LNC from the selected input variables with three machine learning techniques.

Model	Features	Preferred Variables	Calibration Set		Validation Set	
			R^2	RMSE	R^2	RMSE
PLS	VIs	8	0.791	0.448	0.708	0.439
	PFs	7	0.812	0.421	0.722	0.392
	DFs	20	0.867	0.352	0.794	0.330
	FFs	35	0.895	0.313	0.814	0.328
SVR	VIs	8	0.791	0.442	0.659	0.449
	PFs	7	0.809	0.448	0.703	0.416
	DFs	20	0.897	0.325	0.780	0.367
	FFs	35	0.954	0.209	0.842	0.312
GBDT	VIs	8	0.848	0.148	0.717	0.384
	PFs	7	0.853	0.137	0.77	0.386
	DFs	20	0.927	0.084	0.832	0.303
	FFs	35	0.975	0.01	0.861	0.263

Where VIs represents vegetation indices, PFs represents position features, DFs represents deep features, DFs represents deep features, and FFs represents fusion features.

From the comparison of the models based on different features shown in Figure 7, we could find that the R^2 of the model ranges from 0.791 to 0.848 based on VIs, from 0.809 to 0.853 based on PFs, from 0.867 to 0.927 based on DFs, and from 0.895 to 0.975 based on FFs for the calibration set. The corresponding R^2 of the model are 0.659–0.717, 0.703–0.77, 0.78–0.832, and 0.814–0.861 for the validation set. Therefore, regardless of the validation set or calibration set, the estimation results of model with a combination of features perform better than a single feature.

4. Discussion

4.1. Deep Features and Spectral Features

With the improvement of the ability to collect and store data, this has brought about many difficulties in data reduction and analysis. Feature extraction methods for higher-resolution spectral information are becoming increasingly important. In this study, the vegetation index provided important information for quantifying wheat LNC. However, only a few wavelengths of reflectance are used in the vegetation index, which affects the robustness of the model for monitoring wheat nitrogen nutrition [70]. In order to highlight the difference between the spectral absorption features of LNC, the continuum removal transform is used to mine more potential information of spectral position features, which can not only solve the saturation problem of the existing index, but also effectively reduce the influence of the background on the spectral features [14].

Moreover, the expression of features in high-dimensional data is also increasingly important, especially for non-linear features. Traditional features only focus on fewer and relatively obvious features, and these features are not reliable for different input data. In contrast, the deep features extracted from hyperspectral images based on deep learning could express the detailed information of spatial features. To estimate wheat LNC more accurately, spectral information and spatial information were extracted to realize the comprehensive features of hyperspectral images, which overcome the limitations of the traditional single feature.
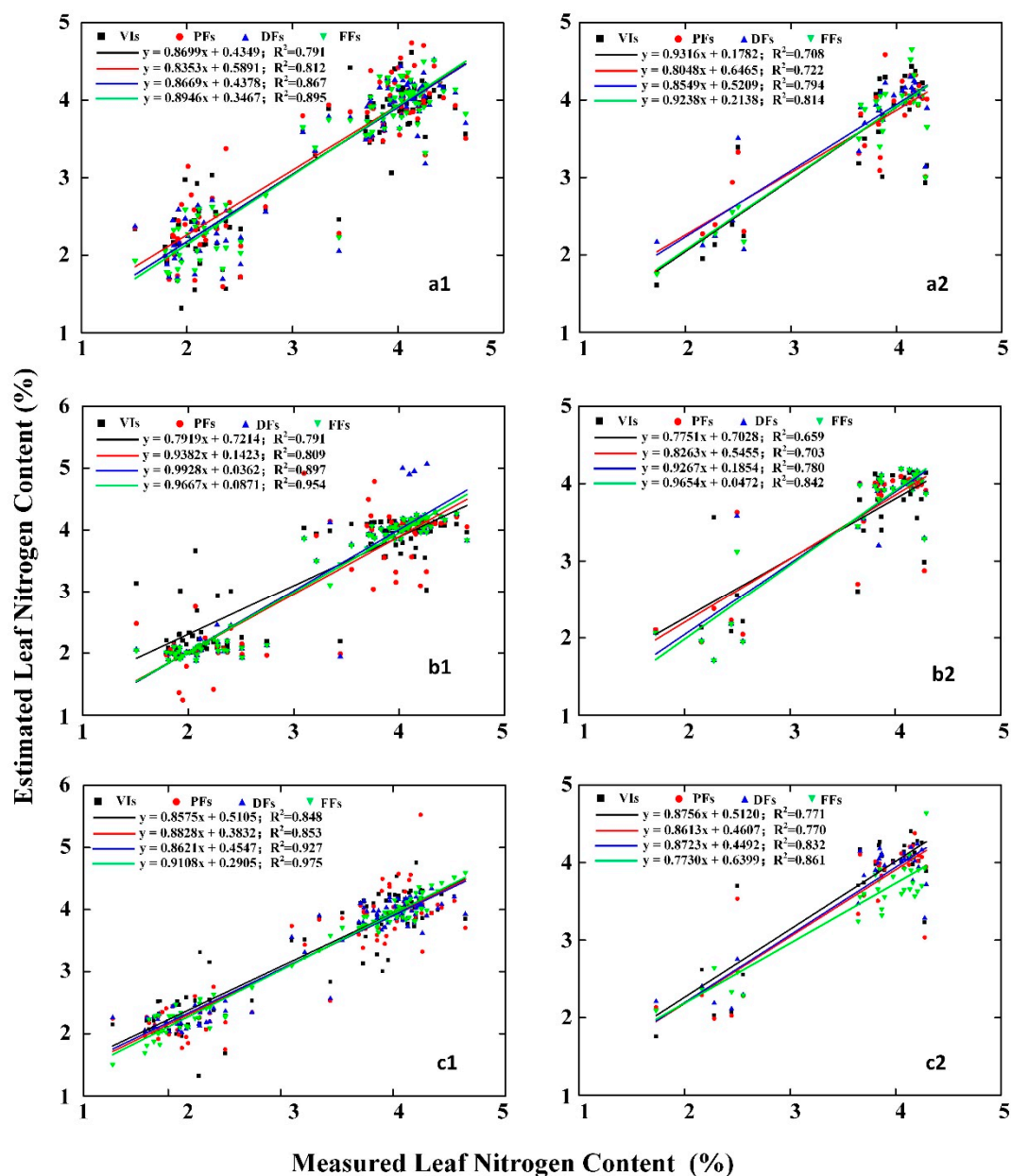
Figure 7. Estimated and measured leaf nitrogen content (%) in wheat. Left: validation set, right: calibration set with PLS (**a1**,**a2**), SVR (**b1**,**b2**), and GBDT (**c1**,**c2**). VIs: vegetation indices, PFs: position features, DFs: deep features, FFs: fusion features.

4.2. The Necessity of Extracting Deep Features from Hyperspectral Images

Previous studies have shown that convolutional neural networks (CNN) can extract deep features. The feature extractor at each layer can use the convolutional layer and the pooling layer to convert the input raw data to complex deep features, thereby reducing the data noise problem caused by external environmental interference [71]. Especially when traditional methods cannot collect enough features to support accurate detection, CNN can still extract more detailed features, which will help to increase the detection potential [23]. Cheng et al. proposed a deep learning model that combines spectral and spatial features, which can effectively extract complex hyperspectral features [72]. Therefore, to extract the features which play a key role in the quantitative analysis of the nitrogen content of wheat leaves from the hyperspectral image, it is necessary to maintain the spatial topological structure of the hyperspectral image of the wheat canopy, and the

deep features are extracted layer by layer through CNN and gradually abstracted based on the constructed convolutional neural network.

Although the CNN method can fully obtain the high-dimensional deep features of the sample, not all of the detected deep features are useful, and the high-dimensional data lead to the complexity of calculation and analysis [73]. In particular, the excessive increase in the network depth of the convolutional neural network will lead to some negative effects, such as overfitting, disappearance of gradients, and decreased accuracy [74]. Therefore, to solve the general problem of imbalance between the limitation and availability of high-dimensional deep features, the random forest algorithm was used to eliminate irrelevant and redundant features of deep features. Thus, the complexity of the model was reduced, and the accuracy and generalization ability of the model were improved.

4.3. Different Models and Different Features

Fusion features including deep features and spectral features extracted from hyperspectral imagery can successfully estimate wheat LNC. Fan et al. used the position features and vegetation index to estimate wheat LNC [61]. However, the extracted features only included spectral features. In this study, not only the spectral features (VIs and PFs) were extracted, but also the spatial features (deep features) were extracted based on the convolutional neural network. Table 2 showed that the accuracy of the GBDT model based on fusion features was 8.2% higher than that of the PLS model, and 2.2% higher than that of the SVR model.

In addition, the estimation performances of different regression models (PLSR, SVR, GBDT) were good. Specially, the effect based on the GBDT model was better than the other two regression models. Despite the experimental design of different models and different features, the proposed model achieved good estimation results. In the future, it is necessary to collect more data to verify the method proposed in this work, which can provide technical support for estimating wheat LNC.

5. Conclusions

In this study, a method based on fusion features was proposed to estimate wheat LNC, including spectral features and deep features. On the one hand, fusion features include both spectral features and spatial features, and the extraction of spatial features (deep features) was based on CNN, which could learn spatial hierarchy, including basic features and semantic features. On the other hand, PLS, GBDT, and SVR models were constructed to estimate wheat LNC. The GBDT model has the highest accuracy ($R^2 = 0.975$ for calibration set, $R^2 = 0.861$ for validation set), which provides a new approach for quantitative estimation of wheat LNC from hyperspectral imagery. In addition, it is necessary to further try more deep learning models in future research to provide technical support for crop growth monitoring.

Author Contributions: B.Y. performed the experiments and wrote the paper; J.M. processed the data; X.Y., W.C. and Y.Z. revised the paper. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

References

1. Zhu, Y.; Tian, Y.; Yao, X.; Liu, X.; Cao, W. Analysis of Common Canopy Reflectance Spectra for Indicating Leaf Nitrogen Concentrations in Wheat and Rice. *Plant Prod. Sci.* **2007**, *10*, 400–411. [CrossRef]
2. Yang, B.; Wang, M.; Sha, Z.; Wang, B.; Chen, J.; Yao, X.; Cheng, T.; Cao, W.; Zhu, Y. Evaluation of Aboveground Nitrogen Content of Winter Wheat Using Digital Imagery of Unmanned Aerial Vehicles. *Sensors* **2019**, *19*, 4416. [CrossRef] [PubMed]
3. Rabatel, G.; Al Makdessi, N.; Ecarnot, M.; Roumet, P. A spectral correction method for multi-scattering effects in close range hyperspectral imagery of vegetation scenes: Application to nitrogen content assessment in wheat. *Adv. Anim. Biosci.* **2017**, *8*, 353–358. [CrossRef]
4. He, L.; Zhang, H.-Y.; Zhang, Y.-S.; Song, X.; Feng, W.; Kang, G.-Z.; Wang, C.-Y.; Guo, T.-C. Estimating canopy leaf nitrogen concentration in winter wheat based on multi-angular hyperspectral remote sensing. *Eur. J. Agron.* **2016**, *73*, 170–185. [CrossRef]
5. Vigneau, N.; Ecarnot, M.; Rabatel, G.; Roumet, P. Potential of field hyperspectral imaging as a non destructive method to assess leaf nitrogen content in Wheat. *Field Crop. Res.* **2011**, *122*, 25–31. [CrossRef]
6. Wang, R.; Song, X.; Li, Z.; Yang, G.; Guo, W.; Tan, C.; Chen, L. Estimation of winter wheat nitrogen nutrition index using hyperspectral remote sensing. *Trans. Chin. Soc. Agric. Eng.* **2014**, *30*, 191–198.
7. Liu, H.; Zhu, H.; Li, Z.; Yang, G. Quantitative analysis and hyperspectral remote sensing of the nitrogen nutrition index in winter wheat. *Int. J. Remote. Sens.* **2019**, *41*, 858–881. [CrossRef]
8. Feng, W.; Zhang, H.; Zhang, Y.; Qi, S.; Heng, Y.; Guo, B.; Ma, D.; Guo, T. Remote detection of canopy leaf nitrogen concen-tration in winter wheat by using water resistance vegetation indices from in-situ hyperspectral data. *Field Crop. Res.* **2016**, *198*, 238–246. [CrossRef]
9. Ye, M.; Ji, C.; Chen, H.; Lei, L.; Qian, Y. Residual deep PCA-based feature extraction for hyperspectral image classifica-tion. *Neural Comput. Appl.* **2020**, *32*, 14287–14300. [CrossRef]
10. Uddin, M.P.; Al Mamun, M.; Hossain, M.A. Effective feature extraction through segmentation-based folded-pca for hyper-spectral image classification. *Int. J. Remote Sens.* **2019**, *40*, 7190–7220. [CrossRef]
11. Li, Y.; Ge, C.; Sun, W.; Peng, J.; Du, Q.; Wang, K. Hyperspectral and lidar data fusion classification using superpixel segmen-tation-based local pixel neighborhood preserving embedding. *Remote Sens.* **2019**, *11*, 550. [CrossRef]
12. Bandos, T.V.; Bruzzone, L.; Camps-Valls, G. Classification of Hyperspectral Images With Regularized Linear Discriminant Analysis. *IEEE Trans. Geosci. Remote. Sens.* **2009**, *47*, 862–873. [CrossRef]
13. Leemans, V.; Marlier, G.; Destain, M.-F.; Dumont, B.; Mercatoris, B. Estimation of leaf nitrogen concentration on winter wheat by multispectral imaging. *Hyperspectral Imaging Sens. Innov. Appl. Sens. Stand.* **2017**, 102130I. [CrossRef]
14. Mutanga, O.; Skidmore, A.K. Hyperspectral band depth analysis for a better estimation of grass biomass (Cenchrus ciliaris) measured under controlled laboratory conditions. *Int. J. Appl. Earth Obs. Geoinf.* **2004**, *5*, 87–96. [CrossRef]
15. Ghasemzadeh, A.; Demirel, H. 3D discrete wavelet transform-based feature extraction for hyperspectral face recognition. *IET Biom.* **2018**, *7*, 49–55. [CrossRef]
16. Cao, X.; Xu, L.; Meng, D.; Zhao, Q.; Xu, Z. Integration of 3-dimensional discrete wavelet transform and Markova random field for hyperspectral image classification. *Neurocomputing* **2017**, *226*, 90–100. [CrossRef]
17. Li, H.-C.; Zhou, H.; Pan, L.; Du, Q. Gabor feature-based composite kernel method for hyperspectral image classification. *Electron. Lett.* **2018**, *54*, 628–630. [CrossRef]
18. Jia, S.; Shen, L.; Li, Q. Gabor Feature-Based Collaborative Representation for Hyperspectral Imagery Classification. *IEEE Trans. Geosci. Remote. Sens.* **2015**, *53*, 1118–1129. [CrossRef]
19. Li, W.; Chen, C.; Su, H.; Du, Q. Local binary patterns and extreme learning machine for hyperspectral imagery classifica-tion. *IEEE Trans. Geoence Remote Sens.* **2015**, *53*, 3681–3693. [CrossRef]
20. Deng, Z.P.; Sun, H.; Zhou, S.L.; Zhao, J.P.; Lei, L.; Zou, H.X. Multi-scale object detection in remote sensing imagery with convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 3–22. [CrossRef]
21. Zheng, H.; Li, W.; Jiang, J.; Liu, Y.; Cheng, T.; Tian, Y.; Zhu, Y.; Cao, W.; Zhang, Y.; Yao, X. A Comparative Assessment of Different Modeling Algorithms for Estimating Leaf Nitrogen Content in Winter Wheat Using Multispectral Images from an Unmanned Aerial Vehicle. *Remote. Sens.* **2018**, *10*, 2026. [CrossRef]
22. Alam, F.I.; Zhou, J.; Liew, W.C.; Jia, X.; Chanussot, J.; Gao, Y. Conditional random field and deep feature learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1612–1628. [CrossRef]
23. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. *IEEE Trans. Geosci. Remote. Sens.* **2016**, *54*, 6232–6251. [CrossRef]

24. Liu, B.; Yu, X.; Zhang, P.; Yu, A.; Fu, Q.; Wei, X. Supervised Deep Feature Extraction for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote. Sens.* **2017**, *56*, 1909–1921. [CrossRef]

25. Pan, B.; Shi, Z.W.; Xu, X. MugNet: Deep learning for hyperspectral image classification using limited samples. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 108–119. [CrossRef]

26. Chen, Y.; Lin, Z.; Zhao, X.; Wang, G.; Gu, Y. Deep Learning-Based Classification of Hyperspectral Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **2014**, *7*, 2094–2107. [CrossRef]

27. Xu, S.; Sun, X.; Lu, H.; Zhang, Q. Detection of Type, Blended Ratio, and Mixed Ratio of Pu'er Tea by Using Electronic Nose and Visible/Near Infrared Spectrometer. *Sensors* **2019**, *19*, 2359. [CrossRef]

28. Yang, B.; Gao, Y.; Yan, Q.; Qi, L.; Zhu, Y.; Wang, B. Estimation Method of Soluble Solid Content in Peach Based on Deep Features of Hyperspectral Imagery. *Sensors* **2020**, *20*, 5021. [CrossRef]

29. Hasan, M.; Chopin, J.P.; Laga, H.; Miklavcic, S.J. Detection and analysis of wheat spikes using Convolutional Neural Networks. *Plant Methods* **2018**, *14*, 100. [CrossRef]

30. Condori, R.H.M.; Romualdo, L.M.; Bruno, O.M.; de Cerqueira Luz, P.H. Comparison Between Traditional Texture Methods and Deep Learning Descriptors for Detection of Nitrogen Deficiency in Maize Crops. In Proceedings of the 2017 Workshop of Computer Vision (WVC), Natal, Brazil, 30 October–1 November 2017; pp. 7–12.

31. Moghimi, A.; Yang, C.; Anderson, J.A. Aerial hyperspectral imagery and deep neural networks for high-throughput yield phenotyping in wheat. *Comput. Electron. Agric.* **2020**, *172*, 105299. [CrossRef]

32. Huang, P.; Luo, X.; Jin, J.; Wang, L.; Zhang, L.; Liu, J.; Zhang, Z. Improving High-Throughput Phenotyping Using Fusion of Close-Range Hyperspectral Camera and Low-Cost Depth Sensor. *Sensors* **2018**, *18*, 2711. [CrossRef] [PubMed]

33. Zhou, K.; Deng, X.; Yao, X.; Tian, Y.; Cao, W.; Zhu, Y.; Ustin, S.L.; Cheng, T. Assessing the Spectral Properties of Sunlit and Shaded Components in Rice Canopies with Near-Ground Imaging Spectroscopy Data. *Sensors* **2017**, *17*, 578. [CrossRef] [PubMed]

34. Yao, X.; Ren, H.; Cao, Z.; Tian, Y.; Cao, W.; Zhu, Y.; Cheng, T. Detecting leaf nitrogen content in wheat with canopy hyperspectral under different soil backgrounds. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *32*, 114–124. [CrossRef]

35. Hansen, P.; Schjoerring, J.K. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote. Sens. Environ.* **2003**, *86*, 542–553. [CrossRef]

36. Chen, P.; Haboudane, D.; Tremblay, N.; Wang, J.; Vigneault, P.; Li, B. New spectral indicator assessing the efficiency of crop nitrogen treatment in corn and wheat. *Remote. Sens. Environ.* **2010**, *114*, 1987–1997. [CrossRef]

37. Adams, M.L.; Philpot, W.D.; Norvell, W.A. Yellowness index: An application of spectral second derivatives to estimate chlorosis of leaves in stressed vegetation. *Int. J. Remote. Sens.* **1999**, *20*, 3663–3675. [CrossRef]

38. Serrano, L.; Penuelas, J.; Ustin, S. Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: De-composing biochemical from structural signals. *Remote Sens. Environ.* **2002**, *81*, 355–364. [CrossRef]

39. Gitelson, A.A.; Kaufman, Y.J.; Stark, R.; Rundquist, D. Novel algorithms for remote estimation of vegetation fraction. *Remote. Sens. Environ.* **2002**, *80*, 76–87. [CrossRef]

40. Huete, A. A soil-adjusted vegetation index (SAVI). *Remote. Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]

41. Fitzgerald, G.; Rodriguez, D.; Christensen, L.K.; Belford, R.; Sadras, V.O.; Clarke, T.R. Spectral and thermal sensing for nitrogen and water status in rainfed and irrigated wheat environments. *Precis. Agric.* **2006**, *7*, 233–248. [CrossRef]

42. Richardson, A.; Wiegand, C. Distinguishing vegetation from soil background information. *Photogramm. Eng. Remote Sens.* **1977**, *43*, 1541–1552.

43. Liu, H.; Huete, A. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 457–465. [CrossRef]

44. Rouse, J.W., Jr.; Haas, R.; Schell, J.; Deering, D. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ.* **1974**, *351*, 309.

45. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [CrossRef]

46. Rondeaux, G.; Steven, M.; Baret, F. Optimization of soil-adjusted vegetation indices. *Remote. Sens. Environ.* **1996**, *55*, 95–107. [CrossRef]

47. Schuerger, A.C.; Capelle, G.A.; Di Benedetto, J.A.; Mao, C.; Thai, C.N.; Evans, M.D.; Richards, J.T.; A Blank, T.; Stryjewski, E.C. Comparison of two hyperspectral imaging and two laser-induced fluorescence instruments for the detection of zinc stress and chlorophyll concentration in bahia grass (Paspalum notatum Flugge.). *Remote. Sens. Environ.* **2003**, *84*, 572–588. [CrossRef]

48. Tang, S.; Zhu, Q.; Wang, J.; Zhou, Y.; Zhao, F. Theoretical bases and application of three gradient difference vegetation index. *Sci. China Ser. D* **2003**, *33*, 1094–1102.

49. Broge, N.; Leblanc, E. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for esti-mation of green leaf area index and canopy chlorophyll density. *Remote Sens. Environ.* **2001**, *76*, 156–172. [CrossRef]

50. Haboudane, D.; Miller, J.; Pattey, E.; Zarco-Tejada, P.; Strachan, I. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* **2004**, *90*, 337–352. [CrossRef]

51. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote. Sens. Environ.* **1996**, *58*, 289–298. [CrossRef]

52. Chen, J.M. Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. *Can. J. Remote. Sens.* **1996**, *22*, 229–242. [CrossRef]

53. Kaufman, Y.J.; Tanre, D. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Trans. Geosci. Remote. Sens.* **1992**, *30*, 261–270. [CrossRef]

54. Vogelmann, J.E.; Rock, B.N.; Moss, D.M. Red edge spectral measurements from sugar maple leaves. *Int. J. Remote Sens.* **1993**, *14*, 1563–1575. [CrossRef]

55. Gamon, J.; Peñuelas, J.; Field, C. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote. Sens. Environ.* **1992**, *41*, 35–44. [CrossRef]

56. Gamon, J.A.; Serrano, L.; Surfus, J.S. The photochemical reflectance index: An optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. *Oecologia* **1997**, *112*, 492–501. [CrossRef]

57. Peñuelas, J.; Gamon, J.; Fredeen, A.; Merino, J.; Field, C. Reflectance indices associated with physiological changes in nitro-gen-and water-limited sunflower leaves. *Remote Sens. Environ.* **1994**, *48*, 135–146. [CrossRef]

58. Penuelas, J.; Baret, F.; Filella, I. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica* **1995**, *31*, 221–230.

59. Merzlyak, M.N.; Gitelson, A.; Chivkunova, O.B.; Rakitin, V.Y. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiol. Plant.* **1999**, *106*, 135–141. [CrossRef]

60. Strachan, I.; Pattey, E.; Boisvert, J. Impact of nitrogen and environmental conditions on corn as detected by hyperspectral re-flectance. *Remote Sens. Environ.* **2002**, *80*, 213–224. [CrossRef]

61. Fan, L.; Zhao, J.; Xu, X.; Liang, D.; Yang, G.; Feng, H.; Yang, H.; Wang, Y.; Chen, G.; Wei, P. Hyperspectral-based Estimation of Leaf Nitrogen Content in Corn Using Optimal Selection of Multiple Spectral Variables. *Sensors* **2019**, *19*, 2898. [CrossRef]

62. Fu, Y.; Yang, G.; Wang, J.; Song, X.; Feng, H. Winter wheat biomass estimation based on spectral indices, band depth analysis and partial least squares regression using hyperspectral measurements. *Comput. Electron. Agric.* **2014**, *100*, 51–59. [CrossRef]

63. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM.* **2017**, *6*, 84–90. [CrossRef]

64. Rigatti, S.J. Random Forest. *J. Insur. Med.* **2017**, *47*, 31–39. [CrossRef] [PubMed]

65. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]

66. Höskuldsson, A. PLS regression methods. *J. Chemom.* **1988**, *2*, 211–228. [CrossRef]

67. Lu, C.-J.; Lee, T.-S.; Chiu, C.-C. Financial time series forecasting using independent component analysis and support vector regression. *Decis. Support Syst.* **2009**, *47*, 115–125. [CrossRef]

68. Joharestani, M.Z.; Cao, C.; Ni, X.; Bashir, B.; Talebiesfandarani, S. PM$_{2.5}$ Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data. *Atmosphere* **2019**, *10*, 373. [CrossRef]

69. Yang, B.; Qi, L.; Wang, M.; Hussain, S.; Wang, H.; Wang, B.; Ning, J. Cross-Category Tea Polyphenols Evaluation Model Based on Feature Fusion of Electronic Nose and Hyperspectral Imagery. *Sensors* **2020**, *20*, 50. [CrossRef]

70. Davide, C.; Glenn, F.; Raffaele, C.; Bruno, B. Assessing the robustness of vegetation indices to estimate wheat N in Mediterranean environments. *Remote Sens.* **2014**, *6*, 2827–2844.

71. Iglesias-Puzas, Á.; Boixeda, P. Deep Learning and Mathematical Models in Dermatology. *Actas Dermo-Sifiliogr.* **2020**, *111*, 192–195. [CrossRef]

72. Cheng, G.; Li, Z.; Han, J.W.; Yao, X.W.; Guo, L. Exploring hierarchical convolutional features for hyperspectral image classifica-tion. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6712–6722. [CrossRef]

73. Mirzaei, A.; Pourahmadi, V.; Soltani, M.; Sheikhzadeh, H. Deep feature selection using a teacher-student net-work. *Neurocomputing* **2020**, *383*, 396–408. [CrossRef]

74. Jia, M.; Li, W.; Wang, K.; Zhou, C.; Cheng, T.; Tian, Y.; Zhu, Y.; Cao, W.; Yao, X. A newly developed method to extract the optimal hyperspectral feature for monitoring leaf biomass in wheat. *Comput. Electron. Agric.* **2019**, *165*, 104942. [CrossRef]