

Article

Oilseed Rape Yield Prediction from UAVs Using Vegetation Index and Machine Learning: A Case Study in East China

Hao Hu ^{1,2}, Yun Ren ³, Hongkui Zhou ^{1,2}, Weidong Lou ^{1,2}, Pengfei Hao ⁴, Baogang Lin ⁴, Guangzhi Zhang ⁵, Qing Gu ^{1,2} and Shuijin Hua ^{4,*} 

- ¹ Institute of Digital Agriculture, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China; huh@zaas.ac.cn (H.H.); zhouhongkui@zaas.ac.cn (H.Z.); louwd@zaas.ac.cn (W.L.); guq@zaas.ac.cn (Q.G.)
- ² Key Laboratory of Information Traceability for Agricultural Products, Ministry of Agriculture of China, Hangzhou 310021, China
- ³ Huzhou Agricultural Science and Technology Development Center, Huzhou 313000, China; yunhuaren@163.com
- ⁴ Institute of Crops and Nuclear Technology Utilization, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China; 11816004@zju.edu.cn (P.H.); linbg@mail.zaas.ac.cn (B.L.)
- ⁵ Zhejiang Institute of Hydraulics and Estuary, Hangzhou 310020, China; guangzhi0125@163.com
- * Correspondence: huasj@mail.zaas.ac.cn

Abstract: Yield prediction is an important agriculture management for crop policy making. In recent years, unmanned aerial vehicles (UAVs) and spectral sensor technology have been widely used in crop production. This study aims to evaluate the ability of UAVs equipped with spectral sensors to predict oilseed rape yield. In an experiment, RGB and hyperspectral images were captured using a UAV at the seedling (S1), budding (S2), flowering (S3), and pod (S4) stages in oilseed rape plants. Canopy reflectance and spectral indices of oilseed rape were extracted and calculated from the hyperspectral images. After correlation analysis and principal component analysis (PCA), input spectral indices were screened to build yield prediction models using random forest regression (RF), multiple linear regression (MLR), and support vector machine regression (SVM). The results showed that UAVs equipped with spectral sensors have great potential in predicting crop yield at a large scale. Machine learning approaches such as RF can improve the accuracy of yield models in comparison with traditional methods (e.g., MLR). The RF-based training model had the highest determination coefficient (R^2) (0.925) and lowest relative root mean square error (RRMSE) (5.91%). In testing, the MLR-based model had the highest R^2 (0.732) and lowest RRMSE (11.26%). Moreover, we found that S2 was the best stage for predicting oilseed rape yield compared with the other growth stages. This study demonstrates a relatively accurate prediction for crop yield and provides valuable insight for field crop management.

Keywords: oilseed rape; UAV; yield; vegetation index; machine learning



Citation: Hu, H.; Ren, Y.; Zhou, H.; Lou, W.; Hao, P.; Lin, B.; Zhang, G.; Gu, Q.; Hua, S. Oilseed Rape Yield Prediction from UAVs Using Vegetation Index and Machine Learning: A Case Study in East China. *Agriculture* **2024**, *14*, 1317. <https://doi.org/10.3390/agriculture14081317>

Academic Editors: Josef Eitzinger and Juan Ignacio Arribas

Received: 22 May 2024

Revised: 26 July 2024

Accepted: 6 August 2024

Published: 8 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Reliable prediction of crop yield is important for farmers to determine agricultural management. An accurate crop yield prediction model can help farmers decide on what to plant, when to grow, and when to harvest. Different approaches have been developed to predict the yields of various crops [1–5]. Many of these studies have used traditional methods to estimate crop yield by collecting soil and climate parameters [6]. However, these methods are time-consuming and do not fully describe the spatial variability of yield [7]. In contrast to traditional methods, remote sensing techniques are relatively economical and can provide quick and convenient solutions to predicting crop yields. Recently, vegetation indices (VIs) calculated from spectral reflectance in remote sensing have been developed to predict crop yields; their use requires less labor and avoids crop destruction [8,9].

Generally, near-infrared reflectance (NIR) and the visible spectrum are considered to be related to crop yield. VIs are calculated mainly by a combination of the above reflectance in two or more different bands. The reliability and potential of many indices have been investigated in various studies to estimate the crop yield of different crops. Several VIs, such as water index (WI), soil-adjusted vegetation index (SAVI), green index (GI), and green vegetation index (GVI), have been proposed [10–12]. The normalized difference vegetation index (NDVI) is widely used to estimate crop yield due to the significant correlation between crop growth conditions and productivity [13].

Usually, a single index alone may not provide excellent prediction all the time and everywhere. Combining multiple indices may lead to better predictive performance. However, a large number of VIs may also make data analysis and model building complex and difficult [14]. Therefore, many researchers have investigated the optimal combination of VIs for the most accurate yield estimation. In 2022, Kayad et al. [15] used normalized difference red-edge (NDRE) and NDVI to predict maize yield, and the NDRE- and NDVI-based model's determination coefficient (R^2) was 0.85 and 0.92, respectively. VIs may contain a large amount of information, and they usually have a nonlinear relationship with crop yield. Feature filtering, such as correlation analysis, importance ranking, and principal component analysis (PCA), can be applied to help us select model input variables.

Machine learning has big advantages in processing large amounts of nonlinear data and is capable of analyzing spectral big data. Many machine learning algorithms, such as artificial neural networks (ANNs) [16], support vector machines (SVM) [17], random forests (RFs) [18], and AdaBoost [19], have successfully predicted the yields of various crops. They each have different advantages. Rajković et al. [20] found that ANNs have a high ability to model nonlinear complex relationships between dependent and independent variables. Meanwhile, support vector machines (SVMs) have the potential to solve the overfitting problem when using high-dimensional data such as spectral data [14]. Spectral indices have been used in machine learning methods for the yield estimation of rice [21], potatoes [22,23], wheat [4,24], and cotton [25]. Marques Ramos et al. [18] employed five different machine learning methods with vegetation indices to predict maize yield, and they found that RF predicted maize yield more accurately than other models.

UAVs can be equipped with many portable devices and have been widely used in various agricultural tasks. A UAV can accomplish seeding, fertilization, spraying, and crop monitoring for a low cost and taking little time [26]. UAVs mounted with spectral sensors can collect crop canopy spectral reflectance and then obtain vegetation indices through spectral analysis. Crop yields have been predicted using such VIs by many researchers [12,27]. However, there are limited studies on predicting oilseed rape yield at different growth stages using spectral VIs obtained by UAVs and machine learning methods.

Considering all the above, the aims of the present work are (1) to investigate the effect of the growth stage on oilseed rape yield model performance, (2) to compare the oilseed rape yield model performances of four different machine learning methods and reveal the best solution, and (3) to investigate the effects of different input strategies on oilseed rape model performance. Our study improves the relationships between oilseed rape yield and predictive VI variables obtained by UAV proximal sensing.

2. Materials and Methods

2.1. Study Area and Experimental Design

The test site was located at the experiment field base of the Zhejiang Academy of Agricultural Sciences, Liangzhu town, Hangzhou city, Zhejiang province, China (Figure 1). It has a monsoon climate and is on the southern edge of the northern subtropical zone. Its winter and summer are long, while spring and autumn are short, warm, and humid; the region experiences abundant sunlight and rainfall. The annual average temperature ranges from 15.3 °C to 16.2 °C, and the annual average rainfall ranges from 1150 to

1550 mm (<https://upimg.baikex.com/doc/5366320-5602031.html>. In Chinese, accessed on 1 July 2024).

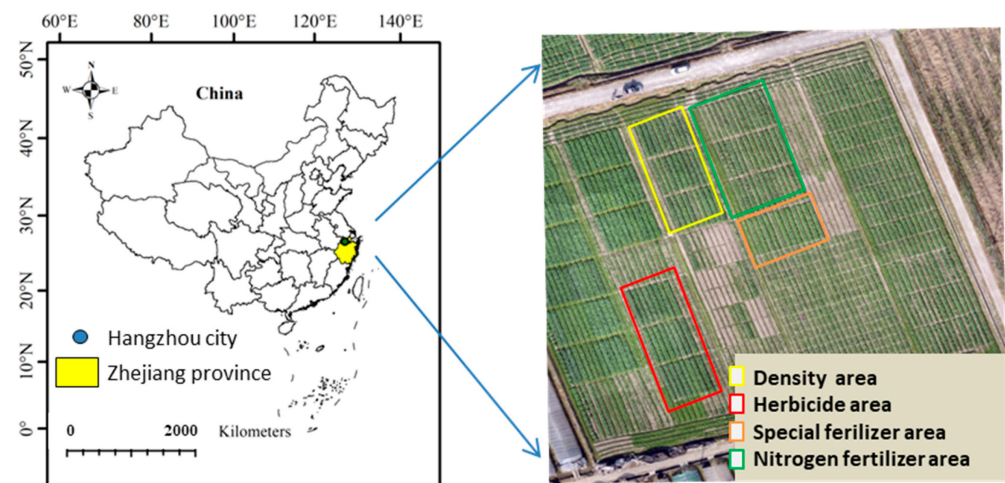


Figure 1. Location of study area and experiment field.

Details on the experimental design are shown in Table 1. In order to obtain gradient yield data, we investigated nitrogen fertilizer application (F1, F2, and F3), herbicide (NHB and HB), planting density (D1, D2, and D3), and oilseed rape special fertilizer application (CF and SF). The field experiment area could be seen in Figure 1. The study (the above four experiment treatments) involved a total of 84 field plots. Each field plot was 10.0 m long and 2.0 m wide. Oilseed rape seedlings with three leaves were transplanted on October 1 in 2021 at a depth of 10–15 cm. Except where otherwise stated, oilseed rape was planted with a density of 4.5×10^5 seedlings/ha, compound fertilizer (N:P₂O₅:K₂O = 16:16:16) at 750 kg/ha was applied, and pre-emergence herbicide (S)-Metolachlor (1.5 L/ha) was sprayed. Oilseed rape was manually harvested and dry-weighted plot by plot on 20 May 2022. Yield was expressed as kg/ha.

Table 1. Experimental design details.

Notation	Treatment	Note
F1	N fertilizer 125 kg/ha	Repeated 12 times
F2	N fertilizer 500 kg/ha	
F3	N fertilizer 750 kg/ha	
NHB	Herbicide (S)-Metolachlor sprayed with 1.5 L/ha	Repeated 12 times
HB	Without herbicide (S)-Metolachlor sprayed	
D1	Density of 1.5×10^5 seedlings/ha	Repeated 6 times
D2	Density of 4.5×10^5 seedlings/ha	
D3	Density of 7.5×10^5 seedlings/ha	
CF	Compound fertilizer (N:P ₂ O ₅ :K ₂ O = 16:16:16)	Repeated 3 times
SF	Special fertilizer for oilseed rape plant	

2.2. Data Acquisition

We chose two UAVs as platforms to obtain the RGB and hyperspectral image collection: a DJ Phantom-4-RTK and a DJ M600 Pro (DJI Technology Co., Shenzhen, China). An aerial survey was conducted under conditions of a low wind speed and clear skies from 11:00 AM to 2:00 PM local time on February 4 (seedling, (S1), March 11 (budding, S2), April 3 (flowering, S3), and April 25 (pod, S4), 2022. The mission was planned at a flight height of 60 m and with a flight speed of 6 m/s, with a front overlap of 80% and a side overlap of 70%.

A visible and near-infrared hyperspectral imaging system, the Pika XC2 camera (Resnon Inc., Bozeman, MT, USA), with a spectral range of 400–1100 nm, a spectral resolu-

tion of 1.3 nm, 447 spectral channels, and 1600 spatial channels with a 12-bit depth, was mounted and fixed on the DJ Matrice 600 Pro (DJI, Shenzhen, China) to capture the spectral images of oilseed rape plants. The maximum payload of the UAV was 6.0 kg and that of the Pika XC2 camera was 2.51 kg. Captured images were stored in memory cards. Reflectance of the oilseed rape was extracted using the software Spectronon Pro 1.0 (Resonon Inc., Bozeman, MT, USA), and spectral VIs were calculated as shown in Table 2.

Table 2. Summary of the spectral reflectance indices used in this study.

Index	Equation	Reference
NDVI	$(R_{800} - R_{670}) / (R_{800} + R_{670})$	[28]
NDVY	$(R_{560} - R_{450}) / (R_{560} + R_{450})$	[29]
CI _{re}	R_{800} / R_{720-1}	[30]
CI _{green}	R_{800} / R_{550-1}	[30]
RVI	R_{800} / R_{670}	[31]
TVI	$0.5 \times (120 \times (R_{750} - R_{550}) - 200 \times (R_{670} - R_{550}))$	[32]
BNDVI	$(R_{860} - R_{450}) / (R_{860} + R_{450})$	[33]
DVI	$R_{800} - R_{670}$	[34]
RDVI	$[NDVI \times (R_{800} + R_{670})^2]^0.5$	[35]
NIR	R_{860}	/

R_x refers to the leaf reflectance at a wavelength *x* in nanometers; NDVI is normalized difference vegetation index; NDVY is the normalized yellow difference vegetation index; CI_{re} is the red-edge chlorophyll index; CI_{green} is the green chlorophyll index; RVI is the ratio vegetation index; TVI is triangular vegetation index; BNDVI is the blue normalized difference vegetation index; DVI is the difference vegetation index; RDVI is the ratio difference vegetation index; NIR is the near-infrared reflectance.

2.3. Data Reduction

Hyperspectral images containing huge amounts of data may make data analysis complex and difficult. Thus, it is necessary to reduce the data and select the appropriate spectral VIs to construct the yield prediction model. We applied correlation analysis and PCA to reduce the dimension and amount of large spectral data. A two-tailed test with $p < 0.05$ (*) was used in the correlation analysis between reflectance, VIs, and oilseed rape yield. After correlation analysis, PCA with two principal components was conducted. The resulting scores allowed us to screen VIs with a close relationship with oilseed rape yield; these VIs were then used as model input variables.

2.4. Data Modeling

For modeling, two-thirds of the data (56 data points) were used for training, while the remaining third (28 data points) were used independently for testing. To build the oilseed rape yield prediction model, we used four machine learning regression methods: RF, multiple linear regression (MLR), and SVM. RF is a commonly used machine learning algorithm. It can combine the output of multiple decision trees to reach a single result. In the RF, the number of decision trees was set to 500, and other parameters were set to the algorithm's default values. For MLR, "input" is the model construction method. SVM is commonly used for regression analysis due to its powerful supervised learning algorithms. The employed SVM used default parameters.

2.5. Model Performance

The flowchart of the entire process, starting from image collection to data analysis and modeling, is shown in Figure 2. Before modeling, data of all the input variables were normalized using Equation (1):

$$x_n = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Here, x_i is an original value, x_n is a normalized value, and $x_{\max}$ and $x_{\min}$ are the maximum and minimum values, respectively.

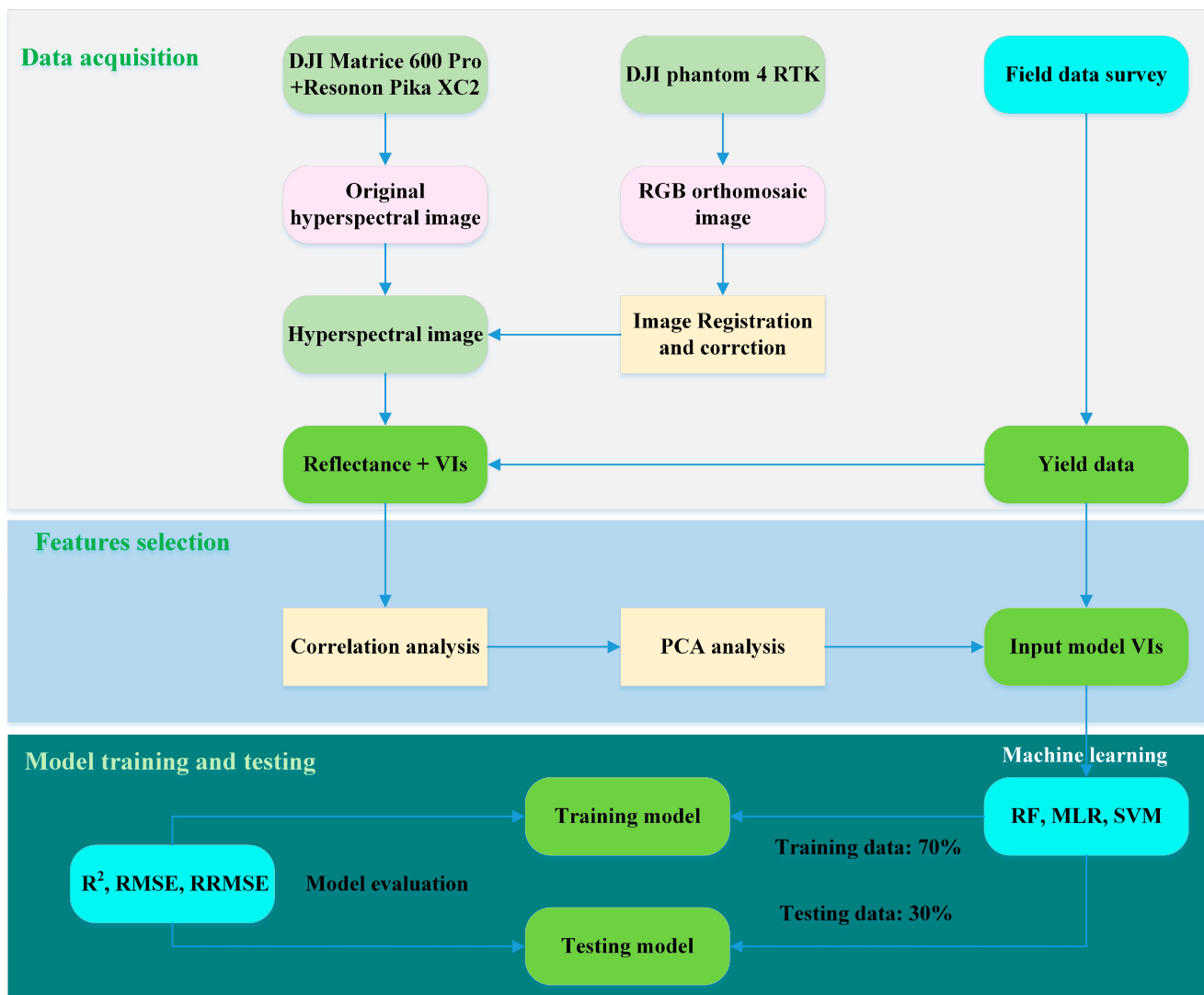


Figure 2. Development workflow for the yield prediction model.

The R^2 , root mean square error (RMSE), and relative root mean squared error (RRMSE) were used to evaluate training and testing performance (Equations (2)–(4)):

$$R^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

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

$$RRMSE = \sqrt{\frac{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2}{\sum_{i=1}^N (y_i)^2}} \times 100\% \quad (4)$$

where x and y are the measured and predicted values, respectively; and $\bar{x}$ and $\bar{y}$ are the average values of measured and predicted yield, respectively. N is the number of observations.

We conducted a Pearson correlation analysis and PCA using IBM SPSS Statistics 25.0 (SPSS Inc., New York, NY, USA, 2017). The correlation, scatter, fitted line, and PCA plot

were drawn by Origin 2022 (Origin Lab Corporation, Northampton, MA, USA, 2022). The machine learning algorithms were run in MATLAB R2022a (The MathWorks, Inc., Natick, MA, USA, 2022).

3. Results

3.1. Oilseed Rape Yield

The yield of oilseed rape plants following various treatments is shown in Figure 3. It can be seen that oilseed rape yield increased from F1 to F3, and the yield with F3 was the highest (3302.5 kg/ha). Regarding herbicide treatments, the yield with HB was 2930.04 kg/ha; this was higher than that seen with NHB treatment. The yield for the density of 4.5×10^5 seedlings/ha (D2) was the highest among the density treatments. The average yield across all treatments was 2757.2 kg/ha; the minimum and maximum yields were 1071.96 and 4152.08 kg/ha, respectively.

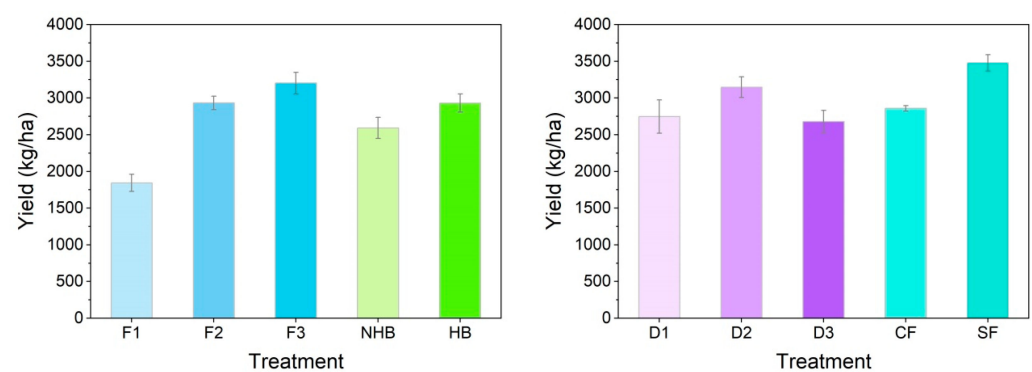


Figure 3. Oilseed rape yield with N fertilizer (F1, F2, and F3), herbicide (NHB and HB), density (D1, D2, and D3), and special fertilizer (CF and SF) treatments.

3.2. Reflectance of Oilseed Rape Plants

The reflectance of oilseed rape plants at various growth stages is shown in Figure 4. There are obvious reflectance differences in visible and near-infrared wavelength across the S1, S2, S3, and S4 stages (Figure 4A). We note that the reflectance in the NIR wavelength was the highest in the S2 stage and the lowest in the S4 stage. Hence, reflectance in the NIR wavelength may be used to differentiate the growth stages of oilseed rape.

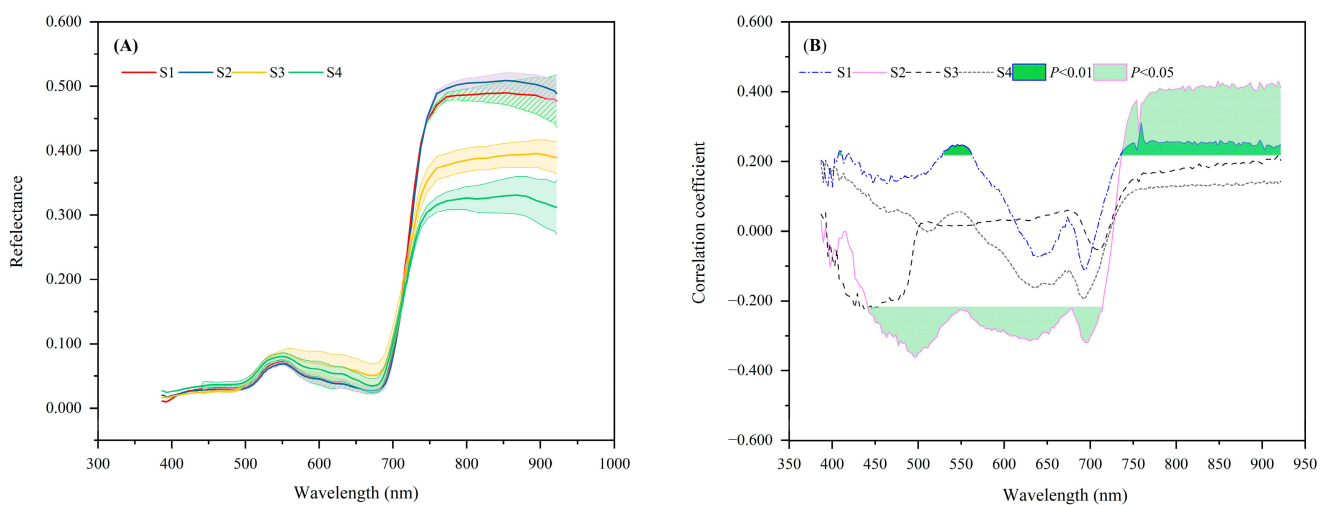


Figure 4. Reflectance of various growth stages in oilseed rape plants (A) and correlation between reflectance and yield at various growth stages (B).

The relation between reflectance and oilseed rape yield is shown in Figure 4B. The reflectance from 442 to 713 nm at the S2 stage and from 737 to 921 nm at the S1 stage had a significant relationship ($p < 0.05$) with oilseed rape yield. Meanwhile, reflectance from 529 to 561 nm in the S1 stage and from 529 to 561 nm in the S2 stage had a greatly significant relationship ($p < 0.01$) with oilseed rape yield. This indicates that S2 has a greater effect on oilseed rape in comparison with the S1, S3, and S4 stages. Compared with the visible region, the NIR region was more sensitive, and NIR reflectance had a closer relationship with the yield. Therefore, R860 in NIR region was selected as another input variable for yield modeling.

3.3. VIs of Oilseed Rape Plants

We performed a correlation analysis to evaluate the relationships between VIs and oilseed rape yield. This analysis may be helpful to ultimately determine the variables that should be used as the inputs in yield prediction models. The results are shown in Figure 5. It can be seen that all VIs at S1; all VIs except NDVY at S2; Clre and RDVI at S3; and NDVI, Clre, and RVI at S4 had significant relationships with oilseed rape yield. NDVI and BNDVI at S2 had the highest correlation coefficients: 0.896 and 0.856, respectively. S1 and S2 had more VIs with a closer relationship to oilseed rape yield than other growth stages.

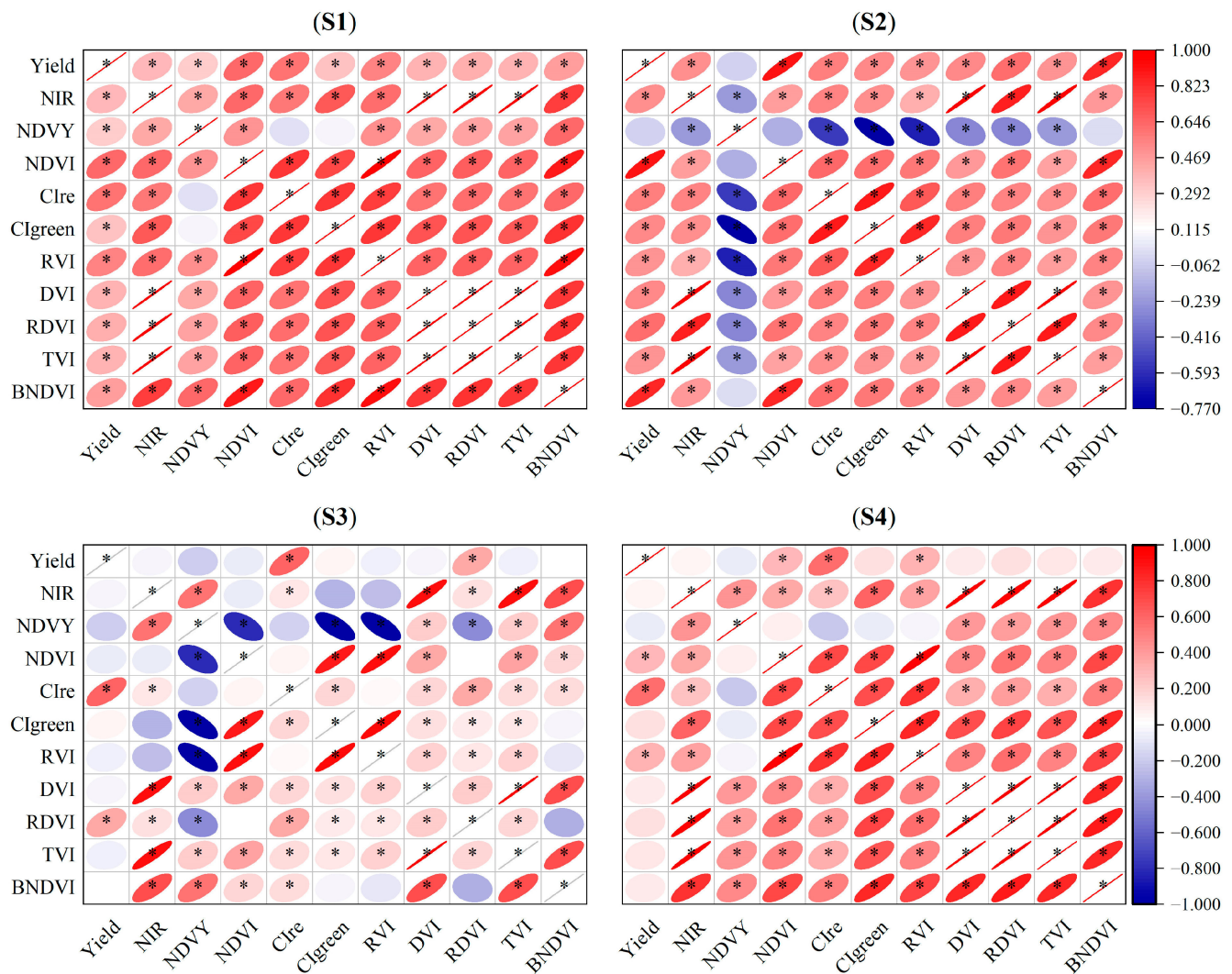


Figure 5. Relationship between VIs and oilseed rape yield at (S1), (S2), (S3), and (S4) growth stages. * indicates a significant correlation at $p < 0.05$.

3.4. PCA of VIs

As seen from Figure 5, a large number of VIs had a significant relationship with oilseed rape yield. VIs also had significant correlations among themselves (Figure 5). The use of highly correlated variables as inputs in prediction models may cause multicollinearity problems. Meanwhile, too many data make the analysis difficult. To overcome these problems, we can reduce the number of VIs. PCA of VIs which had a very significant relationship with yield was performed. The PCA result is shown in Figure 6. VIs which were in the same quadrant with yield and the first two most closely related to yield were selected and considered as the oilseed rape yield model input variables. So, we can see that NDVI and Clre at S1, RVI and Clgreen at S2, Clre at S3, and Clre at S4 were ultimately selected.

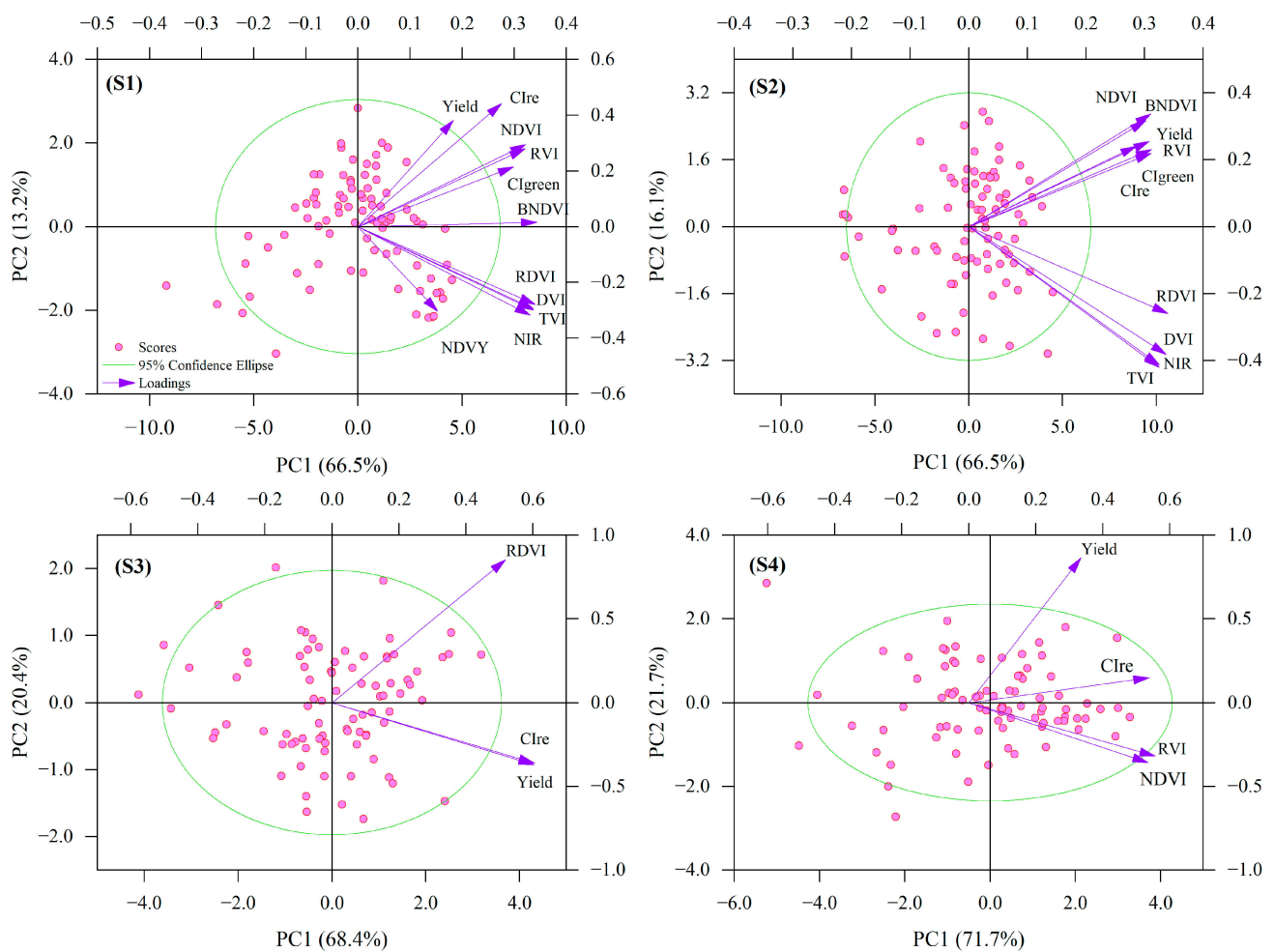


Figure 6. PCA plot of VIs and oilseed rape yield at (S1), (S2), (S3), and (S4) stages.

3.5. Performance of Oilseed Rape Yield Prediction Model

The aforementioned selected VIs were used as the input variables to build oilseed rape yield models based on the RF, MLR, and SVM. For one stage, more than one model may be established because there may be more than one input variable and combinations. Hence, the models with the best R^2 were selected (see Figures 7–9 and Table 3). During model training, the RF model had the highest R^2 (0.925) and lowest RMSE and RRMSE (167.8 kg/ha and 5.91%, respectively) across growth stages. For testing, the MLR model had the highest R^2 (0.732) and lowest RMSE and RRMSE (309.5 kg/ha and 11.26%, respectively). The RF models performed second and third best. Regarding the various oilseed rape

growth stages, S2 had the highest R^2 and lowest RMSE and RRMSE across both model training and testing.

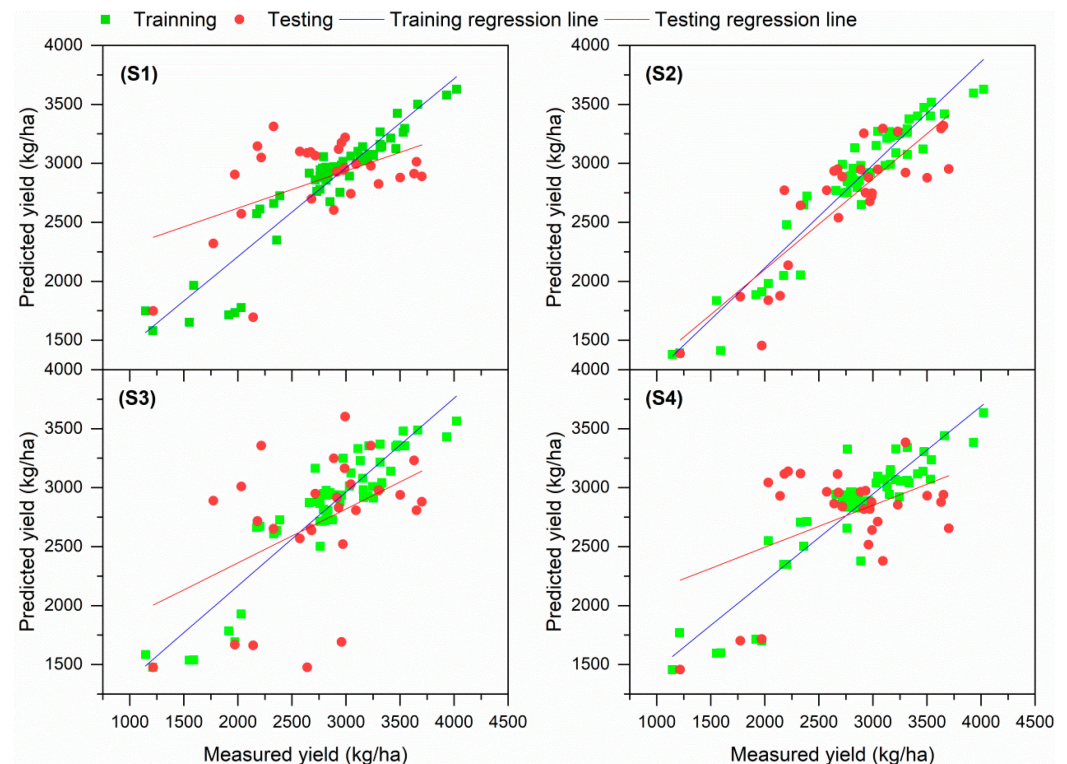


Figure 7. Measured and predicted oilseed rape yield obtained by the RF model at (S1), (S2), (S3), and (S4) stages.

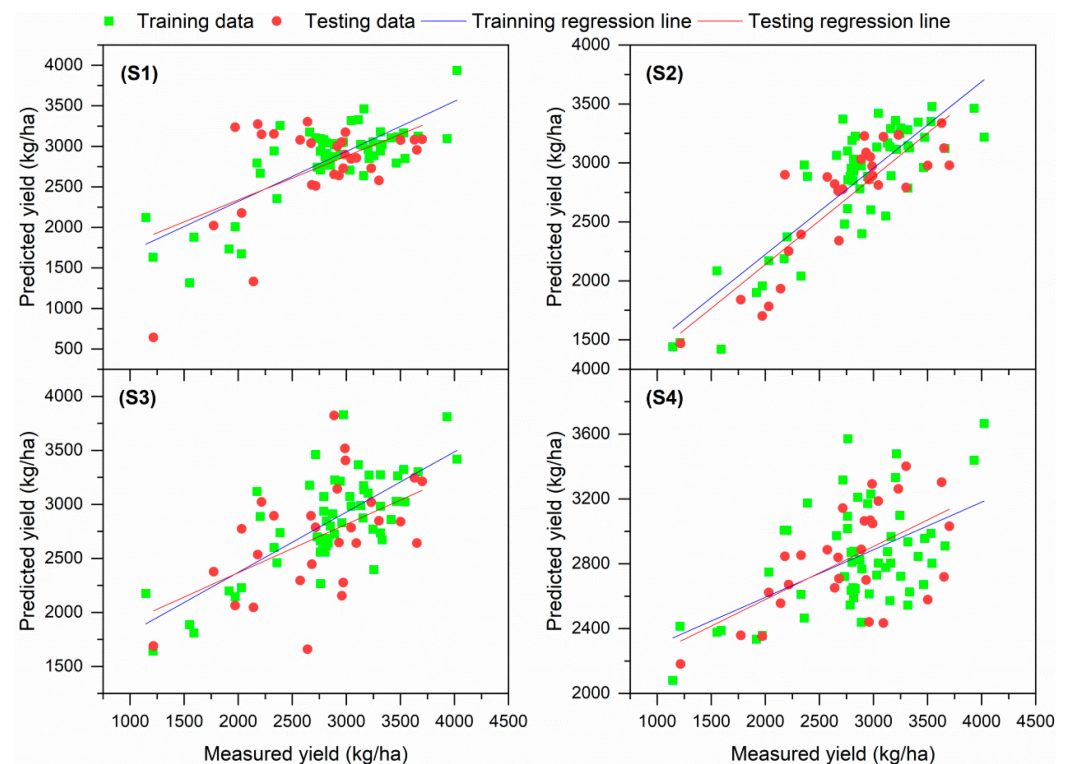


Figure 8. Measured and predicted oilseed rape yield obtained by the MLR model at (S1), (S2), (S3), and (S4) stages.

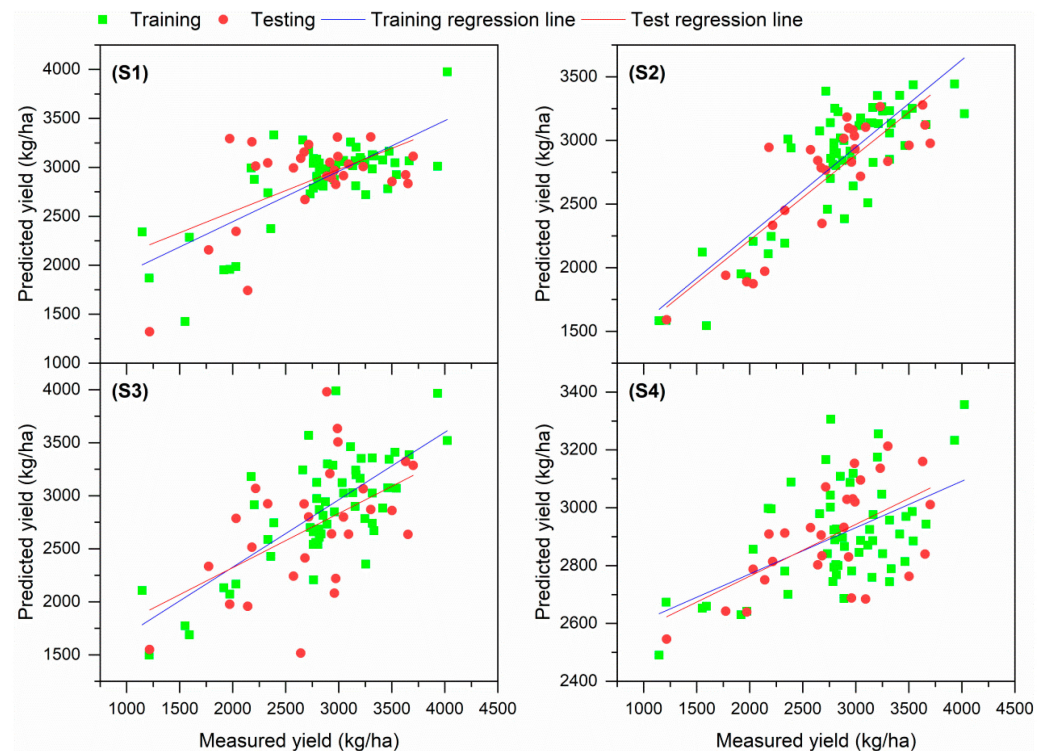


Figure 9. Measured and predicted oilseed rape yield obtained by the SVM model at (S1), (S2), (S3), and (S4) stages.

Table 3. Performance of models utilizing various machine learning methods.

Stage	Model	Training Performance			Testing Performance		
		RMSE (kg/ha)	RRMSE (%)	R ²	RMSE (kg/ha)	RRMSE (%)	R ²
S1	RF	214.9	7.57	0.893	531.3	19.34	0.243
	MLR	370.7	13.05	0.618	559.6	20.37	0.297
	SVM	398.2	14.02	0.568	518.5	18.87	0.314
S2	RF	167.8	5.91	0.925	319.7	11.64	0.723
	MLR	309.7	10.90	0.733	309.5	11.26	0.732
	SVM	316.9	11.15	0.722	318.2	11.58	0.714
S3	RF	227.7	8.02	0.861	601.3	21.88	0.220
	MLR	398.6	14.03	0.558	546.6	19.89	0.268
	SVM	404.0	14.22	0.558	577.0	21.00	0.268
S4	RF	247.4	8.71	0.842	543.0	19.76	0.227
	MLR	503.8	17.74	0.293	477.0	17.36	0.367
	SVM	529.0	18.62	0.293	525.4	19.12	0.367

3.6. Comparisons of Predicted and Measured Yield

We compared the yield predicted through machine learning and the measured values, and the results are shown in Figure 10 and Table 4. The violin plots show that the median value of the predicted yield was near the measured value across all four stages. Except for MLR and SVM for S3, the maximum yield predicted across stages was smaller than the measured yield. Moreover, except for MLR for S1, all minimum yield values predicted through machine learning were higher than the measured values across all stages. In addition, except for MLR for S1, the yield range of values predicted through machine learning were smaller than the manually measured treatment. Finally, the skewness and kurtosis (i.e., absolute value) of values predicted through machine learning were higher than those of the measured values.

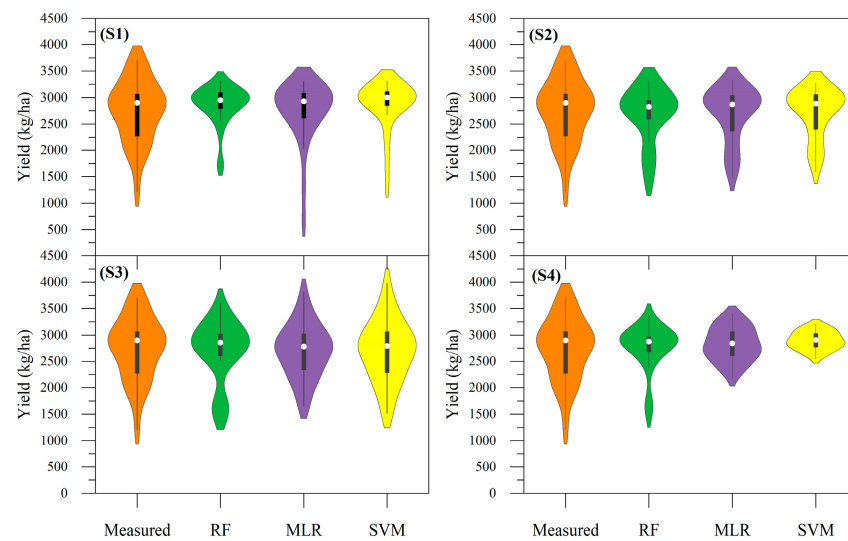


Figure 10. Measured and predicted yields at (S1), (S2), (S3), and (S4) stages.

Table 4. Data characteristics of the predicted and measured yield.

Stage		Median (kg/ha)	Average (kg/ha)	Max (kg/ha)	Min (kg/ha)	Range	Skewness	Kurtosis
S1	Measured	2901.3	2747.7	3701.9	1215.0	2486.8	−0.53	0.20
	RF	2950.9	2856.9	3313.8	1695.9	1617.9	−2.00	4.11
	MLR	2928.4	2744.0	3304.4	643.6	2660.8	−2.23	5.59
	SVM	3001.7	2871.5	3310.4	1322.2	1988.2	−2.12	4.67
S2	RF	2825.2	2676.1	3318.4	1388.7	1929.8	−2.00	4.11
	MLR	2870.2	2695.4	3338.1	1470.1	1868.0	−2.23	5.59
	SVM	2885.1	2718.4	3279.0	1591.5	1687.5	−2.12	4.67
S3	RF	2856.4	2704.8	3603.4	1477.4	2125.9	−2.00	4.11
	MLR	2779.7	2703.3	3822.5	1658.9	2163.6	−2.23	5.59
	SVM	2792.8	2705.8	3980.7	1516.0	2464.8	−2.12	4.67
S4	RF	2877.9	2761.6	3384.0	1458.5	1925.5	−2.00	4.11
	MLR	2842.7	2825.1	3401.8	2182.3	1219.5	−2.23	5.59
	SVM	2907.4	2897.8	3213.3	2546.2	667.1	−2.12	4.67

4. Discussion

In recent years, UAVs, as a near-ground-based platform to collect crop data, have been widely used in agriculture management. UAV-derived imagery data are geometrically registered, and they may have multiple sensors such as those used to collect spectral, thermal, radar, and fluorescence information. Previous studies proved that crop yield can be well predicted with UAV-collected data [36,37]. In this study, a hyperspectral imagery system was mounted on a UAV, and VIs were used to predict oilseed rape yield. R^2 of model training and testing reached 0.925 and 0.768, respectively. This suggests that UAVs may have a great ability to predict crop yield on a large scale.

VIs, including NDVI, are often used for predicting crop yield. Some studies indicated that NDVI had a significant relationship with yield and that crop yield could be predicted using NDVI [13,38,39]. A similar result was seen in our study. In our experiment, the correlation coefficient between NDVI and oilseed rape yield was 0.896 at S2. However, Lukas et al. [40] reported that the R^2 of NDVI-based oilseed rape yield was very low (only up to 0.40). In our study, NDVI was not selected as a model input variable because it had a lower correlation with oilseed rape yield than RVI and Cigreen in PCA. Other VIs may be better than NDVI in oilseed rape yield prediction models. In 2022, Lukas et al. found that R^2 of BNDVI and NDYI for oilseed rape yield reached 0.95 and 0.90, respectively. In our

study, R^2 of RVI+CIgreen with an RF-based oilseed rape yield model reached 0.925. This suggests that RVI+CIgreen has great potential for predicting crop yield.

We selected RVI, CIgreen, CIre, and their combinations as model input variables. The models during training had good performance (i.e., high R^2 and low RMSE and RRMSE). However, the testing performance was not excellent. The reason for this may be that VI data in actual field conditions are often disturbed by background noise, image shadows, and other factors. Some researchers reported that VIs combined with crop height, leaf area index, and morphological and physiological data can improve the yield model accuracy [41–44]. Which VIs and what combinations with other data should be used to help build a better yield model need to be further investigated.

Generally, hyperspectral imagery with rich spatial and spectral information can comprise hundreds of VIs or more. These VIs may have auto-collinearity problems. This makes subsequent data analysis complex and difficult. Feature screening to reduce the number of data and increase data quality is necessary for building models. To accomplish this, we employed Pearson correlations and PCA, and we obtained three features (VIs) to keep in our models. Feature screening methods have also been used by Taşan et al. [27] for predicting eggplant yield. Other data dimensionality reduction methods have been employed by many authors [44–46]. Appropriate feature selection methods for predicting crop yield models require future research.

Machine learning has been widely employed in agriculture management in recent years. The RF, MLR, and SVM are the main bases for model construction. Their suitability in predicting oilseed rape yield was demonstrated in the present work. We found that, during training, the RF-based model had the highest R^2 and lowest RMSE and RRMSE across all four stages: 0.925, 167.8 kg/ha, and 5.91%, respectively. A similar result was obtained in a study estimating pasture aboveground biomass by Freitas et al. [46]. During testing, the MLR-based model had the best performance, with its R^2 , RMSE, and RRMSE at 0.732, 309.5 kg/ha, and 11.26%, respectively. The RF-based models were more successful than other models in analyzing the complex relationships between oilseed rape yield and spectral indices. We also note that the SVR model had the lowest performance among the four investigated models.

We investigated oilseed rape growth from S1 to S4 for yield prediction. The S2 stage was determined as the best growth stage to predict yield. The earlier (S1) and latter stages (S3 and S4) may be unhelpful for predicting yield. However, we note that S2 is a growth stage and not a specific time point. Obtaining the best date for predicting yield with the best accuracy needs further study.

The data collection and model application need to follow certain standards. Our data include UAV-based imagery and ground-based measurements alongside various modeling approaches. Our data may thus encounter errors from uncontrollable factors. Therefore, standardized, unified, clear, and controlled procedures must be carefully developed and integrated into an intelligent and automated system. Uncertainty, transferability, and further application of the developed models should be assessed in future work.

There are some limitations to our research. First, we hypothesized that oilseed rape yield was determined only by spectral information. Crop yield prediction can be explained in part by spectral information. This may be the reason that our testing model had relatively weak performance. Second, in our study, spectral reflectance was extracted by averaging all the pixel reflectance of the sampling area without considering each pixel in the spectral image. Calculating the reflectance pixel by pixel may be a better approach [47]. Finally, our research result was based on a one-year field experiment. Two or more years of experiments may help in building a stable and universal yield prediction model.

5. Conclusions

In recent years, studies on the prediction of crop yield at various growth stages using spectral VIs captured by UAVs and machine learning methods have been scarce. Therefore, the current research investigated the impact of growth stage on the performance of oilseed

rape yield models, compared the efficacy of four distinct machine learning techniques, and analyzed the influence of different input strategies on model performance.

The key findings of this investigation are as follows: (1) oilseed rape yield can be accurately predicted using spectral VIs obtained from UAVs and machine learning methods, achieving an excellent model performance, with an R^2 value of 0.925 and an RRMSE of 5.91%. (2) Among the four studied growth stages, the budding stage (S2) is the most optimal stage for predicting oilseed rape yield. (3) Among the four machine learning algorithms, random forest emerged as the most effective method for predicting oilseed rape yield. (4) The combination of RVI and CIgreen served as the optimal input VIs for the oilseed rape yield model.

Looking ahead to the practical application of this yield model, several critical challenges remain to be addressed. Notably, the accurate identification of specific days and time points, rather than a broad growth stage, for accurate yield prediction demands further investigation. Additionally, rigorous evaluations of the uncertainty, transferability, and wider applicability of the developed yield models are imperative. In summary, the integration of spectral sensors with UAVs presents immense potential for large-scale crop yield forecasting. Moreover, machine learning approaches, particularly RF, significantly enhance the accuracy of yield models, surpassing traditional methods such as MLR. This study contributes a relatively precise prediction model for crop yield, offering profound insights that can drive advancements in field crop management practices.

Author Contributions: Writing—original draft, validation, methodology, formal analysis, data curation, H.H.; validation, methodology, formal analysis, Y.R.; validation, formal analysis, H.Z.; data curation, W.L.; methodology, P.H.; validation, methodology, B.L. and Q.G.; formal analysis, G.Z.; data curation, writing—review and editing, supervision, software, resources, project administration, funding acquisition, conceptualization, S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by China Agriculture Research System (CARS12) and the Three Agriculture and Nine Unit Projects of Zhejiang Province, China (2023SNJF005).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be addressed to the corresponding author.

Acknowledgments: We would like to thank the editors and reviewers for their professional advice on this manuscript.

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

References

1. Clark, R.; Dahlhaus, P.; Robinson, N.; Larkins, J.-A.; Morse-McNabb, E. Matching the model to the available data to predict wheat, barley, or canola yield: A review of recently published models and data. *Agric. Sys.* **2023**, *211*, 103749. [[CrossRef](#)]
2. Habibi, L.N.; Matsui, T.; Tanaka, T.S.T. Critical evaluation of the effects of a cross-validation strategy and machine learning optimization on the prediction accuracy and transferability of a soybean yield prediction model using UAV-based remote sensing. *J. Agric. Food Res.* **2024**, *16*, 101096. [[CrossRef](#)]
3. Luo, L.; Sun, S.; Xue, J.; Gao, Z.; Zhao, J.; Yin, Y.; Gao, F.; Luan, X. Crop yield estimation based on assimilation of crop models and remote sensing data: A systematic evaluation. *Agric. Sys.* **2023**, *210*, 103711. [[CrossRef](#)]
4. Pantazi, X.E.; Moshou, D.; Alexandridis, T.; Whetton, R.L.; Mouazen, A.M. Wheat yield prediction using machine learning and advanced sensing techniques. *Comput. Electron. Agric.* **2016**, *121*, 57–65. [[CrossRef](#)]
5. Ren, S.; Chen, H.; Hou, J.; Zhao, P.; Dong, Q.G.; Feng, H. Based on historical weather data to predict summer field-scale maize yield: Assimilation of remote sensing data to WOFOST model by ensemble Kalman filter algorithm. *Comput. Electron. Agric.* **2024**, *219*, 108822. [[CrossRef](#)]
6. Cheema, M.J.M.; Bakhsh, A.; Mahmood, T.; Liaqat, M.U. Assessment of water allocations using remote sensing and GIS modeling for Indus Basin, Pakistan. In Proceedings of the PSSP Working Paper, Washington, DC, USA, 19 February 2016; pp. 1–42.
7. Rhebergen, T.; Yeates, S.J. Climate and soil-based constraints to rainfed cotton yield in the Northern Territory, Australia—A modelling approach using APSIM-OZCOT. *Eur. J. Agron.* **2023**, *151*, 126998. [[CrossRef](#)]

8. Nakano, H.; Tanaka, R.; Guan, S.; Ohdan, H. Predicting rice grain yield using normalized difference vegetation index from UAV and GreenSeeker. *Crop Environ.* **2023**, *2*, 59–65. [\[CrossRef\]](#)
9. Xiong, X.; Zhong, R.; Tian, Q.; Huang, J.; Zhu, L.; Yang, Y.; Lin, T. Daily DeepCropNet: A hierarchical deep learning approach with daily time series of vegetation indices and climatic variables for corn yield estimation. *ISPRS J. Photogramm. Remote Sens.* **2024**, *209*, 249–264. [\[CrossRef\]](#)
10. Fernandes, J.L.; Ebecken, N.F.F.; Esquerdo, J.C.D.M. Sugarcane yield prediction in Brazil using NDVI time series and neural networks ensemble. *Int. J. Remote Sens.* **2017**, *38*, 4631–4644. [\[CrossRef\]](#)
11. Sulik, J.J.; Long, D.S. Spectral considerations for modeling yield of canola. *Remote Sens. Environ.* **2016**, *184*, 161–174. [\[CrossRef\]](#)
12. Wang, Z.; Chen, J.; Zhang, J.; Fan, Y.; Cheng, Y.; Wang, B.; Wu, X.; Tan, X.; Tan, T.; Li, S.; et al. Predicting grain yield and protein content using canopy reflectance in maize grown under different water and nitrogen levels. *Field Crops Res.* **2021**, *260*, 107988. [\[CrossRef\]](#)
13. Panek, E.; Gozdowski, D. Analysis of relationship between cereal yield and NDVI for selected regions of Central Europe based on MODIS satellite data. *Remote Sens. Appl. Soc. Environ.* **2020**, *17*, 100286. [\[CrossRef\]](#)
14. Chlingaryan, A.; Sukkariéh, S.; Whelan, B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Comput. Electron. Agric.* **2018**, *151*, 61–69. [\[CrossRef\]](#)
15. Kayad, A.; Rodrigues, F.A., Jr.; Naranjo, S.; Sozzi, M.; Pirotti, F.; Marinello, F.; Schulthess, U.; Defourny, P.; Gerard, B.; Weiss, M. Radiative transfer model inversion using high-resolution hyperspectral airborne imagery—Retrieving maize LAI to access biomass and grain yield. *Field Crops Res.* **2022**, *282*, 108449. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Kaul, M.; Hill, R.L.; Walthall, C. Artificial neural networks for corn and soybean yield prediction. *Agric. Sys.* **2005**, *85*, 1–18. [\[CrossRef\]](#)
17. Kamir, E.; Waldner, F.; Hochman, Z. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS J. Photogramm. Remote Sens.* **2020**, *160*, 124–135. [\[CrossRef\]](#)
18. Marques Ramos, A.P.; Prado Osco, L.; Elis Garcia Furuya, D.; Nunes Gonçalves, W.; Cordeiro Santana, D.; Pereira Ribeiro Teodoro, L.; Antonio da Silva Junior, C.; Fernando Capristo-Silva, G.; Li, J.; Henrique Rojo Baio, F.; et al. A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Comput. Electron. Agric.* **2020**, *178*, 105791. [\[CrossRef\]](#)
19. Jeevaganesh, R.; Harish, D.; Priya, B. A machine learning-based approach for crop yield prediction and fertilizer recommendation. In Proceedings of the 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 28–30 April 2022; pp. 1330–1334.
20. Rajković, D.; Marjanović Jeromela, A.; Pezo, L.; Lončar, B.; Zanetti, F.; Monti, A.; Kondić Špika, A. Yield and quality prediction of winter rapeseed—Artificial neural network and random forest models. *Agronomy* **2021**, *12*, 58. [\[CrossRef\]](#)
21. Liu, Z.; Ju, H.; Ma, Q.; Sun, C.; Lv, Y.; Liu, K.; Wu, T.; Cheng, M. Rice yield estimation using multi-temporal remote sensing data and machine learning: A case study of Jiangsu, China. *Agriculture* **2024**, *14*, 14040638. [\[CrossRef\]](#)
22. Gómez, D.; Salvador, P.; Sanz, J.; Casanova, J.L. Potato yield prediction using machine learning techniques and sentinel 2 Data. *Remote Sens.* **2019**, *11*, 1745. [\[CrossRef\]](#)
23. Sun, C.; Feng, L.; Zhang, Z.; Ma, Y.; Crosby, T.; Naber, M.; Wang, Y. Prediction of end-of-season tuber yield and tuber set in potatoes using in-season UAV-based hyperspectral imagery and machine learning. *Sensors* **2020**, *20*, 5293. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Johnson, M.D.; Hsieh, W.W.; Cannon, A.J.; Davidson, A.; Bédard, F. Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agric. For. Meteorol.* **2016**, *218–219*, 74–84. [\[CrossRef\]](#)
25. Prasad, N.R.; Patel, N.R.; Danodia, A. Crop yield prediction in cotton for regional level using random forest approach. *Spat. Inf. Res.* **2020**, *29*, 195–206. [\[CrossRef\]](#)
26. Matese, A.; Toscano, P.; Di Gennaro, S.; Genesio, L.; Vaccari, F.; Primicerio, J.; Belli, C.; Zaldei, A.; Bianconi, R.; Gioli, B. Intercomparison of UAV, aircraft and satellite remote sensing platforms for precision viticulture. *Remote Sens.* **2015**, *7*, 2971–2990. [\[CrossRef\]](#)
27. Taşan, S.; Cemek, B.; Taşan, M.; Cantürk, A. Estimation of eggplant yield with machine learning methods using spectral vegetation indices. *Comput. Electron. Agric.* **2022**, *202*, 107367. [\[CrossRef\]](#)
28. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ETRS. In Proceedings of the Third Earth Resources Technology Satellite-1 Symposium, Washington, DC, USA, 10–14 December 1973; pp. 309–317.
29. Sulik, J.J.; Long, D.S. Spectral indices for yellow canola flowers. *Int. J. Remote Sens.* **2015**, *36*, 2751–2765. [\[CrossRef\]](#)
30. Gitelson, A.A.; Viña, A.; Ciganda, V.; Rundquist, D.C.; Arkebauer, T.J. Remote estimation of canopy chlorophyll content in crops. *Geophys. Res. Lett.* **2005**, *32*, L08403. [\[CrossRef\]](#)
31. Jordan, C.F. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* **1969**, *50*, 663–666. [\[CrossRef\]](#)
32. Broge, N.H.; Leblanc, E. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sens. Environ.* **2001**, *76*, 156–172. [\[CrossRef\]](#)
33. Hancock, D.W.; Dougherty, C.T. Relationships between blue-and red-based vegetation indices and leaf area and yield of alfalfa. *Crop Sci.* **2007**, *47*, 2547–2556. [\[CrossRef\]](#)
34. Hadjimitsis, D.G.; Papadavid, G.; Agapiou, A.; Themistocleous, K.; Hadjimitsis, M.G.; Retalis, A.; Michaelides, S.; Chrysoulakis, N.; Toullos, L.; Clayton, C.R.I. Atmospheric correction for satellite remotely sensed data intended for agricultural applications: Impact on vegetation indices. *Nat. Hazards Earth Syst. Sci.* **2010**, *10*, 89–95. [\[CrossRef\]](#)

35. Haboudane, D.; Miller, J.R.; Tremblay, N.; Zarco-Tejada, P.J.; Dextraze, L. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens. Environ.* **2002**, *81*, 416–426. [[CrossRef](#)]
36. Alabi, T.R.; Abebe, A.T.; Chigeza, G.; Fowobaje, K.R. Estimation of soybean grain yield from multispectral high-resolution UAV data with machine learning models in West Africa. *Remote Sens. Appl. Soc. Environ.* **2022**, *27*, 100782. [[CrossRef](#)]
37. Ge, H.; Ma, F.; Li, Z.; Du, C. Estimating rice yield by assimilating UAV-derived plant nitrogen concentration into the DSSAT model: Evaluation at different assimilation time windows. *Field Crops Res.* **2022**, *288*, 108705. [[CrossRef](#)]
38. Barriguiha, A.; Jardim, B.; de Castro Neto, M.; Gil, A. Using NDVI, climate data and machine learning to estimate yield in the Douro wine region. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *114*, 103069. [[CrossRef](#)]
39. Zsebő, S.; Bede, L.; Kukorelli, G.; Kulmány, I.M.; Milics, G.; Stencinger, D.; Teschner, G.; Varga, Z.; Vona, V.; Kovács, A.J. Yield Prediction Using NDVI Values from GreenSeeker and MicaSense Cameras at Different Stages of Winter Wheat Phenology. *Drones* **2024**, *8*, 88. [[CrossRef](#)]
40. Lukas, V.; Huňady, I.; Kintl, A.; Mezera, J.; Hammerschmiedt, T.; Sobotková, J.; Brtnický, M.; Elbl, J. Using UAV to Identify the optimal vegetation index for yield prediction of oil seed rape (*Brassica napus* L.) at the flowering stage. *Remote Sens.* **2022**, *14*, 4953. [[CrossRef](#)]
41. Huang, J.; Tian, L.; Liang, S.; Ma, H.; Becker-Reshef, I.; Huang, Y.; Su, W.; Zhang, X.; Zhu, D.; Wu, W. Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. *Agric. For. Meteorol.* **2015**, *204*, 106–121. [[CrossRef](#)]
42. Sadenova, M.; Beisekenov, N.; Varbanov, P.S.; Pan, T. Application of machine learning and neural networks to predict the yield of cereals, legumes, oilseeds and forage crops in Kazakhstan. *Agriculture* **2023**, *13*, 1195. [[CrossRef](#)]
43. Xie, Y.; Wang, P.; Bai, X.; Khan, J.; Zhang, S.; Li, L.; Wang, L. Assimilation of the leaf area index and vegetation temperature condition index for winter wheat yield estimation using Landsat imagery and the CERES-Wheat model. *Agric. For. Meteorol.* **2017**, *246*, 194–206. [[CrossRef](#)]
44. Zhou, H.; Yang, J.; Lou, W.; Sheng, L.; Li, D.; Hu, H. Improving grain yield prediction through fusion of multi-temporal spectral features and agronomic trait parameters derived from UAV imagery. *Front. Plant Sci.* **2023**, *14*, 1217448. [[CrossRef](#)] [[PubMed](#)]
45. Adnan, R.M.; Liang, Z.; Heddam, S.; Zounemat-Kermani, M.; Kisi, O.; Li, B. Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. *J. Hydrol.* **2020**, *586*, 124371. [[CrossRef](#)]
46. Freitas, R.G.; Pereira, F.R.S.; Dos Reis, A.A.; Magalhães, P.S.G.; Figueiredo, G.K.D.A.; do Amaral, L.R. Estimating pasture aboveground biomass under an integrated crop-livestock system based on spectral and texture measures derived from UAV images. *Comput. Electron. Agric.* **2022**, *198*, 107122. [[CrossRef](#)]
47. He, D.; Zamora, M.; Oto, A.; Karczmar, G.S.; Fan, X. Comparison of region-of-interest-averaged and pixel-averaged analysis of DCE-MRI data based on simulations and pre-clinical experiments. *Phys. Med. Biol.* **2017**, *62*, N445–N459. [[CrossRef](#)] [[PubMed](#)]

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.