agronomy

MDPI

Article

Application of UAV Multispectral Imaging to Monitor Soybean Growth with Yield Prediction through Machine Learning

Sadia Alam Shammi [1,2], Yanbo Huang [1,*], Gary Feng [1], Haile Tewolde [1], Xin Zhang [3], Johnie Jenkins [1] and Mark Shankle [4]

1 USDA-ARS Genetics and Sustainable Agriculture Research Unit, Starkville, MS 39762, USA; sadia.shammi@usda.gov (S.A.S.); gary.feng@usda.gov (G.F.); haile.tewolde@usda.gov (H.T.); johnie.jenkins@usda.gov (J.J.)
2 Oak Ridge Institute for Science and Education, Oak Ridge, TN 37830, USA
3 Department of Agricultural and Biological Engineering, Mississippi State University, Starkville, MS 39762, USA; xzhang@abe.msstate.edu
4 Pontotoc Ridge-Flatwoods Branch Experiment Station, Mississippi State University, Pontotoc, MS 38863, USA; mark.shankle@msstate.edu
* Correspondence: yanbo.huang@usda.gov

Abstract: The application of remote sensing, which is non-destructive and cost-efficient, has been widely used in crop monitoring and management. This study used a built-in multispectral imager on a small unmanned aerial vehicle (UAV) to capture multispectral images in five different spectral bands (blue, green, red, red edge, and near-infrared), instead of satellite-captured data, to monitor soybean growth in a field. The field experiment was conducted in a soybean field at the Mississippi State University Experiment Station near Pontotoc, MS, USA. The experiment consisted of five cover crops (Cereal Rye, Vetch, Wheat, Mustard plus Cereal Rye, and native vegetation) planted in the winter and three fertilizer treatments (Fertilizer, Poultry Liter, and None) applied before planting the soybean. During the soybean growing season in 2022, eight UAV imaging flyovers were conducted, spread across the growth season. UAV image-derived vegetation indices (VIs) coupled with machine learning (ML) models were computed for characterizing soybean growth at different stages across the season. The aim of this study focuses on monitoring soybean growth to predict yield, using 14 VIs including CC (Canopy Cover), NDVI (Normalized Difference Vegetation Index), GNDVI (Green Normalized Difference Vegetation Index), EVI2 (Enhanced Vegetation Index 2), and others. Different machine learning algorithms including Linear Regression (LR), Support Vector Machine (SVM), and Random Forest (RF) are used for this purpose. The stage of the initial pod development was shown as having the best predictability for earliest soybean yield prediction. CC, NDVI, and NAVI (Normalized area vegetation index) were shown as the best VIs for yield prediction. The RMSE was found to be about 134.5 to 511.11 kg ha^{-1} in the different yield models, whereas it was 605.26 to 685.96 kg ha^{-1} in the cross-validated models. Due to the limited number of training and testing samples in the K-fold cross-validation, the models' results changed to some extent. Nevertheless, the results of this study will be useful for the application of UAV remote sensing to provide information for soybean production and management. This study demonstrates that VIs coupled with ML models can be used in multistage soybean yield prediction at a farm scale, even with a limited number of training samples.

Keywords: crop yield prediction; UAV; multispectral imaging; machine learning

check for updates

Citation: Shammi, S.A.; Huang, Y.; Feng, G.; Tewolde, H.; Zhang, X.; Jenkins, J.; Shankle, M. Application of UAV Multispectral Imaging to Monitor Soybean Growth with Yield Prediction through Machine Learning. *Agronomy* **2024**, *14*, 672. https://doi.org/10.3390/agronomy14040672

Academic Editor: Gniewko Niedbała

Received: 28 February 2024
Revised: 17 March 2024
Accepted: 25 March 2024
Published: 26 March 2024

1. Introduction

The relationship between agriculture and food security is intertwined. Ongoing advancements in technology and tools are streamlining agricultural operations, facilitating a daily increase in food production. Despite these developments, challenges persist in managing the field environment, obtaining current information on crops and fields, and addressing uncertainties related to weather and unforeseen natural events. In pursuit of the national objective of a sustainable agricultural production system, numerous researchers are actively working to enhance monitoring and forecasting systems for crop production.

Soybean is the fourth leading crop cultivated globally, and the most-traded agricultural commodity, at about 9 percent of the total value of agricultural trade [1]. It is the second most-cultivated crop in the United States, comprising an about 31.2% share of the total area planted (about 31.1 million hectares) and providing a net return of about 544.89 USD per hectare [2]. The national average soybean total production costs per bushel were 9.85 USD for the 2021 crop in the United States [1]. According to a 2022 report, U.S. soybean production has increased and reached 4465 million bushels (121.5 million metric tons) in 2022 [3]. However, in 2022, an estimated 4.3 billion bushels of soybeans were produced in the United States, a decrease of almost 200 million bushels compared to the previous year [4].

To reduce production costs, the adoption of new technologies can be beneficial. To provide sufficient food and fibers for increasing human demands, an increase in agricultural production is urgently needed. However, sustainable food production is always a demanding and challenging issue. With the challenges of global climate change, natural and anthropogenic pollution, erosion, and disturbances, the agriculture sector needs new, additional technologies to overcome the limitations of crop production. Remote sensing as an advanced technology is, nowadays, widely applied in precision agriculture [5–7]. It is a method of acquiring spatial information by measuring electromagnetic radiation that interacts with the atmosphere and with objects. Nowadays, multispectral and hyperspectral space-borne and airborne images are available from different sensors on different platforms like MODIS, Landsat, SPOT, Sentinel, crewed aircraft, and uncrewed unmanned aerial vehicles (UAVs), with different resolutions and wavelengths [8].

However, resolution and scale are two important factors in the application of remote sensing (RS) for precision agriculture. On the farm scale, UAVs are successfully used for corn yield prediction at different growth stages [9]. UAV-captured multispectral images are widely applied for soybean grain yield prediction [10–12]. Satellite-captured images are also applied for monitoring soybean growth stages [13] and yield prediction [1]. Other researchers are also trying to improve crop cultivation and monitoring systems using RS technology [14–19].

Crop yield prediction using the vegetation index (VI) obtained from RS data is challenging [10]. This study has been conducted to use some commonly used VIs, i.e., the CC (Canopy Cover), NDVI (Normalized Difference Vegetation Index), GNDVI (Green Normalized Difference Vegetation Index), EVI2 (Enhanced Vegetation Index 2), NDRE (Normalized Difference Red Edge Index), ARVI (Atmospherically resistant vegetation index), CCCI (Canopy Chlorophyl Content Index), GRRI (Green–Red ratio vegetation Index), CARI (Chlorophyll Absorption Ratio Index), NAVI (Normalized Area Vegetation Index), SCCCI (Simplified Canopy Chlorophyll Content Index), CIRE (Chlorophyll Index Red edge), CVI (Chlorophyll Vegetation Index), and GCVI (Green Chlorophyll Vegetation Index) to monitor crop yield at different stages of soybean growth. Monitoring soybean growth at the vegetative and reproductive stages would be helpful for improving crop yield prediction levels.

Furthermore, this study used some popular ML models, i.e., Linear Regression (LR), Random Forest (RF), and Support Vector Machine (SVM) that have been previously applied for crop yield prediction [9–11]. Different stages of soybean growth have been modelled with LR models [13]. The VIs have been used to study soybean V4 stages (25 days after emergence), using decision trees for yield prediction [10]. Besides this, the vegetative (V6) and reproductive (R5) growth stages of corn have been explored using different ML models [9]. Different phenotypes of soybean have also been examined using machine learning for yield prediction, which is helpful for crop breeding assessment [11]. Data fusion and ML models also show good predictions for soybean yield estimation [12]. This study monitored both the vegetative and reproductive stages of soybean using ML models for yield estimation. However, identifying detailed information for the different stages of soybean growing in a field is a difficult and time-consuming task. For better yield prediction, this information is necessary. VIs made from remotely sensed images would be helpful for this purpose.

The purpose of this study was to identify soybean yield under the influence of nutrient management in a field. This will provide advantages during decision making for farm management, crop economics, and market management. The objectives of this study were to identify the optimal VIs that are useful at a farm scale and the optimal stage for predicting soybean yield by the processing and analyzing of UAV-captured multispectral images.

2. Materials and Methods

2.1. Study Site

The study field was located at the Pontotoc Ridge-Flatwoods Branch Experiment Station of the Mississippi Agricultural and Forestry Experiment Station near Pontotoc, MS, USA (34.2478831° N, 88.998673° W) (Figure 1).

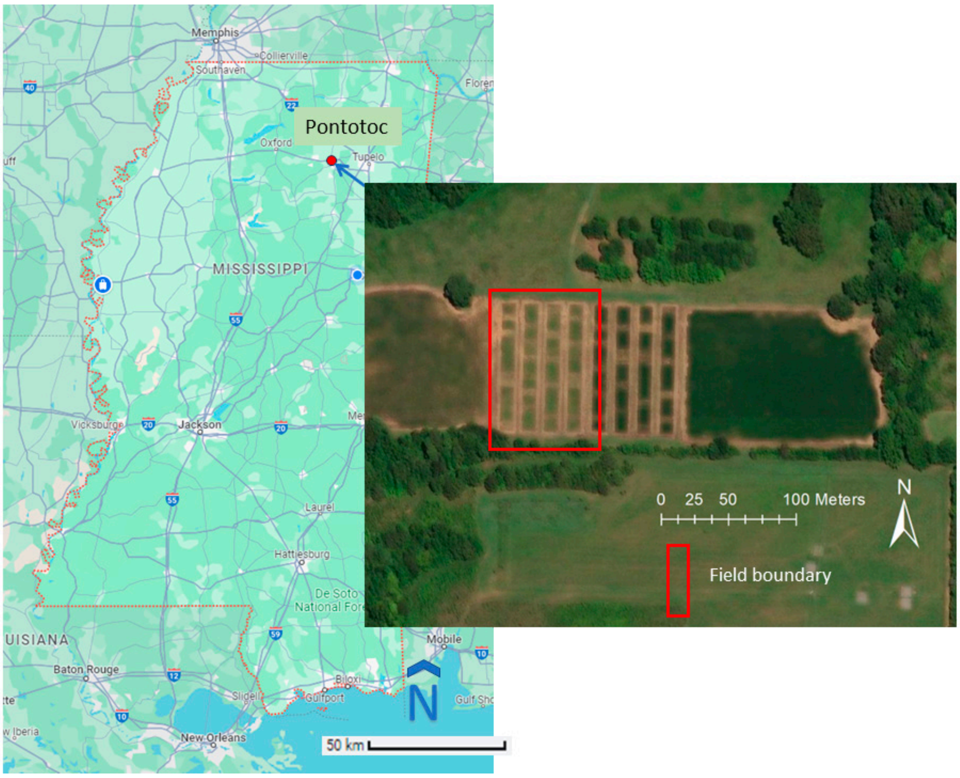


Figure 1. The field in the study site for the soybean planting experiment.

2.2. Experiment Design

The study consisted of five cover crops (Cereal Rye, Vetch, Wheat, Mustard plus Cereal Rye, and native vegetation) planted in the winter and three fertilizer treatments (Synthetic Fertilizer, Poultry Liter, and None) applied before planting the soybean in a full factorial combination. The design was a randomized complete block with four replications. The cover crops were planted in the fall of 2021 and killed about two weeks before planting soybean in the spring of 2022. The design of the experimental plots for the cover crops and fertilizer treatments are mentioned in Table A1 and the layout as shown in Figure 2. The size of each plot was 6.1 m by 9.1 m. The synthetic fertilizer treatment was recommended based on standard practice; it contained 125 kg ha^{-1} yr^{-1} (Kilogram per hectare per year) of Phosphorous (P), 45 kg ha^{-1} yr^{-1} of K (Potassium), 22.4 45 kg ha^{-1} yr^{-1} of S (Sulphur), and 4.5 of Zn (Zinc). Poultry litter (PL) was used as a substitute for synthetic fertilizers. The rate of PL was 4500 kg ha^{-1} yr^{-1}. Hence, the soybean was not irrigated. Therefore, the soybean growth stages mentioned in Table 1 were examined during the growth period. The soybean yield was measured after harvesting and compared with different stages of soybean growth in the models described in later sections.

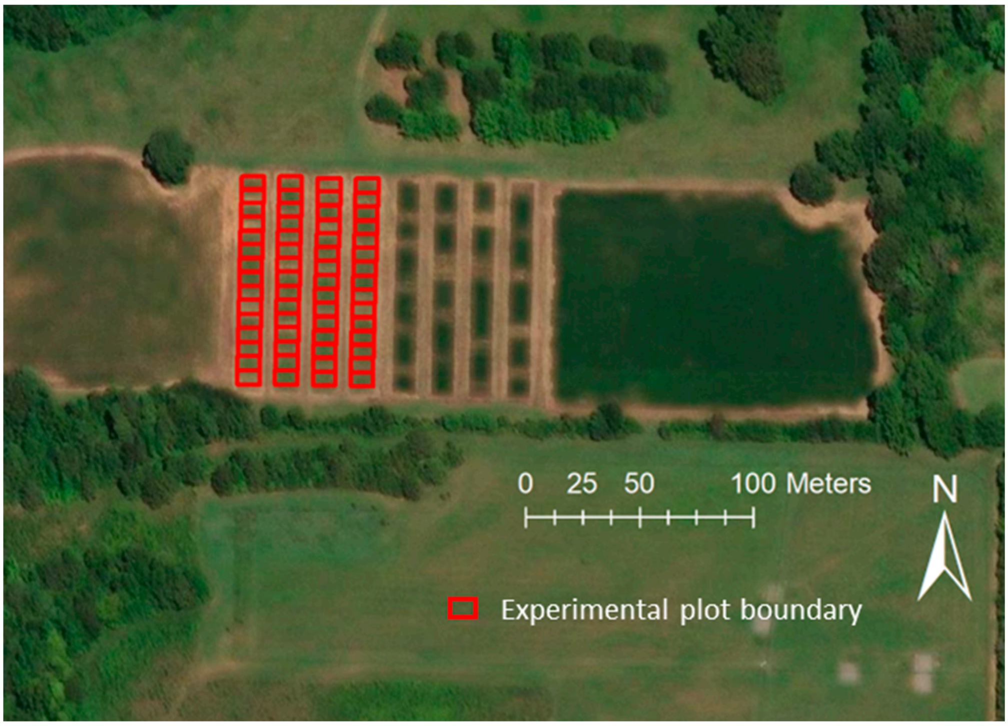


Figure 2. Field plot arrangement in the experimental design.

Table 1. Soybean growth stages with respect to UAV flyovers for field imaging.

No.	Date of UAV Imaging Flyovers	Growth Stage Label and Description
a	8 June 2022	VE: Vegetative Emergence
b	17 June 2022	V1: 1st node develops
c	7 July 2022	V7: 8th node develops
d	21 July 2022	R3: Initial Pod develops
e	3 August 2022	R4: Full Pod develops
f	16 August 2022	R5: Initial seed develops
g	31 August 2022	R6: Full seed develops
h	13 September 2022	R7: Initial Maturity

2.3. UAV Imaging

The soybean field images were acquired using a DJI Phantom 4 quadcopter UAV with a built-in multispectral camera (DJI, Shenzhen, China). On the UAV, the camera was mounted on a gimble with a −90° to +30° tilt controllable range, with six 1/2.9″ 2.08 MP CMOS sensors with a 1600 × 1300 image size and 62.7° field of view. The sensors included one broadband RGB sensor for visible light imaging and five narrowband monochrome sensors (blue (B): 450 nm ± 16 nm; green (G): 560 nm ± 16 nm; red (R): 650 nm ± 16 nm; red edge (RE): 730 nm ± 16 nm; and near-infrared (NIR): 840 nm ± 26 nm) for multispectral imaging. For image calibration from digital counts to percent reflectance, the images of a calibrated reflectance panel were captured prior to and after each flight. The camera operation was automatically synchronized for global position system (GPS) positions with the global navigation satellite system (GNSS; GPS + GLONASS + Galileo) built-in on the UAV.

According to the USDA's field crop usual planting and harvesting dates (2010), the most active season for planting is between 24 March–27 April and the harvesting period is between 23 August–23 September for Mississippi (https://downloads.usda.library.cornell.edu/usda-esmis/files/vm40xr56k/dv13zw65p/w9505297d/planting-10-29-2010.pdf, accessed on 10 January 2024). We have followed the period of late May for planting and mid-September for the harvesting of soybean. During this period, the monthly average low and high temperatures were reported as 19–30 °C for June, 21–32 °C for July, 20–32 °C for August, and 16–29 °C for September in the study area. The average precipitation during these months was 123, 110, 102, and 93 mm, respectively. The UAV flights were conducted between 10:30 a.m. and 12:00 p.m. to avoid cloud shadows, as weather permitted, with a flight altitude of 50 m above the canopy surface to acquire high-resolution (~4 cm/pixel) images along the progress of the soybean's growth. Flight routes were preset using the mission planning tool of Pix4DCapture software (Pix4D, Lausanne, Switzerland), with an image front overlap of 80% and a side overlap of 70%.

The collected images were Importe" to 'ix4DMapper (Pix4D, Lausanne, Switzerland) to generate broadband RGB orthomosaic images and narrow-band green, red, red edge, and NIR orthomosaic images, which were orthorectified to correct for geometric and vignetting distortion. Figure 3 shows the color infrared (CIR) orthomosaic image of the soybean field. Orthomosaic images were imported to ArcMap (ESRI, Redlands, CA, USA) to draw the boundary of each plot based on the different treatments. A Python (https://www.python.org/, accessed on 10 October 2023) script was written to extract the mean values of the canopy cover and spectral bands within each experimental plot.

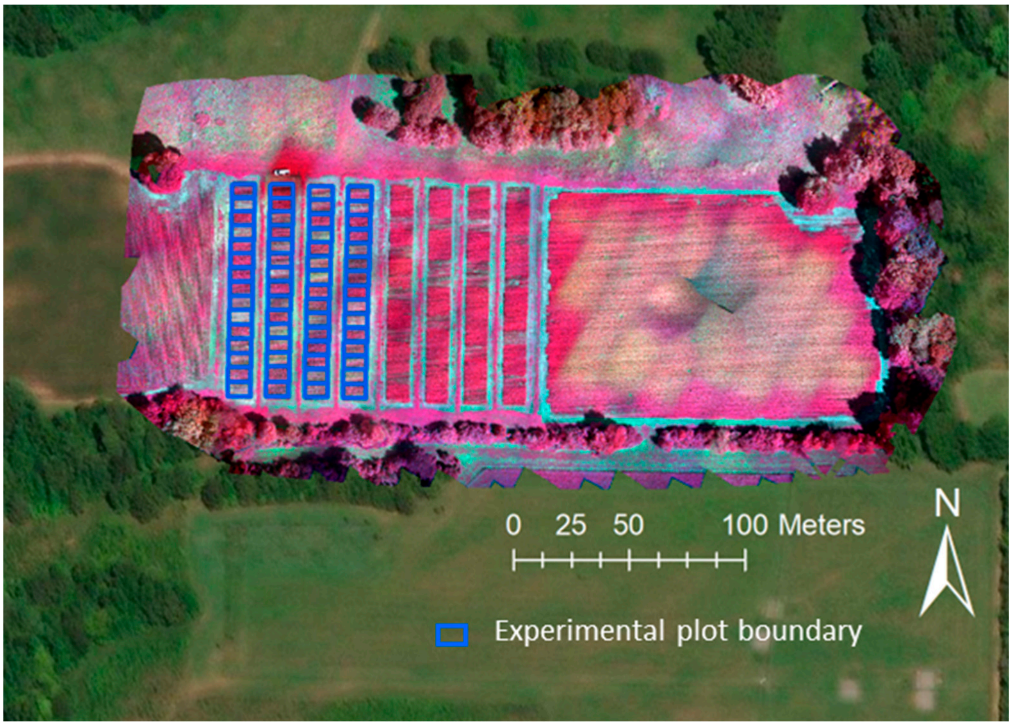
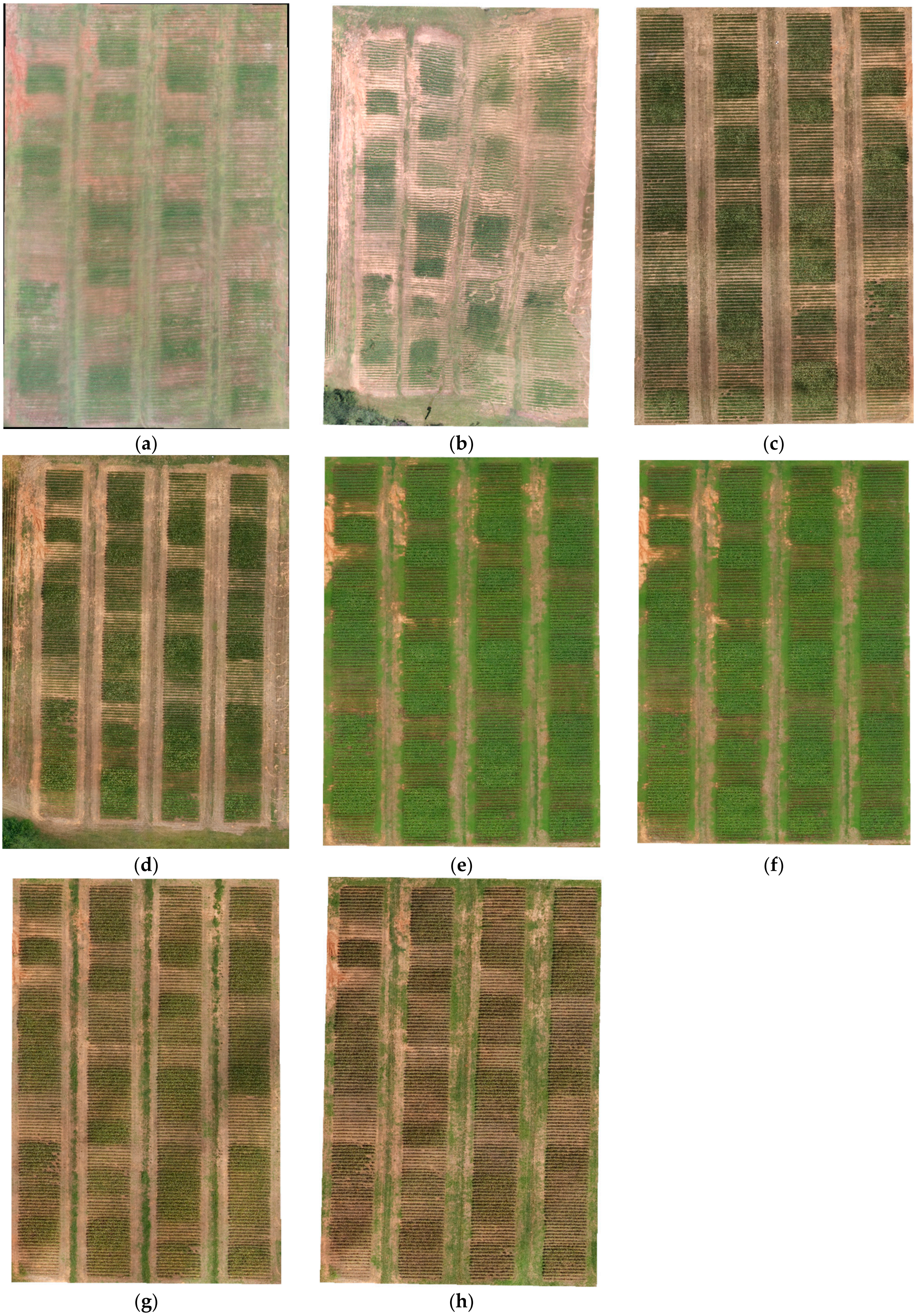


Figure 3. Field experimental plot arrangement over the ortho-mosaicked CIR image of the soybean field.

In total, UAV imaging was conducted using 8 flyovers throughout the soybean's growth stages (Table 1). The dates were selected based on weather conditions and the soybean growth stages that were being monitored in the field. Figure 4 shows RGB UAV images from each flight on changes in the condition of crops.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 4. The cropped RGB images of the soybean field plots for each flight date from (**a**–**h**) in Table 1, respectively.

2.4. Vegetation Indices

Vegetation indices (VIs), which are formulated by combining image band data, are indicators of crop greenness and health. Among the various vegetation indices, the normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI) are frequently used for crop growth and yield-related research as a remote sensing parameter [20]. Other studies have shown that the vegetation index, reflecting peak greenness, is the most active parameter in forecasting crop yield [21,22]. In our study, the acquired multispectral datasets for different stages of soybean growth were used in deriving thirteen VIs (Table 2).

Table 2. Description of different VIs (Blue, Green, Red, NIR, Red Edge are the five multispectral images used for deriving the VIs).

Vegetation Index	Description	Formula	Reference
CC	Canopy cover	No. of Vegetative Pixels/No. of Total Pixels	[23]
NDVI	Normalized difference vegetation index	$\frac{NIR-RED}{NIR+RED}$	[10]
GNDVI	Green normalized difference vegetation index	$\frac{NIR-GREEN}{NIR+GREEN}$	[10]
EVI2	Two-band enhanced vegetation index	$\frac{2.5(NIR-RED)}{NIR+2.4RED+1}$	[19]
NDRE	Normalized difference red edge index	$\frac{NIR-RED\ EDGE}{NIR+RED\ EDGE}$	[10]
ARVI	Atmospherically resistant vegetation index	$\frac{GREEN-RED}{GREEN+RED-BLUE}$	[24]
CCCI	Canopy chlorophyl content index	$\frac{NDRE-NDRE_{MIN}}{NDRE_{max}-NDRE_{MIN}}$	[25]
GRRI	Green–red ratio vegetation index	$\frac{GREEN}{RED}$	[26]
CARI	Chlorophyll absorption ratio index	$\frac{RED\ EDGE}{RED}$	[27]
NAVI	Normalized area vegetation index	$\frac{NIR-RED}{NIR}$	[28]
SCCCI	Simplified canopy chlorophyll content index	$\frac{NDVI}{NDRE}$	[24]
CIRE	Chlorophyll index red edge	$\frac{NIR}{RED\ EDGE}-1$	[29]
CVI	Chlorophyll vegetation index	$\frac{NIR\times RED}{GREEN^2}$	[28]
GCVI	Green chlorophyll vegetation index	$\frac{NIR}{GREEN}-1$	[30]

In addition, canopy cover (CC) [23] was included as an index by dividing the number of vegetative pixels (0.5 NDVI thresholding) from the total number of pixels in the unit area (each plot for this study). Table 3 provides a list of the derived VIs and their mathematical formulas used in this study.

Table 3. Correlation between VIs.

	CC	NDVI	GNDVI	EVI2	NDRE	ARVI	CCCI	GRRI	CARI	NAVI	SCCCI	CIRE	CVI	GCVI
CC	1.00													
NDVI	0.80	1.00												
GNDVI	0.66	0.79	1.00											
EVI2	0.23	0.37	−0.18	1.00										
NDRE	0.19	0.17	−0.27	0.61	1.00									
ARVI	0.20	0.49	0.10	0.69	0.12	1.00								
CCCI	0.19	0.17	−0.27	0.61	1.00	0.12	1.00							
GRRI	0.52	0.68	0.14	0.85	0.47	0.76	0.47	1.00						
CARI	0.67	0.84	0.84	0.11	−0.28	0.42	−0.28	0.50	1.00					
NAVI	0.79	0.99	0.77	0.38	0.19	0.48	0.19	0.68	0.81	1.00				
SCCCI	0.08	0.22	0.50	−0.43	−0.87	0.15	−0.87	−0.13	0.59	0.21	1.00			
CIRE	−0.14	−0.24	0.32	−0.95	−0.73	−0.60	−0.73	−0.78	0.04	−0.25	0.54	1.00		
CVI	0.18	0.16	0.68	−0.76	−0.41	−0.43	−0.41	−0.55	0.30	0.16	0.43	0.74	1.00	
GCVI	0.64	0.71	0.96	−0.29	−0.23	−0.04	−0.23	0.06	0.80	0.70	0.46	0.37	0.77	1.00

2.5. Data Modeling and Evaluation

2.5.1. ML Models

Three commonly used models i.e., linear regression (LR), support vector machine (SVM), and random forest (RF) were applied in this study to evaluate soybean yield prediction at a farm scale [9,11,13,14]. Linear regression is a simple and interpretable statistical model that describes the linear relationship between dependent and independent variables. LR makes the following assumptions: homogeneity of variance (i.e., training samples have similar variances), training samples are normally distributed and statistically independent, and there is linearity between dependent and independent variables. For yield prediction, SVM is employed as support vector regression (SVR) to enable an optimal hyperplane to be obtained and minimizes the difference between predicted and observed values. The availability of kernel functions, such as linear, polynomial, radial basis function, and sigmoid, facilitates the development of optimal hyperplanes to produce higher accuracy. RF uses a bagging technique, where many decision trees (DTs) are developed to obtain an ensemble model for accurate classification or prediction results. The implemented RF algorithm has one hyperparameter, noted as 'mtry', which describes the number of input variables randomly selected at each split while developing different DTs.

2.5.2. Model Performance Evaluation

We used the coefficient of determination (R^2) and Root Mean Square Error ($RMSE$) in this study, as these are commonly used in assessing the prediction performance of ML models. The mathematical equations are shown in Equations (1) and (2), respectively:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum(y - \hat{y})^2}{\sum(y - \overline{y})^2} \tag{1}$$

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

where $\hat{y}$ is the model predicted yield, y is the measured yield, $\overline{y}$ is the average yield, n is the total number of samples, SSE is the sum square error, and SST is the total sum square.

2.5.3. K-Fold Cross Validation

We used a K-fold cross validation scheme to evaluate the ML model's prediction to deal with the limited data generated by the experimental design. The schematic diagram for the K-fold cross validation is described in Figure 5. First, the dataset is divided into K-folds. Then, one-fold is used for testing, and the k-1 folds are used for training. A K-number of iterations are completed to find the mean error in the models. For this study, the data were split into 4 groups. One group was set as the test data and the remaining 3 groups were set as the training and validation data. As such, a 4-fold cross validation was conducted with the scheme.

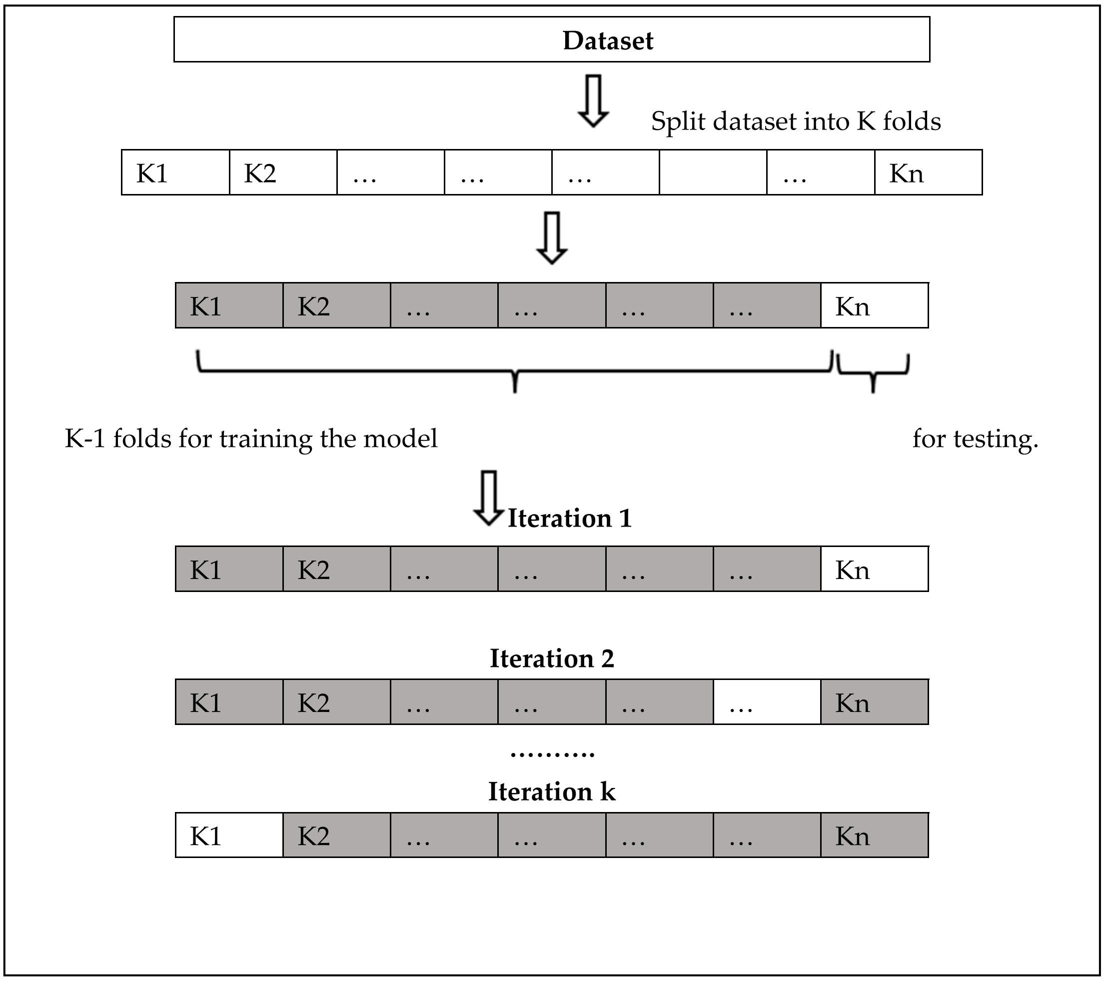


Figure 5. A schematic diagram for K-fold cross validation.

3. Results

3.1. VI Explorative Analysis

This study formed 14 Vis from the images collected by the UAV. The ranges of the Vis are plotted in Figure 6. Some Vis, i.e., CC, NDVI, GNDVI, EVI2, NDRE, ARVI, CCCI, CIRE were within the range of 0–1. Other Vis, i.e., GRRI, CARI, NAVI, SCCCI, CVI, and GCVI were within the range of 0–8, with some outliers. The correlation among the Vis is shown in Table 3. While 13 VI pairs were highly positively correlated, 6 VI pairs were negatively correlated. It is notable that we did not find any direct influence of cover crops and fertilizer treatments on the Vis.

3.2. Impact of Fertilizer Treatments and Cover Crops

This study had three fertilizer and five cover crop treatments that had an interactive impact on the soybean yield. As Figure 7 shows, there was an increase in soybean yield due to fertilizer and poultry litter treatment compared to no fertilizer treatment in the field. The range of crop yields with fertilizer treatment (fert.) was 1443.61–2421.71 kg ha^{-1}, with a mean of 1902.26 kg ha^{-1} and a standard deviation (std dev) of 292.07 kg ha^{-1}; for the poultry litter treatment (PL.) the range was 1526.73–2801.68 kg ha^{-1}, with a mean of 2136.9 kg ha^{-1} and a std dev of 289.38 kg ha^{-1}, respectively. The range of the crop yield with no treatment (None) was 529.6–1453.77 kg ha^{-1}, with a mean of 1116.37 kg ha^{-1} and a std dev of 246.81 kg ha^{-1}. However, there was no significant change in the soybean yield with different cover crops. The study found that the range of the crop yields for Cereal Rye (CR) was 529.6–2171.47 kg ha^{-1}, with a mean of 1567.29 kg ha^{-1} and a std dev of 522.41 kg ha^{-1}; for Vetch (VE) this was 724.83–2338.12 kg ha^{-1} with a mean of 1747.45 kg ha^{-1} and a std dev of 471.83 kg ha^{-1}; for Wheat (WH) the range was 993.1–2801.68 kg ha^{-1} with a mean of 1898.9 kg ha^{-1} and a std dev of 533.37 kg ha^{-1}; for Mustard plus Cereal Rye (CRm) the range was 765.12–2783.32 kg ha^{-1} with a mean of 1711.2 kg ha^{-1} and a std dev of 553.75 kg ha^{-1}; and for native vegetation (NV), the range was 885.09–2421.71 kg ha^{-1} with a mean of 1667.69 kg ha^{-1} and a std dev of 448.23 kg ha^{-1}, respectively. The dry weight of the CR, CRm, VE, and WH were measured and the average values were 1885 kg ha^{-1}, 2009 kg ha^{-1}, 1652 kg ha^{-1}, and 3242 kg ha^{-1},respectively. The NV's dry weight was not measured.

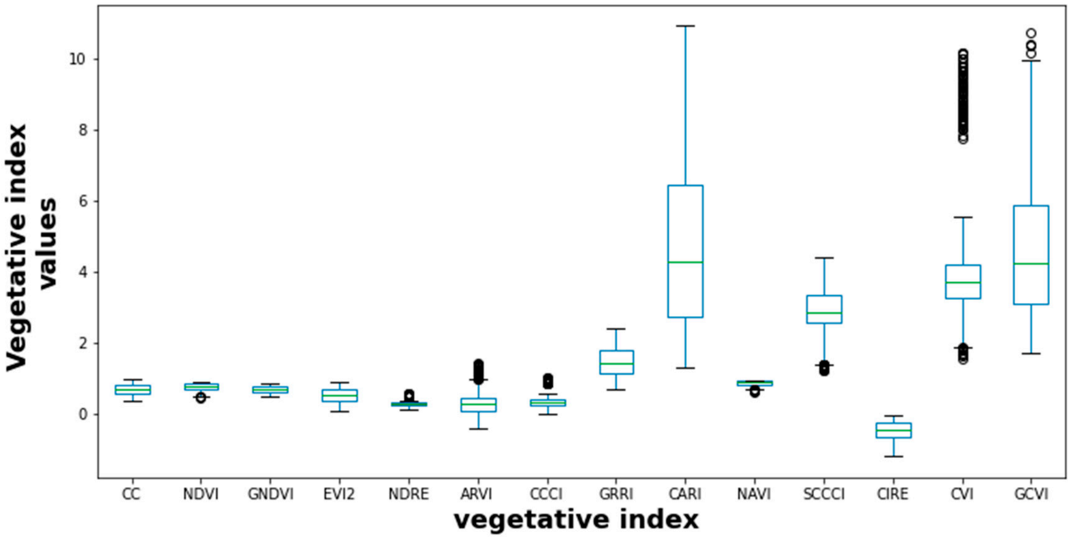


Figure 6. Ranges of Vis.

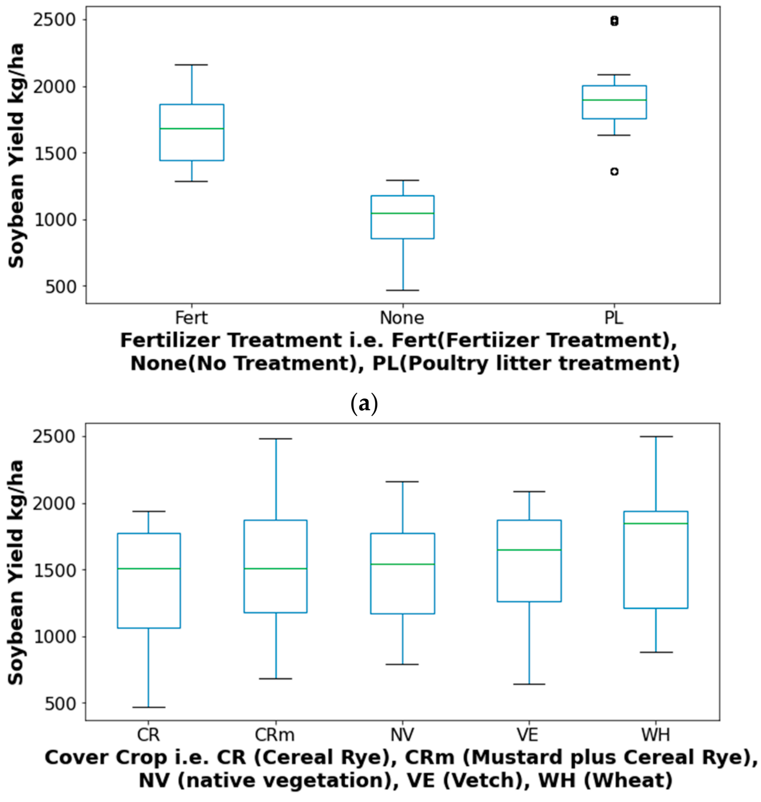


Figure 7. Soybean yield with field experiment treatments: (**a**) yield vs. fertilizers; (**b**) yield vs. cover crops.

3.3. Yield Prediction Modeling

3.3.1. Crop Yield Modeling

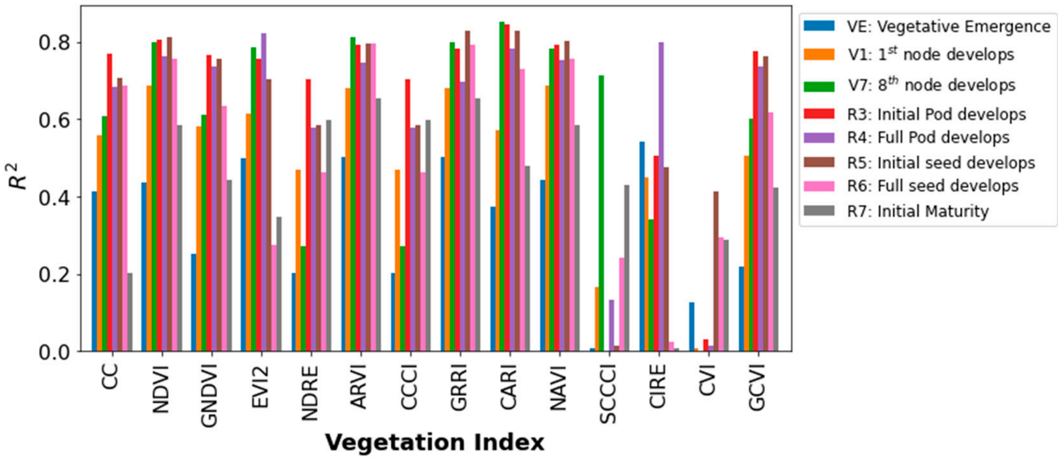
Obtained using VIs, the crop yield modeling results for the total growth season are listed in Table 4. From the LR model, CC, NDVI, and NAVI showed better predictability, with a RMSE of 404.65, 447.22, and 447.76 kg ha^{-1}, respectively. The R^2 in the LR model ranged from 0 to 38% using different VIs. A similar trend was found in the SVR model. The RMSE ranged between 409.29 and 529.33 kg ha^{-1} in the SVR model. However, the RF model showed CC, NDVI, EVI2, GRRI, and NAVI as being better VIs compared to the rest of the VIs. The R^2 values ranged between 80 and 87% and the RMSE ranged between 184.6 and 228.12 kg ha^{-1} in the RF models. The CC metrics showed the least RMSE and highest R^2 value of all the ML models. Overall, all models agreed that the CC, NDVI and NAVI were the best indices for soybean yield modelling from UAV-derived multispectral images.

Table 4. Yield prediction from three ML models.

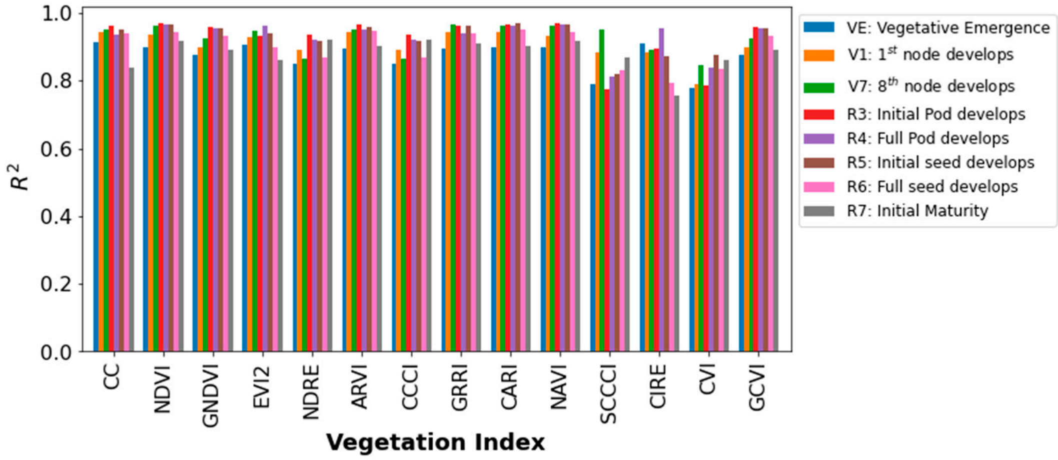
VI	LR Model		RF Model		SVR Model	
	R^2	RMSE (kg ha^{-1})	R^2	RMSE (kg ha^{-1})	R^2	RMSE (kg ha^{-1})
CC	0.386	404.65	0.872	184.60	0.372	409.29
NDVI	0.250	447.22	0.849	200.54	0.245	448.63
GNDVI	0.105	488.44	0.816	221.39	0.098	490.46
EVI2	0.046	504.45	0.863	191.33	0.059	500.75
NDRE	0.025	509.76	0.811	224.35	−0.039	526.37
ARVI	0.044	504.79	0.839	207.00	0.033	507.75
CCCI	0.025	509.76	0.811	224.35	0.040	506.00
GRRI	0.174	469.21	0.852	198.46	0.176	468.87
CARI	0.152	475.60	0.835	209.55	0.197	462.82
NAVI	0.248	447.76	0.849	200.54	0.247	447.96
SCCCI	0.000	516.35	0.824	216.88	−0.051	529.33
CIRE	0.017	511.85	0.840	206.66	−0.037	525.84
CVI	0.002	515.75	0.805	228.12	0.001	516.02
GCVI	0.089	492.75	0.816	221.39	0.117	485.35

3.3.2. Soybean Yield Modeling at Different Growth Stages

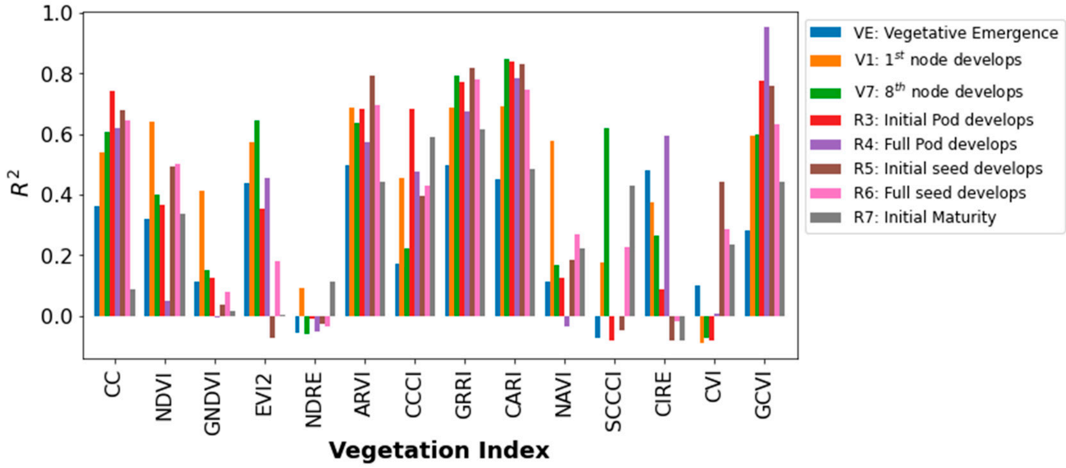
In this study, UAV imaging flyovers spanned soybean growth phenology across different stages, i.e., VE, V1, V7, R3, R4, R5, R6, and R7, as shown in Table 1. For VE, V1, V7, R3, R4, R5, R6, and R7, the crop yield modeling results (R^2, and RMSE) are plotted in Figures 8 and 9, respectively. With the LR model, the soybean growth stages of R3, R4, and R5 were indicated as good stages for yield prediction, with high R^2 and low RMSE values. All the VIs showed the same trend in these stages. However, the variation in the prediction of the RF model was not noticeable. Besides the LR model, the SVR model indicated variations in yield prediction at different growth stages. This study found that the LR and SVR models were more consistent than the RF model for the soybean yield prediction scenarios. Therefore, the R3 stage could be used for early soybean yield prediction and R5, R6, and R7 could be used for later yield prediction using any ML models.



(**a**)



(**b**)



(**c**)

Figure 8. The R^2 values for soybean yield prediction modeling at different growth stages: (**a**) for the LR model, (**b**) for the RF model, and (**c**) for the SVR model, respectively.

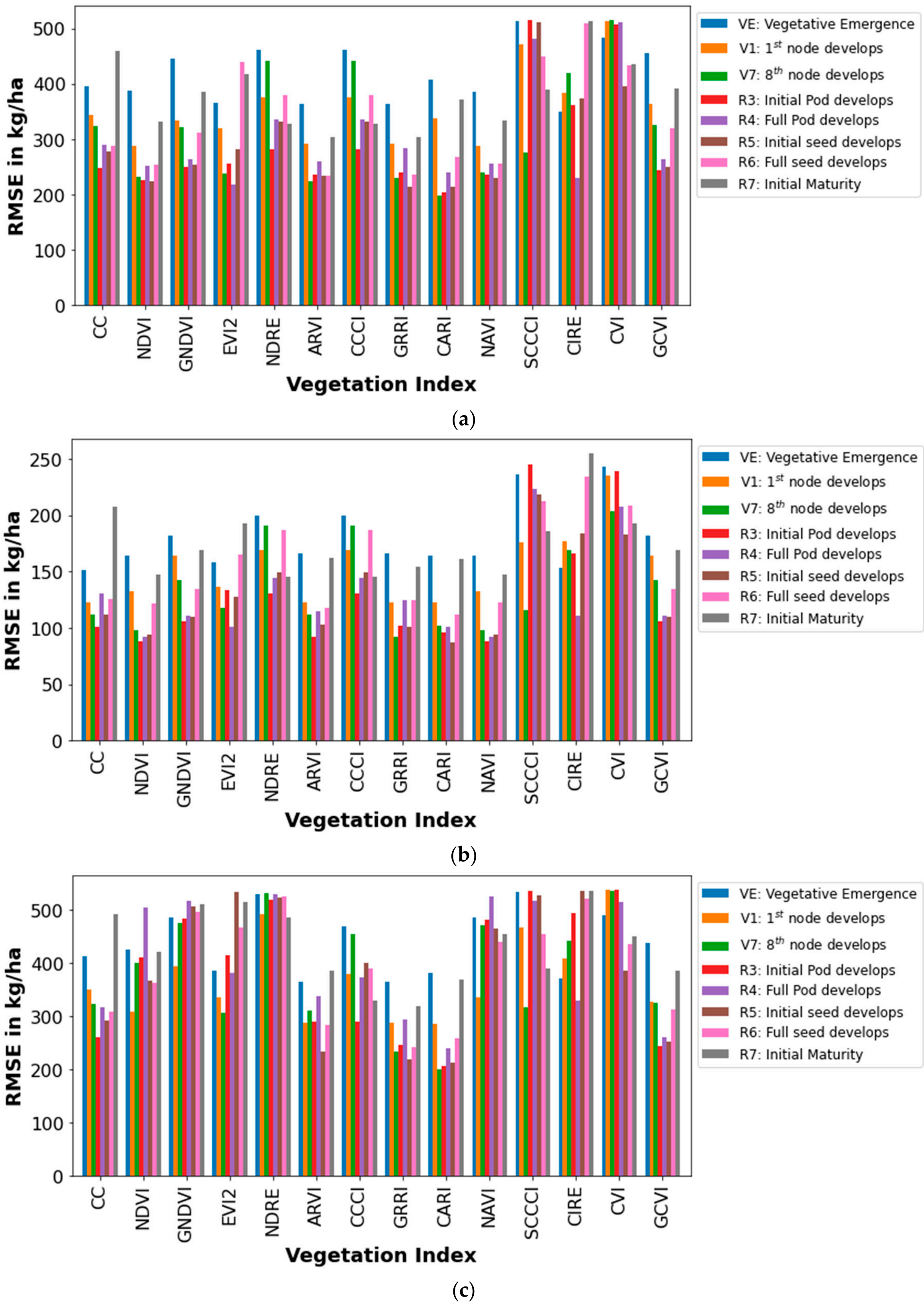


Figure 9. The RMSE values for soybean yield prediction modeling at different growth stages: (**a**) for the LR model, (**b**) for the RF model, and (**c**) for the SVR model, respectively.

3.4. Cross-Validated Yield Prediction Model

Due to the limited sample size of the dataset, we cross-validated the total season soybean yield model using a K-fold cross validation scheme. Here, we have four-fold cross validation results, as shown in Table 5. This study found a range of RMSEs of between 618.71 and 679.24 kg ha^{-1} from all the models, with varying R^2 values. Different VIs showed different values of R^2 and RMSE in different models but with similar trends. We found similar trends in the LR and SVR cross-validated models. However, the RF models showed an ability to capture data patterns, which might be due to the limited number of training and testing samples. The negative R^2 value is a reflection of this. The Lowest RMSE of 595.64 kg ha^{-1} was found in the SVR model using the CARI index. The SVR model showed less sensitivity for soybean yield prediction. However, this study simplified the VI metrics from these cross-validated ML models, which may have been useful for soybean yield modeling with limited data.

Table 5. Cross-validated yield prediction by three ML models.

VI	LR Model		RF Model		SVR Model	
	R^2	RMSE (kg ha^{-1})	R^2	RMSE (kg ha^{-1})	R^2	RMSE (kg ha^{-1})
CC	0.063	621.40	−0.117	678.56	0.082	615.14
NDVI	0.012	638.01	0.015	625.95	0.036	630.34
GNDVI	0.017	636.67	−0.155	677.89	0.023	634.58
EVI2	0.013	637.74	−0.003	642.92	0.025	633.91
NDRE	0.025	634.04	−0.060	660.95	0.024	634.18
ARVI	−0.004	643.26	−0.020	648.37	0.002	641.44
CCCI	0.025	634.04	−0.060	660.95	0.031	632.16
GRRI	0.061	622.07	−0.129	670.26	0.069	619.45
CARI	0.076	617.16	−0.115	677.95	0.109	595.64
NAVI	0.002	641.24	0.016	625.82	0.008	639.56
SCCCI	0.020	635.59	−0.138	684.96	0.056	623.62
CIRE	0.008	639.29	−0.105	675.01	0.007	628.73
CVI	0.002	641.51	−0.158	678.97	0.048	615.55
GCVI	0.053	624.63	−0.149	688.13	0.098	599.00

4. Discussion

This study developed a remote sensing method for monitoring soybean growth and predicting yields using UAV multispectral image data by using machine learning approaches. Compared to satellite sensors, UAV sensors have the advantage of providing high-resolution data with low atmospheric interference and showed good predictive abilities in this study, which will be informative for reducing crop management costs and labor. In this study, we formed and used 14 VI metrics. Of these 14 VIs, CC, NDVI, and NAVI showed good performance in predicting soybean yield at different growth stages. Previous studies have also emphasized the use of NDVI and NDVI-derived metrics for soybean yield prediction using remote sensing data [13,14]. This study found that the R3 stage, i.e., the initial pod development stage, could be the earliest stage for good yield prediction derived from the ML models. A previous study conducted on the V6 and R5 stages for corn yield prediction showed a similar performance [9]. The RMSE found in this study, which was a range of between 605.26 and 685.96 kg ha^{-1} in the cross validated models, followed the results found in previous studies [13]. Our study also found the impacts of nutrient treatments, i.e., poultry litter and fertilizer, on the soybean yield.

Among the ML models used in this study, the LR model was easiest to implement. However, this model was found to be sensitive to the outliers, so it was difficult to monitor the crop growth and yield relationship. However, the SVR model was found to be less sensitive to the crop growth and yield interaction and was good for yield prediction. However, while the RF model is usually good for identifying feature interactions, due to the small sample size it could not identify the proper interactions during cross validation.

We found a negative R-square in the cross validated RF model. Therefore, all models showed good applicability for soybean yield prediction.

This study showed a real field scenario for soybean yield prediction using UAV multispectral data. These results could be helpful in implementing UAVs in agricultural crop management and could also help in national and global crop production management in cooperation with satellite sensors for large-scale studies. The use of UAVs in the field will reduce labor and costs for seasonal crop production in fields. In crop science studies, field-scale experiments are highly desirable; however, the costs of the experimental setups prevent repetition over multiple years.

Therefore, this study maintains an actual field scenario of soybean production with nutrient management and agrochemical treatment. The results found in this study will provide a baseline for future crop studies using UAVs. In addition to this, if we could collect more UAV images, the soybean growth stages could be well monitored and might improve the prediction scenario. Additionally, considering meteorological factors in the ML model could provide more realistic predictions.

5. Conclusions

This study was conducted on soybean crops with nutrient management in a field. The use of UAVs to monitor soybean growth and predict soybean yield from VI metrics showed fruitful behaviors. The CC, NDVI, and NAVI metrics showed the best predictability at different stages of soybean growth. The earliest time for soybean yield prediction was at initial pod development, according to UAV-derived VI metrics. This study showed the LR, RF and SVR model's applicability for soybean yield prediction. However, any other ML model could fit in this study. Considering the interaction of fertilizers and cover crops, different-yield ML models produced an RMSE ranging from 134.5 to 511.11 kg ha^{-1} in training models, in contrast to the 605.26 to 685.96 kg ha^{-1} in the cross-validated models. This study will help us meet national crop management goals and will assist decision makers in crop production and management.

Author Contributions: Conceptualization, Y.H.; methodology, S.A.S. and Y.H.; software, S.A.S.; validation, S.A.S.; formal analysis, S.A.S.; investigation, S.A.S. and Y.H.; resources, Y.H. and J.J.; data curation, Y.H., H.T. and M.S.; writing—original draft preparation, S.A.S. and Y.H.; writing—review and editing, G.F, H.T., X.Z., J.J. and M.S.; visualization, S.A.S.; supervision, Y.H.; project administration, Y.H.; funding acquisition, Y.H., G.F. and J.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by a postdoctoral research fellow appointment to the Agricultural Research Service (ARS) Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the U.S. Department of Agriculture (USDA). ORISE is managed by Oak Ridge Associated Universities (ORAU) under DOE contract number DE-SC0014664. All opinions expressed in this paper are the author's and do not necessarily reflect the policies and views of USDA, DOE, or ORAU/ORISE. This research was also supported in part by the Mississippi Soybean Promotion Board, project no. 21-2022 and 62-2023.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to United States Department of Agriculture, Agricultural Research Service guidelines.

Acknowledgments: Thanks to Wyatt Young for his travels to Pontotoc, MS, for the operation of the UAV for imaging the fields, and for the processing of the collected multispectral images in Pix4D for further analyses in this study.

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

Abbreviations

ARVI	Atmospherically resistant vegetation index
B	Blue
CARI	Chlorophyll Absorption Ratio Index
CC	Canopy Cover
CCCI	Canopy Chlorophyl Content Index
CIR	Color infrared
CIRE	Chlorophyll Index Red edge
CR	Cereal Rye
CRm	Mustard plus Cereal Rye
CVI	Chlorophyll Vegetation Index
EVI2	Enhanced Vegetation Index 2
Fert	Fertilizer
G	Green
GCVI	Green Chlorophyll Vegetation Index
GNDVI	Green Normalized Difference Vegetation Index
GNSS	Global navigation satellite system
GPS	Global position system
GRRI	Green–Red ratio vegetation Index
kg ha^{-1} yr^{-1}	Kilogram per hectare per year
LR	Linear Regression
ML	Machine Learning
NAVI	Normalized Area Vegetation Index
NDRE	Normalized Difference Red Edge Index
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
NV	Native vegetation
PL	Poultry Liter
R	Red
RE	Red Edge
RF	Random Forest
RMSE	Root Mean Square Error
R2	Coefficient of Determination
RS	Remote sensing
SCCCI	Simplified Canopy Chlorophyll Content Index
SVM	Support Vector Machine
UAV	Unmanned aerial vehicle
VE	Vetch
WH	Wheat

Appendix A

Table A1. Design of field experiments for different treatments and varieties.

Cover Crop	Experimental Plot Design			
Cereal Rye	101 Fert	201 Fert	301 None	401 Fert
	102 None	202 PL	302 Fert	402 None
	103 PL	203 None	303 PL	403 PL
Vetch	104 None	204 PL	304 None	404 PL
	105 Fert	205 None	305 PL	405 None
	106 PL	206 Fert	306 Fert	406 Fert
Wheat	107 PL	207 None	307 None	407 Fert
	108 Fert	208 PL	308 PL	408 PL
	109 None	209 Fert	309 Fert	409 None

Table A1. *Cont.*

Cover Crop	Experimental Plot Design			
NRCS Mustard + Cereal Rye	110 None	210 PL	310 None	410 None
	111 PL	211 None	311 Fert	411 PL
	112 Fert	212 Fert	312 PL	412 Fert
Native Vegetation	113 Fert	213 None	313 PL	413 Fert
	114 PL	214 PL	314 None	414 None
	115 None	215 Fert	315 Fert	415 PL

References

1. Valdes, C.; Gillespie, J.; Dohlman, E. *Soybean Production, Marketing Costs, and Export Competitiveness in Brazil and the United States*; (Report No. EIB-262); U.S. Department of Agriculture, Economic Research Service: Washington, DC, USA, 2023.
2. Williams, B.; Pounds-Barnett, G. *Producer Supply Response for Area Planted of Seven Major US Crops*; (Report No ERR-327); U.S. Department of Agriculture, Economic Research Service: Washington, DC, USA, 2023. [CrossRef]
3. Ates, A.M.; Bukowski, M. *Oil Crops Outlook: August 2022*; (Report No OCS-22h); U.S. Department of Agriculture, Economic Research Service: Washington, DC, USA, 2022.
4. US Department of Agriculture, Economic Research Service. United States of America. Soybeans and Oil Crops. 2022. Available online: https://www.ers.usda.gov/topics/crops/soybeans-and-oil-crops/related-data-statistics/ (accessed on 10 November 2023).
5. Huang, Y.; Chen, Z.; Yu, T.; Huang, X.; Gu, X. Agricultural remote sensing big data: Management and applications. *J. Integr. Agric.* **2018**, *7*, 1915–1931. [CrossRef]
6. Sishodia, R.P.; Ray, R.L.; Singh, S.K. Applications of Remote Sensing in Precision Agriculture: A Review. *Remote Sens.* **2020**, *12*, 3136. [CrossRef]
7. Weiss, M.; Jacob, F.; Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* **2020**, *236*, 111402. [CrossRef]
8. Zhu, L.; Suomalainen, J.; Liu, J.; Hyyppä, J.; Kaartinen, H.; Haggren, H. A Review: Remote Sensing Sensors. In *Multi-Purposeful Application of Geospatial Data*; Rustamov, R.B., Hasanova, S., Zeynalova, M.H., Eds.; IntechOpen: London, UK, 2017; Available online: https://www.intechopen.com/books/multi-purposeful-application-of-geospatial-data/a-review-remote-sensing-sensors (accessed on 10 October 2023).
9. Kumar, C.; Mubvumba, P.; Huang, Y.; Dhillon, J.; Reddy, K. Multi-Stage Corn Yield Prediction Using High-Resolution UAV Multispectral Data and Machine Learning Models. *Agronomy* **2023**, *13*, 1277. [CrossRef]
10. Da Silva, E.E.; Baio, F.H.R.; Teodoro, L.P.R.; da Silva Junior, C.A.; Borges, R.S.; Teodoro, P.E. UAV-multispectral and vegetation indices in soybean grain yield prediction based on in situ observation. *Remote Sens. Appl. Soc. Environ.* **2020**, *18*, 100318. [CrossRef]
11. 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]
12. Maimaitijiang, M.; Sagan, V.; Sidike, P.; Hartling, S.; Esposito, F.; Fritschi, F.B. Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sens. Environ.* **2020**, *237*, 111599. [CrossRef]
13. Shammi, S.A.; Meng, Q. Use time series NDVI and EVI to develop dynamic crop growth metrics for yield modeling. *Ecol. Ind.* **2021**, *121*, 107124. [CrossRef]
14. Shammi, S.A.; Meng, Q. Modeling crop yield using NDVI-derived VGM metrics across different climatic regions in the USA. *Int. J. Biometeorol.* **2023**, *67*, 1051–1062. [CrossRef]
15. Doraiswamy, P.C.; Sinclair, T.R.; Hollinger, S.; Akhmedov, B.; Stern, A.; Prueger, J. Application of MODIS derived parameters for regional crop yield assessment. *Remote Sens. Environ.* **2005**, *97*, 192–202. [CrossRef]
16. Doraiswamy, P.C.; Hatfield, J.L.; Jackson, T.J.; Akhmedov, B.; Prueger, J.; Stern, A. Crop conditions and yield simulations using Landsat and MODIS. *Remote Sens. Environ.* **2004**, *92*, 548–559. [CrossRef]
17. Mkhabela, M.S.; Bullock, P.; Raj, S.; Wang, S.; Yang, Y. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. *Agric. For. Meteorol.* **2011**, *151*, 385–393. [CrossRef]
18. Johnson, D.M. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sens. Environ.* **2014**, *141*, 116–128. [CrossRef]
19. Bian, C.; Shi, H.; Wu, S.; Zhang, K.; Wei, M.; Zhao, Y.; Sun, Y.; Zhuang, H.; Zhang, X.; Chen, S. Prediction of field-scale wheat yield using machine learning method and multi-spectral UAV data. *Remote Sens.* **2022**, *14*, 1474. [CrossRef]
20. Liu, J.; Huffman, T.; Qian, B.; Shang, J.; Li, Q.; Dong, T.; Davidson, A.; Jing, Q. Crop yield estimation in the Canadian Prairies using Terra/MODIS-derived crop metrics. *IEEE Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2685–2697. [CrossRef]
21. Bolton, D.K.; Friedl, M.A. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agric. For. Meteorol.* **2013**, *173*, 74–84. [CrossRef]
22. Zhang, Z.; Moore, J.C. Chapter 4—Remote Sensing. In *Mathematical and Physical Fundamentals of Climate Change*; Elsevier: Amsterdam, The Netherlands, 2015; pp. 111–124. ISBN 9780128000663. [CrossRef]

23. Bagaram, M.B.; Giuliarelli, D.; Chirici, G.; Giannetti, F.; Barbati, A. UAV remote sensing for biodiversity monitoring: Are forest canopy gaps good covariates? *Remote Sens.* **2018**, *10*, 1397. [CrossRef]
24. 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]
25. Li, F.; Miao, Y.; Feng, G.; Yuan, F.; Yue, S.; Gao, X.; Liu, Y.; Liu, B.; Ustin, S.L.; Chen, X. Improving estimation of summer maize nitrogen status with red edge-based spectral vegetation indices. *Field Crops Res.* **2014**, *157*, 111–123. [CrossRef]
26. Kanemasu, E.T. Seasonal canopy reflectance patterns of wheat, sorghum, and soybean. *Remote Sens. Environ.* **1974**, *3*, 43–47. [CrossRef]
27. Ballester, C.; Brinkhoff, J.; Quayle, W.C.; Hornbuckle, J. Monitoring the Effects of Water Stress in Cotton Using the Green Red Vegetation Index and Red Edge Ratio. *Remote Sens.* **2019**, *11*, 873. [CrossRef]
28. Venancio, L.P.; Mantovani, E.C.; do Amaral, C.H.; Neale, C.M.U.; Gonçalves, I.Z.; Filgueiras, R.; Eugenio, F.C. Potential of using spectral vegetation indices for corn green biomass estimation based on their relationship with the photosynthetic vegetation sub-pixel fraction. *Agric. Water Manag.* **2020**, *236*, 106155. [CrossRef]
29. Wan, L.; Cen, H.; Zhu, J.; Zhang, J.; Zhu, Y.; Sun, D.; Du, X.; Zhai, L.; Weng, H.; Li, Y. Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer—A case study of small farmlands in the South of China. *Agric. For. Meteorol.* **2020**, *291*, 108096. [CrossRef]
30. Zhang, L.; Zhang, Z.; Luo, Y.; Cao, J.; Xie, R.; Li, S. Integrating satellite-derived climatic and vegetation indices to predict smallholder maize yield using deep learning. *Agric. For. Meteorol.* **2021**, *311*, 108666. [CrossRef]