





Article

Multi-Stage Corn Yield Prediction Using High-Resolution UAV Multispectral Data and Machine Learning Models

Chandan Kumar ^{1,2,*} , Partson Mubvumba ² , Yanbo Huang ³ , Jagman Dhillon ¹  and Krishna Reddy ²

¹ Department of Plant and Soil Sciences, Mississippi State University, Starkville, MS 39762, USA; jsd369@msstate.edu

² Crop Production Systems Research Unit, United States Department of Agriculture, Agriculture Research Service, Stoneville, MS 38776, USA; partson.mubvumba@usda.gov (P.M.); krishna.reddy@usda.gov (K.R.)

³ Genetics and Sustainable Agriculture Research Unit, United States Department of Agriculture, Agriculture Research Service, Starkville, MS 39762, USA; yanbo.huang@usda.gov

* Correspondence: fc407@msstate.edu

Abstract: Timely and cost-effective crop yield prediction is vital in crop management decision-making. This study evaluates the efficacy of Unmanned Aerial Vehicle (UAV)-based Vegetation Indices (VIs) coupled with Machine Learning (ML) models for corn (*Zea mays*) yield prediction at vegetative (V6) and reproductive (R5) growth stages using a limited number of training samples at the farm scale. Four agronomic treatments, namely Austrian Winter Peas (AWP) (*Pisum sativum* L.) cover crop, biochar, gypsum, and fallow with sixteen replications were applied during the non-growing corn season to assess their impact on the following corn yield. Thirty different variables (i.e., four spectral bands: green, red, red edge, and near-infrared and twenty-six VIs) were derived from UAV multispectral data collected at the V6 and R5 stages to assess their utility in yield prediction. Five different ML algorithms including Linear Regression (LR), k-Nearest Neighbor (KNN), Random Forest (RF), Support Vector Regression (SVR), and Deep Neural Network (DNN) were evaluated in yield prediction. One-year experimental results of different treatments indicated a negligible impact on overall corn yield. Red edge, canopy chlorophyll content index, red edge chlorophyll index, chlorophyll absorption ratio index, green normalized difference vegetation index, green spectral band, and chlorophyll vegetation index were among the most suitable variables in predicting corn yield. The SVR predicted yield for the fallow with a Coefficient of Determination (R^2) and Root Mean Square Error (RMSE) of 0.84 and 0.69 Mg/ha at V6 and 0.83 and 1.05 Mg/ha at the R5 stage, respectively. The KNN achieved a higher prediction accuracy for AWP ($R^2 = 0.69$ and RMSE = 1.05 Mg/ha at V6 and 0.64 and 1.13 Mg/ha at R5) and gypsum treatment ($R^2 = 0.61$ and RMSE = 1.49 Mg/ha at V6 and 0.80 and 1.35 Mg/ha at R5). The DNN achieved a higher prediction accuracy for biochar treatment ($R^2 = 0.71$ and RMSE = 1.08 Mg/ha at V6 and 0.74 and 1.27 Mg/ha at R5). For the combined (AWP, biochar, gypsum, and fallow) treatment, the SVR produced the most accurate yield prediction with an R^2 and RMSE of 0.36 and 1.48 Mg/ha at V6 and 0.41 and 1.43 Mg/ha at the R5. Overall, the treatment-specific yield prediction was more accurate than the combined treatment. Yield was most accurately predicted for fallow than other treatments regardless of the ML model used. SVR and KNN outperformed other ML models in yield prediction. Yields were predicted with similar accuracy at both growth stages. Thus, this study demonstrated that VIs coupled with ML models can be used in multi-stage corn yield prediction at the farm scale, even with a limited number of training data.

Keywords: corn (*Zea mays*); corn yield prediction; k-nearest neighbor; multispectral data; support vector regression; agronomic treatment; machine learning; UAV; vegetation indices



Citation: 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. <https://doi.org/10.3390/agronomy13051277>

Academic Editor: Francisco Manzano Agugliaro

Received: 27 March 2023

Revised: 26 April 2023

Accepted: 27 April 2023

Published: 28 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Corn (*Zea mays*) is among the most important crops that play a vital role in global food security. Along with rice and wheat, corn provides ~30% of the food calories to

more than 4.5 billion people worldwide [1]. Corn contributes about 96% of feed grain use in the US [2]. The US is among the leading producers of corn followed by China, Brazil, Argentina, and Ukraine. Agricultural practices such as soil fertility, irrigation, and crop management are aimed at increasing crop yield to meet the demands of the world's population. Despite significant development in agricultural sectors, over 800 million people continue to experience chronic hunger, as reported by the Food and Agriculture Organization (FAO), and issues like climate change, pandemics, and international conflicts make things even worse [3]. In the context of global food security and demand, a timely and reliable crop yield prediction is crucial in effective decision and policy making [4–7].

Conventional yield estimation is labor-intensive, time and cost ineffective, and often accomplished at the end of the season. Remotely sensed data acquired using spaceborne, airborne, and Unmanned Aerial Vehicles (UAVs) have been successfully used for crop yield prediction [4,8–12]. The major advantages of remote sensing-based crop yield prediction are that it is reliable, time and cost-effective, and can be used across different growth stages of crops, facilitating efficient crop management [4,6,13,14]. Furthermore, compared to space and airborne platforms, UAVs offer a high spatio-temporal resolution, flexible acquisition windows, and less atmospheric attenuation, making them more suitable for crop monitoring and yield prediction at the farm or field scale [12,15–18]. Spatio-temporal farm-scale crop monitoring and yield prediction are crucial for growers and insurance agencies to make better-informed decisions.

The multispectral sensor mounted on a UAV consists of suitable spectral bands within the Visible and Near Infrared (VNIR) range, which bands are highly effective in deriving various Vegetation Indices (VIs) sensitive to crop health [19,20]. The UAV-based multispectral data, coupled with Machine Learning (ML) models, has been effectively used in monitoring and predicting the yield of various crops such as corn [21–27], wheat [17,28,29], rice [30–33], soybean [34], cotton [35,36], and others [37–39]. ML is a branch of artificial intelligence that offers several advantages over conventional statistical models, such as the ability to learn complex and non-linear relationships, it can deal with a wide range of variables, higher accuracy, and can handle big data [40,41]. ML learning techniques have also been successfully used in other fields including geology [42,43], geohazards [44], forestry [45], and others [41]. The successful implementation of ML algorithms depends on four crucial steps: suitable training data, relevant variables, hyperparameter optimization, and robust validation approaches [40,42–44].

Crop yield prediction has been successfully achieved using various machine learning (ML) algorithms such as Logistic Regression (LR), k-Nearest Neighbor (KNN), Support Vector Machine/Regression (SVM/SVR), Decision Trees (DTs), Random Forest (RF), Multivariate Adaptive Regression Splines (MARS), Artificial Neural Networks (ANN), Least Absolute Shrinkage and Selection Operator (LASSO), Gradient Boost Regression Tree (GBRT), and others [7,14,17,46–50]. However, the performance of ML models varies based on several factors, including training data, input variables, crop types, and growth stages. For example, Mupangwa et al. [25] evaluated six ML algorithms (LR, LDA, NB, KNN, CART, and SVM) in corn yield prediction and found that KNN produced the best prediction results whereas SVM produced the poorest prediction accuracy. Croci et al. [50] achieved higher accuracy in corn prediction from Gaussian process regression, SVR, and single-layer perceptron feed-forward neural networks as compared to RF, KNN, and cubic regression models. Matsumura et al. [51] obtained higher corn yield prediction using DNN than LR. Kim and Lee [52] obtained accurate corn yield prediction using deep learning methods as compared to SVR, RF, and extremely randomized trees. Shahhosseini et al. [26] achieved a better yield prediction using extreme gradient boosting as compared to LASSO regression, ridge regression, and RF. Recent studies also reported higher accuracy in crop yield prediction using different deep learning models as compared to classical ML models but required a substantially higher amount of training data for model development and validation [53]. Therefore, it is crucial to evaluate the ML model's performance to select the most suitable one for achieving accurate results.

Previous ML-based crop yield prediction studies have utilized a significant number of training samples and a diverse set of variables including VIs, soil nutrients and properties, and meteorological data. However, such comprehensive datasets are often limited to researchers and big consulting companies. This poses a challenge for small farmers and growers as obtaining suitable training data and comprehensive input variables at a farm level can be difficult. Therefore, the main objective of this study was to evaluate the effectiveness of UAV-based VIs for predicting corn yield at vegetative and reproductive growth stages using different ML models with a limited number of training samples. Moreover, we assessed the influence of agronomic treatments on overall yield and their potential impact on yield prediction. We further assessed the influence of the number of variables on ML models' performance. Lastly, we evaluated the performance of five state-of-the-art ML algorithms (i.e., LR, KNN, RF, SVR, and DNN) for predicting corn yield at vegetative and reproductive growth stages.

2. Materials and Methods

2.1. Materials

Experimental Design and Yield Data Collection

The study was conducted at the United State Department of Agriculture-Agricultural Research Service, Crop Production Systems Research Unit farm in Stoneville, Mississippi. Figure 1 displays the geographical location of the experimental site and the design of agronomic treatments applied in this study. The agronomic treatments used were Austrian Winter Peas (AWP) (*Pisum sativum* L.) cover crop, biochar, gypsum, and fallow (i.e., left bare) during the non-growing corn season (i.e., October 2020 to April 2021) to assess their impact on the following corn yield. A seeding rate of 67 kg/ha for AWP and quantities of 15 t/ha and 2 t/ha for biochar and gypsum, respectively, were applied in this experiment. The description of treatments is summarized in Table 1. The experiment was designed with plot sizes of 12.50 × 8.50 m and 16 replications in a completely randomized block design. The corn variety: Dekalb DKC 62-08 was planted at a seeding rate of 55,352 seed/ha using John Deere 1705 Planter at a 1 m row spacing on 7 May 2021. The corn was harvested on 15 September 2021, using a Massey-Ferguson 2065 combine harvester to obtain the yield of each plot.

Table 1. Specification and description of agronomic treatments applied in this experiment.

Treatments	Specification	Description
AWP	Seeding date = 6 November 2020 Seeding rate = 67 kg/ha Harvesting date = 29 April 2021	AWP is a legume cover crop that can improve soil health, thus enhancing the soil's physical, chemical, and biological properties. It improves nutrient cycling, increases nitrogen, sequesters carbon, and enhances soil aggregation, water infiltration, storage capacity, and use efficiency [54,55].
Biochar	Quantity = 15 t/ha Application date = 12 November 2020	Biochar is a soil amendment. It is made from sugarcane bagasse (sugarcane stalk residue after juice extraction). It increases soil organic carbon, soil pH, and microbial activity, improving soil structure, soil porosity, soil water holding capacity, cation exchange capacity, nutrient cycling, and plant growth and yields [56,57].
Gypsum	Quantity = 2 t/ha Application date = 20 November 2020	Gypsum (Calcium sulfate) is a soil amendment. It improves soil fertility by increasing sulfur, phosphorus, calcium, magnesium, manganese, and enhances plant nitrogen use efficiency [58,59].
Fallow	Left bare during the winter period from November 2020 to April 2021.	This is the common practice in the Delta, Mississippi between harvest and planting. The soil is exposed to erosion with essential nutrients susceptible to losses through leaching or runoff.

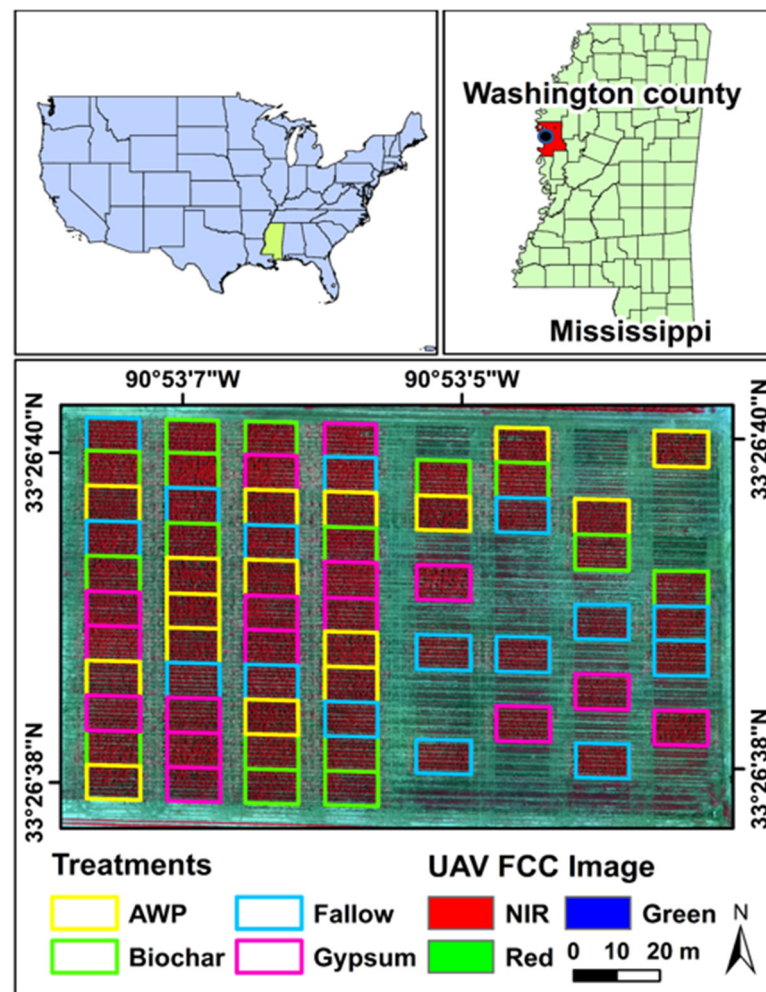


Figure 1. Experimental location and design showing the different agronomic treatments, namely Austrian Winter Peas (AWP) cover crop, biochar, gypsum, and fallow, applied in this experiment. The vector file of treatments wrapped over the False Color Composite (FCC) derived using UAV multispectral data.

2.2. Methods

A workflow diagram of the adopted methodology is shown in Figure 2. This section describes the experimental design and yield data collection, UAV data acquisition and processing, derivation of VIs and training data preparation, suitable variable selection, ML implementation, validation, and performance evaluation of ML models in corn yield prediction.

2.2.1. UAV Data Acquisition and Processing

Corn field images were acquired using a DJI Phantom 3 Pro quadcopter UAV (DJI Phantom). On the UAV, a portable Parrot Sequoia multispectral camera was mounted that captures broadband RGB (14 MP), and narrowband green (550 nm, ± 40 nm), red (660 nm, ± 40 nm), red edge (735 nm, ± 10 nm), and NIR (790 nm, ± 40 nm) wavelengths. The camera automatically corrects brightness and synchronized global position system (GPS) positions with the UAV. The UAV flights were conducted between 10:30 a.m. and 12:00 p.m. to avoid cloud shadows as weather permitted, with flight altitude 30 m above the canopy surface to acquire high-resolution (~ 3 cm/pixel) image data during vegetative (V6) (16 June 2021) and reproductive stages (R5) (9 September 2021). Flight routes were preset using the mission planning tool of Pix4DCapture software, version 4.7.5 with image front overlap of 80% and side overlap of 70%. The collected images were imported to Pix4DMapper

(<https://www.pix4d.com>, accessed on 15 September 2022) to generate RGB orthomosaics in broadband and green, red, red edge, and NIR orthomosaics in narrow band spectrum, which were orthorectified to correct geometric and vignetting distortion. Orthomosaic images were imported to ArcMap to draw the boundary of each plot based on different treatments. An R script was written to extract the mean values of different spectral bands and VIs within each experimental plot.

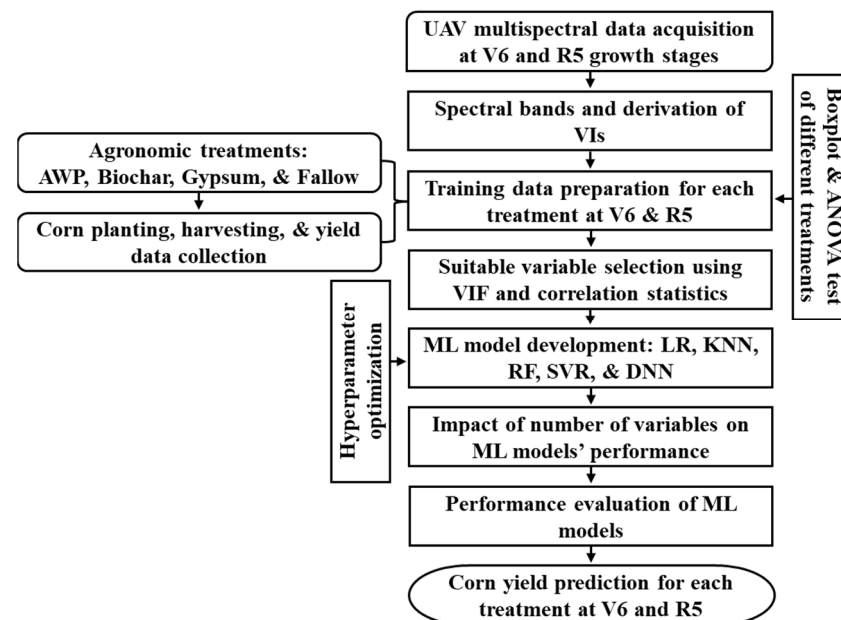


Figure 2. A flowchart diagram presenting the methodology adopted in this study.

2.2.2. Derivation of VIs

VIs optically characterize crop health across their growth stages. The acquired multispectral datasets of vegetative and reproductive stages were used in deriving twenty six different VIs, including the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index 2 (EVI2), Soil Adjusted Vegetation Index (SAVI), Modified Soil Adjusted Vegetation Index (MSAVI), Optimized Soil Adjusted Vegetation Index (OSAVI), Transformed Vegetation Index (TVI), Green Normalized Difference Vegetation Index (GNDVI), Normalized Difference Red Edge Index (NDRE), Normalized Green Red Difference Index (NGRDI), Normalized Crop Management Index (NCMI), Simplified Canopy Chlorophyll Content Index (SCCCI), Canopy Chlorophyll Content Index (CCCI), Renormalized Difference Vegetation Index (RDVI), Transformed Chlorophyll Absorption Reflectance Index (TCARI), Normalized Area Vegetation Index (NAVI), Chlorophyll Index Red Edge (CIRE), Modified Triangular Vegetation Index (MTVI), Modified Triangular Vegetation Index 2 (MTVI2), Red Edge Chlorophyll Vegetation Index (RECI), ratio of NIR and green bands (IV1), ratio of NIR and red bands (IV2), ratio of NIR and red edge bands (IV3), Green Chlorophyll Vegetation Index (GCVI), Chlorophyll Vegetation Index (CVI), Napierian logarithm of the red edge (LNRE), and the ratio of the red edge and the red spectral band produces the Chlorophyll Absorption Ratio Index (CARI). Table 2 provides a list of derived VIs with their mathematical formulas.

The mean values of spectral bands and VIs were extracted based on each treatment at both stages to prepare training data for corn yield prediction. For each treatment, 16 training samples were obtained with thirty independent variables. To prepare the training data for all treatments, the yield data of each treatment were combined, resulting in 64 training samples with thirty independent variables. The spectral bands and VIs are hereafter referred to as variables for improved readability.

Table 2. Implemented Vegetation Indices (VIs) and their mathematical equations. G: green, R: red, RE: red edge, and NIR: near-infrared spectral bands of UAV multispectral data.

VIs	Equation	References
CARI	$CARI = RE/R$	[60]
CCCI	$CCCI = NDRE - NDRE_{min}/NDRE_{max} - NDRE_{min}$	[61]
CIRE	$CIRE = NIR/(RE - 1)$	[31]
CVI	$CVI = NIR/(R/G^2)$	[62]
EVI2	$EVI2 = 2.5(NIR - R)/(NIR + 2.4R + 1)$	[17]
GCVI	$GCVI = NIR/G - 1$	[63]
GNDVI	$GNDVI = (NIR - G)/(NIR + G)$	[34]
IV1	$IV1 = NIR/G$	[34]
IV2	$IV2 = NIR/R$	[34]
IV3	$IV3 = NIR/RE$	[34]
LNRE	$LNRE = 100(\ln(NIR) - \ln(R))$	[34]
MSAVI2	$MSAVI2 = 0.5 \left((2NIR + 1) - \sqrt{2NIR^2 - 8(NIR - R)} \right)$	[34]
MTVI	$MTVI = 1.2(1.2(NIR - G) - 2.5(R - G))$	[5]
MTVI2	$MTVI2 = \left(1.5(1.2(NIR - G) - 2.5(R - G)) / \sqrt{(2NIR + 1)^2 - (6NIR - 5\sqrt{R})} - 0.5 \right)$	[5]
NAVI	$NAVI = (NIR - R)/NIR$	[62]
NCMI	$NCMI = NIR - (G + R)/NIR + (G + R)$	[64]
NDRE	$NDRE = (NIR - RE)/(NIR + RE)$	[34]
NDVI	$NDVI = (NIR - R)/(NIR + R)$	[34]
NGRDI	$GNDVI = (G - R)/(G + R)$	[65]
OSAVI	$OAVI = (1.16(NIR - R))/(NIR + R + 0.16)$	[66]
RDVI	$RDVI = NIR - R/(NIR + R)^{0.5}$	[32]
RECI	$RECI = (NIR - R)/R$	[65]
SAVI	$SAVI = (1.5(NIR - R))/(NIR + R + 0.5)$	[34]
SCCCI	$SCCCI = NDVI/NDRE$	[34]
TCARI	$TCARI = 3((RE - R) - 0.2(RE - G)(RE/R))$	[5]
TVI	$TVI = 0.5(120(RE - G) - 200(R - G))$	[5]

2.2.3. ML Implementation

Five different ML algorithms, namely LR, KNN, SVM, RF, and DNN, were used in this study. We selected these algorithms as they have been widely used in different fields, including crop yield prediction. Furthermore, they represent unique learning mechanics with varied model complexity. The comparison of these algorithms can provide useful insight into their performance in context applied to limited training samples for yield prediction. ML algorithms were implemented using the ‘CARET’ package [67] in R statistical programming language [68] and are briefly described below.

LR

The LR 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 variance), training samples are normally distributed and statistically independent, and there is linearity between dependent and independent variables [69].

KNN

The KNN Is a non-parametric and computationally efficient ML algorithm that stores available training samples and predicts the dependent variable based on similarity measures (i.e., Euclidean, Manhattan, and Minkowski distance) [70]. The samples of similar characteristics produce a lower distance and vice versa. The implemented KNN algorithm

consists of one tuning hyperparameter, i.e., k , which describes the number of neighborhood samples considered in prediction.

RF

The RF uses a bagging technique where many decision trees (DTs) are developed to obtain an ensemble model for accurate classification or prediction results [70]. 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.

SVR

The SVR employs the same principle of maximal margin as SVM in the classification task, which enables an optimal hyperplane to be obtained and minimizes the difference between predicted and observed values [70]. The availability of kernel functions, such as linear, polynomial, radial basis function, and sigmoid, facilitates the development of optimal hyperplanes to produce higher accuracy. The employed SVR consists of two hyperparameters: cost and sigma. Cost denotes penalty and sigma indicates the complexity of the hyperplane.

DNN

The typical architecture of a neural network algorithm consists of input, hidden, and output layers [70]. The hidden layer(s) apply a transformation to learn the structure and pattern from the input data. The neural network employs the backpropagation method to allow hidden layer(s) to adjust the weight of neurons to obtain the desired output [70]. The implemented neural network algorithm consists of three hyperparameters, layer1, layer2, and layer3, as hidden layers, and can be called a Deep Neural Network (DNN).

2.2.4. Model Performance Measures

Considering the small number of training samples (i.e., sixteen for each treatment) in this study, which limits the training and testing data split, a K-fold cross-validation method is an effective approach for the performance evaluation of the different ML models [71]. This approach has been widely used in ML model performance assessment [71–73]. In K-fold cross-validation, the complete dataset is split into a training set with a different set of tests at every iteration to compute the performance measures. We used the coefficient of determination (R^2) (Equation (1)) and Root Mean Square Error (RMSE) (Equation (2)) in this study as these are commonly used in assessing the prediction performance of ML models [17,25]. A 5-fold cross-validation with 10 repetitions (i.e., 50 iterations) was used to compute R^2 and the RMSE (Figure 3) to assess the performance of the different ML models. A fold size of 5 was selected to ensure that approximately 75% of data were used for training and 25% for the validation sample at each iteration to compute the different performance measures.

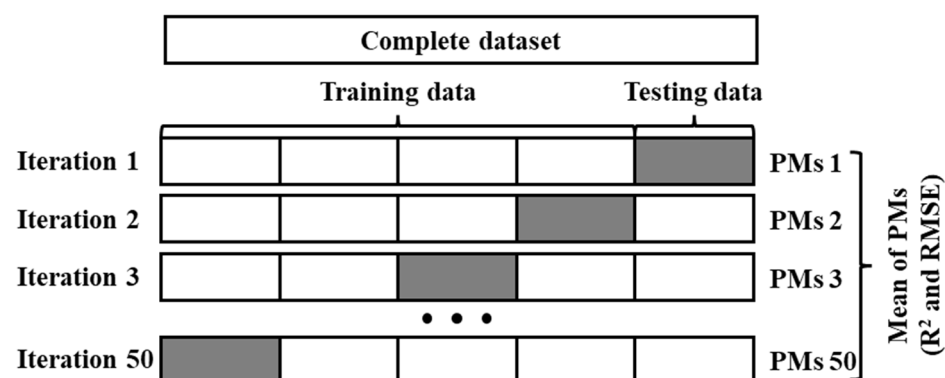


Figure 3. A schematic diagram of a 5-fold cross-validation method for computing the performance measures of different ML models. PMs: performance measures.

$$R^2 = (\sum_{i=1}^k (y_i - \bar{y})(f_i - \bar{f}))^2 / \sum_{i=1}^k (y_i - \bar{y})^2 \sum_{i=1}^k (f_i - \bar{f})^2 \quad (1)$$

$$RMSE = \sqrt{\frac{1}{k} \sum_{i=1}^k (y_i - f_i)^2} \quad (2)$$

2.2.5. Suitable Variable Selection

Pearson's correlation coefficient and the Variance Inflation Factor (VIF) were used in this study to remove highly correlated independent variables [74]. The correlation value between yield and independent variables assists in assessing their relative importance in explaining the yield variability whereas the VIF statistics assist in removing highly correlated variables. The correlation values range from -1 to 1 . X_i and Y_i denote the corresponding values of X and Y for the i -th variable. \bar{X} and \bar{Y} represent the means of X and Y . r_{xy} is the Pearson correlation coefficient between two variables (i.e., X and Y) (Equation (3)). We also used the correlation plot to visually interpret the correlation among different variables. Let $X = X_1, X_2, X_3, \dots, X_N$ represent the independent variable set and R_j^2 represent the multicollinear coefficient between X_j and other variables. The VIF can be computed using Equation (4). A VIF value > 5 indicates higher multicollinearity among independent variables and should be discarded as these variables can negatively impact the model performance and interpretability [75].

$$r_{xy} = \sum_{i=1}^n \frac{X_i - \bar{X}}{\sum_{k=1}^n (X_k - \bar{X})} \times \frac{Y_i - \bar{Y}}{\sum_{k=1}^n (Y_k - \bar{Y})} \quad (3)$$

$$VIF = \frac{1}{(1 - R_j^2)} \quad (4)$$

3. Results

3.1. Relationship between Agronomic Treatments and Yield

A boxplot (Figure 4) and analysis of variance (ANOVA) test (Table 3) confirmed that a one-year experiment with different agronomic treatments did not have a significant impact on corn yields. However, the AWP treatment indicated the lowest impact in improving the total yield (i.e., mean = 18.84 Mg/ha) whereas biochar (i.e., mean = 19.06 Mg/ha), gypsum (i.e., mean = 19.48 Mg/ha), and fallow (i.e., mean = 19.82 Mg/ha) treatments showed a slightly higher yield. The AWP treatment indicated the lowest variability in yield compared to the other treatments. The yield values of AWP and the other treatments are approximately symmetrically distributed and positively skewed, respectively.

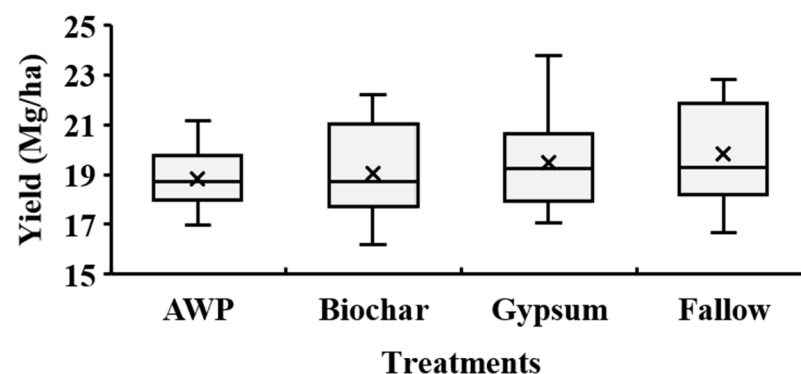


Figure 4. A boxplot of different agronomic treatments and total corn yield.

Table 3. One-way ANOVA test statistics of different agronomic treatments.

	Degrees of Freedom	Sum of Squares	Mean of the Sum Squares	F-Value	Pr(>F)
Treatment	3	9.18	3.06	0.98	0.41
Residuals	60	187.52	3.13		

3.2. Relationship between VIs and Yield

The linear correlation coefficient and VIF statistics were used in selecting suitable variables (i.e., spectral bands and VIs). The variables produce VIF values ≥ 5 were excluded as these indicate multicollinearity among them. The correlation values between different variables and yield at the V6 and R5 stages of different agronomic treatments are presented in Table 4. The correlation between variables and yield varies greatly based on different treatments and growth stages. Variables indicate positive and negative correlations with yield, however, the magnitude of the correlation is of main interest as compared to the positive/negative relationship. The red edge and GNDVI have the maximum correlation with the yield (i.e., 0.40 and -0.31) at V6 and R5, respectively, for the AWP treatment. The CCCI and red edge have the maximum correlation with yield (i.e., -0.63 and -0.60) at V6 and R5 stages for the biochar treatment. The RECI and green have the maximum correlation with the yield (i.e., -0.59 and 0.69) at V6 and R5 for the gypsum treatment. The CARI and CVI have the maximum correlation with the yield (i.e., -0.87 and -0.69) at V6 and R5 for the fallow.

Table 4. Correlation matrix of suitable variables (i.e., spectral bands and vegetation indices) and yield of different treatments at vegetative (V6) and reproductive (R5) stages. The bold font indicates the highest correlation between the variable and yield.

Vegetative (V6)					Reproductive (R5)					
					AWP					
	Rede	NIR	CVI	CARI	Yield	NIR	SCCCI	NCMI	IV3	GNDVI
Yield	0.40	0.32	0.02	0.14	Yield	-0.17	-0.08	-0.25	0.11	-0.31
Rede	1.00	0.60	0.28	0.73	NIR	1.00	-0.10	0.65	0.05	0.71
NIR		1.00	0.61	0.69	SCCI		1.00	0.19	-0.23	-0.19
CVI			1.00	0.69	NCMI			1.00	-0.44	0.74
					IV3				1.00	-0.34
					Biochar					
	Green	NGRDI	CCCI		Yield	Rede	TCARI	SCCCI		
Yield	0.45	-0.33	-0.63		Yield	-0.60	-0.42	-0.31		
Green	1.00	-0.61	-0.16		Rede	1.00	0.74	0.49		
NGRDI		1.00	-0.03		TCARI		1.00	0.60		
					Gypsum					
	Green	RECI	CCCI		Yield	Green	NGRDI	IV3	CVI	
Yield	0.53	-0.59	-0.57		Yield	0.69	0.40	0.53	-0.50	
Green	1.00	-0.67	-0.43		Green	1.00	0.33	0.71	-0.65	
RECI		1.00	0.39		NGRDI		1.00	-0.18	-0.73	
					IV3			1.00	-0.22	
					Fallow					
	Rede	CARI			Yield	Green	SCCCI	NDRE	CVI	CARI
Yield	-0.47	-0.87			Yield	0.36	0.41	-0.11	-0.69	0.47
Rede	1.00	0.65			Green	1.00	-0.32	0.19	-0.13	-0.23
					SCCI		1.00	-0.04	-0.13	0.39
					NDRE			1.00	0.41	-0.47
					CVI				1.00	-0.72
					All treatments					
	Green	MTVI2	CCCI		Yield	CVI	GCVI	Green	IV3	
Yield	0.51	-0.36	-0.28		Yield	-0.40	-0.39	0.35	0.20	
Green	1.00	-0.67	-0.22		CVI	1.00	0.41	-0.24	0.27	
MTVI2		1.00	0.16		GCVI		1.00	-0.39	-0.23	
					Green			1.00	0.49	

Figures 5 and 6 present the correlation between suitable variables and the yield for each treatment at the V6 and R5 stages, respectively. For the AWP treatment, the red edge, NIR, CARI, and CVI are suitable VIs at V6 whereas the GNDVI, NCMI, NIR, IV3, and SCCC are suitable VIs at R5. Similarly, for the biochar, the CCCI, green, and NGRDI are suitable variables at V6 whereas red edge, TCARI, and SCCC are suitable at R5. For the gypsum, the RECI, CCCI, and green are suitable variables at V6 whereas the green, IV3, CVI, and NGRDI are suitable at R5. For the fallow, the CARI and red edge are suitable variables at V6 whereas the CVI, CARI, SCCC, green, and NDRE are suitable variables at R5. The suitable variables for each treatment at both growth stages were used in assessing the impact of the number of variables on ML model performance.

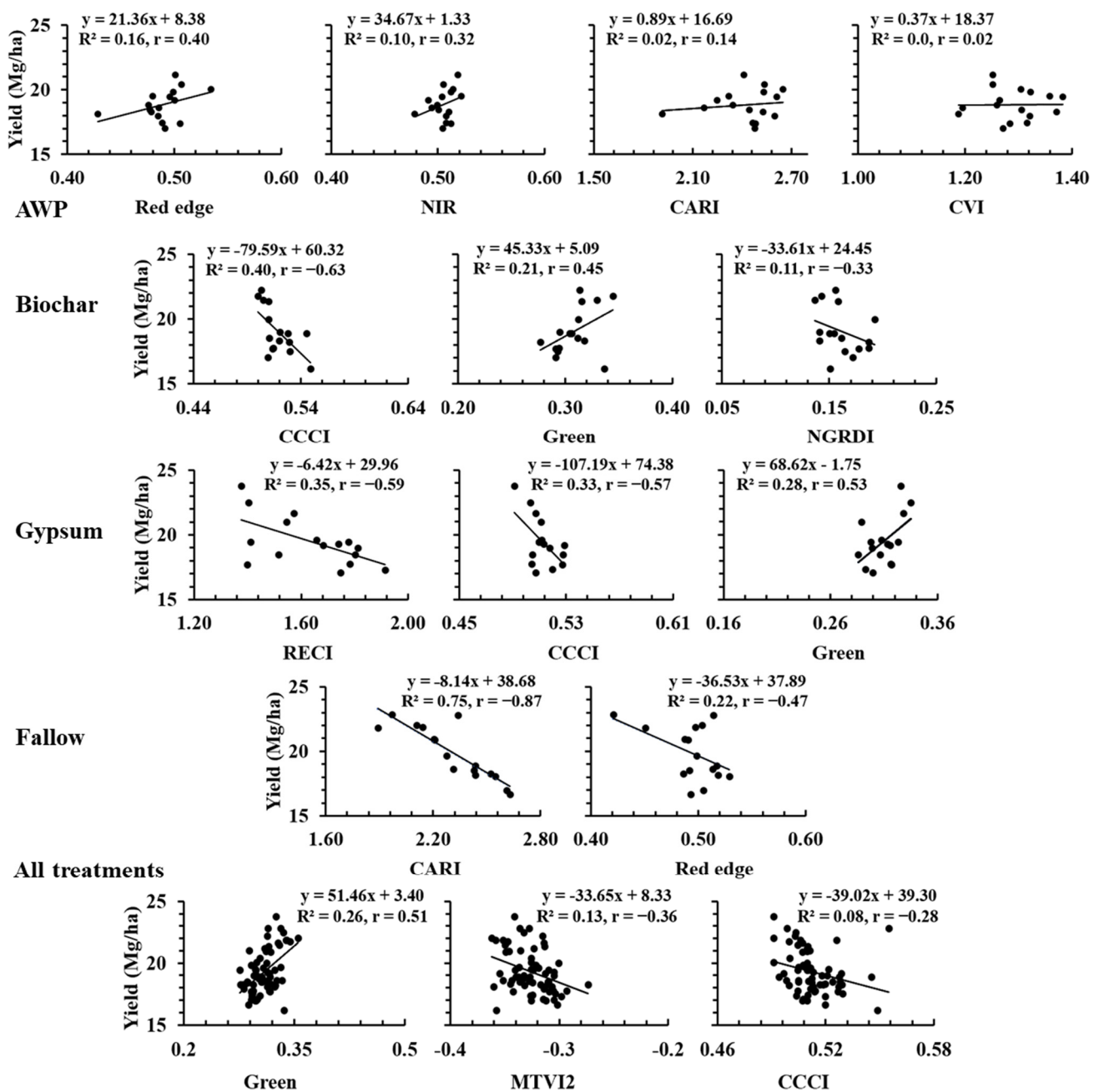


Figure 5. Correlation between suitable variables and yield for each treatment at the V6 stage.

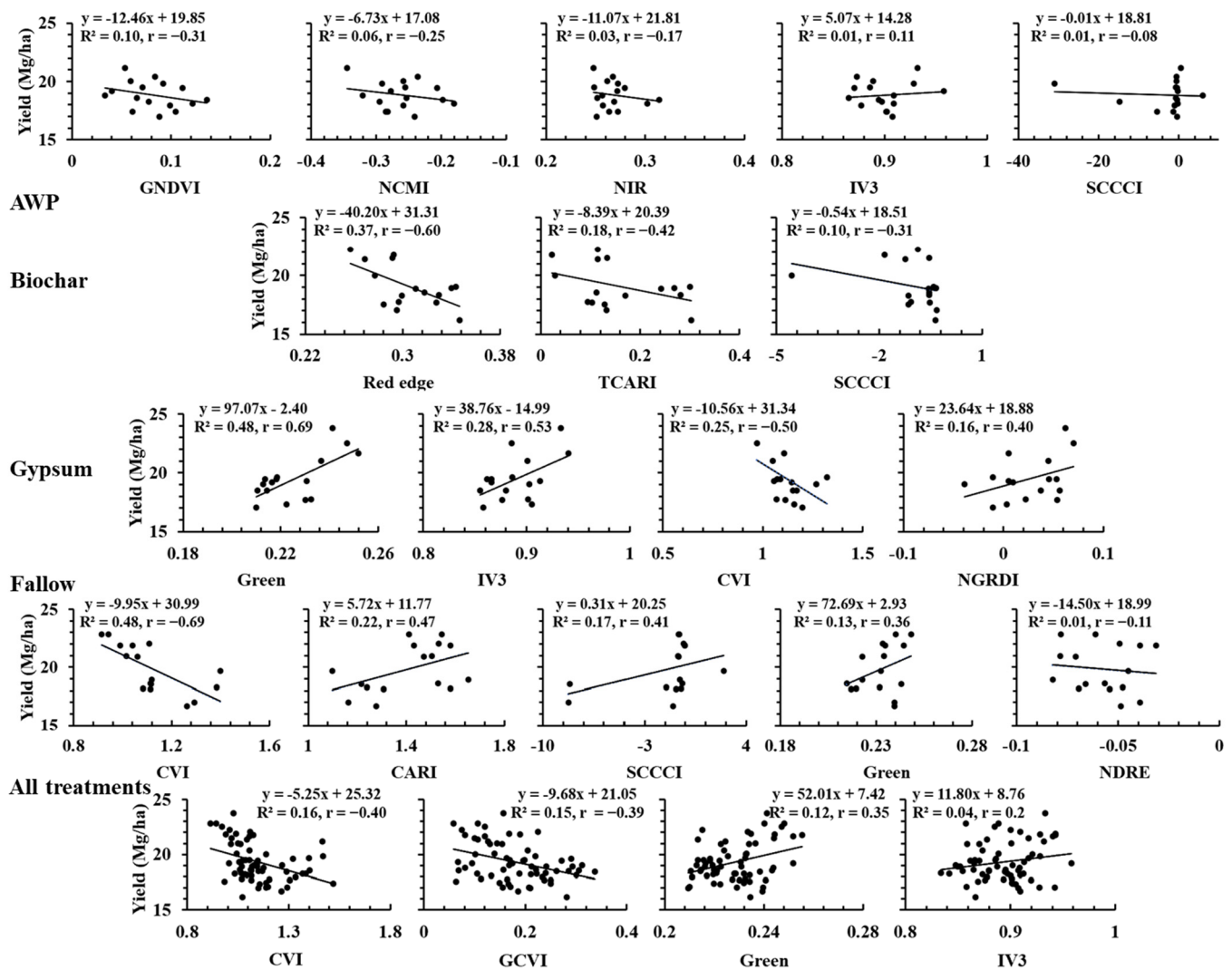


Figure 6. Correlation between suitable variables and corn yield for each treatment at the R5 stage.

It is worth noting that the R^2 of the suitable variables and yield is significantly lower for the AWP than for the other treatments at both growth stages. The R^2 for the AWP treatment is in the range of 0–0.16 at V6 and 0–0.10 at R5. For biochar treatment, the R^2 is in the range 0.11–0.40 at V6 and 0.10–0.37 at R5. The R^2 for the gypsum treatment is in the range of 0.28–0.35 at V6 and 0.16–0.48 at R5. For the fallow, the selected variables indicated a higher correlation with yield as compared with other treatments. The R^2 is in the range of 0.13–0.48 at V6 and 0.01–0.48 at R5. Furthermore, the VIs show a similar correlation at both growth stages across different treatments, indicating their suitability for use in yield prediction irrespective of the growth stage.

3.3. Impact of the Number of Variables on ML Performance

The optimal number of variables is crucial in obtaining accurate results from any ML model. For each treatment, the impact of the number of variables (i.e., one variable, two variables, and so on) on ML performance was assessed to derive the optimal number of variables to use in the final yield prediction at both stages. The hyperparameter of the ML models was optimized using the grid search 5-fold cross-validation method before assessing their performance.

Figures 7 and 8 and Tables 5 and 6 present the prediction accuracies of the ML models using a different number of variables at the V6 and R5 growth stages. In most cases, ML models produced the best prediction results using the top one or two variables across

different treatments at both stages. The top one or two variables having a higher correlation with yield indicate a higher contribution to predicting yield accurately as compared with variables showing a lesser correlation with yield. The addition of more variables did not improve the model's performance significantly, except for a few occasions. As compared with other ML models, the DNN shows a higher sensitivity to the number of variables. For example, for the AWP treatment at V6, the R^2 of DNN increased from 0.58 to 0.67 when the number of variables increased from one to two but showed a decreasing trend in performance when the number of variables increased to three (0.64) or four variables (0.62).

Table 5. Impact of the number of suitable variables on ML models' performance for each treatment at the V6 growth stage. The bold font indicates the best accuracy achieved by the ML models for each treatment.

	Number of Suitable Variables							
	One		Two		Three		Four	
	R ²	RMSE	R ²	RMSE	AWP		R ²	RMSE
LR	0.62	1.12	0.55	1.21	0.53	1.18	0.52	1.32
RF	0.68	1.02	0.66	1.09	0.62	1.09	0.57	1.12
KNN	0.69	1.05	0.65	1.13	0.61	1.23	0.58	1.85
SVR	0.65	1.17	0.58	1.39	0.53	1.41	0.64	2.45
DNN	0.58	1.57	0.67	1.62	0.64	1.67	0.62	1.83
					Biochar			
LR	0.59	1.44	0.55	1.59	0.54	1.70		
RF	0.63	1.27	0.61	1.35	0.62	1.42		
KNN	0.64	1.70	0.66	1.20	0.68	1.29		
SVR	0.62	1.48	0.67	1.19	0.56	1.41		
DNN	0.57	1.74	0.71	1.08	0.67	1.29		
					Gypsum			
LR	0.59	1.56	0.58	1.66	0.57	1.79		
RF	0.51	1.90	0.56	1.74	0.58	1.64		
KNN	0.60	1.73	0.61	1.49	0.65	1.75		
SVR	0.62	1.87	0.58	1.65	0.57	1.52		
DNN	0.55	3.21	0.51	2.80	0.54	2.26		
					Fallow			
LR	0.84	0.96	0.81	1.00				
RF	0.74	1.03	0.73	1.09				
KNN	0.84	1.25	0.79	1.60				
SVR	0.84	0.69	0.78	1.15				
DNN	0.78	0.99	0.68	1.74				
			All treatments (i.e., combined AWP, biochar, gypsum, and fallow)					
LR	0.32	1.52	0.31	1.54	0.28	1.60		
RF	0.16	1.80	0.19	1.69	0.17	1.68		
KNN	0.30	1.62	0.31	1.54	0.28	1.55		
SVR	0.33	1.52	0.36	1.48	0.29	1.57		
DNN	0.16	2.72	0.18	2.29	0.17	1.99		

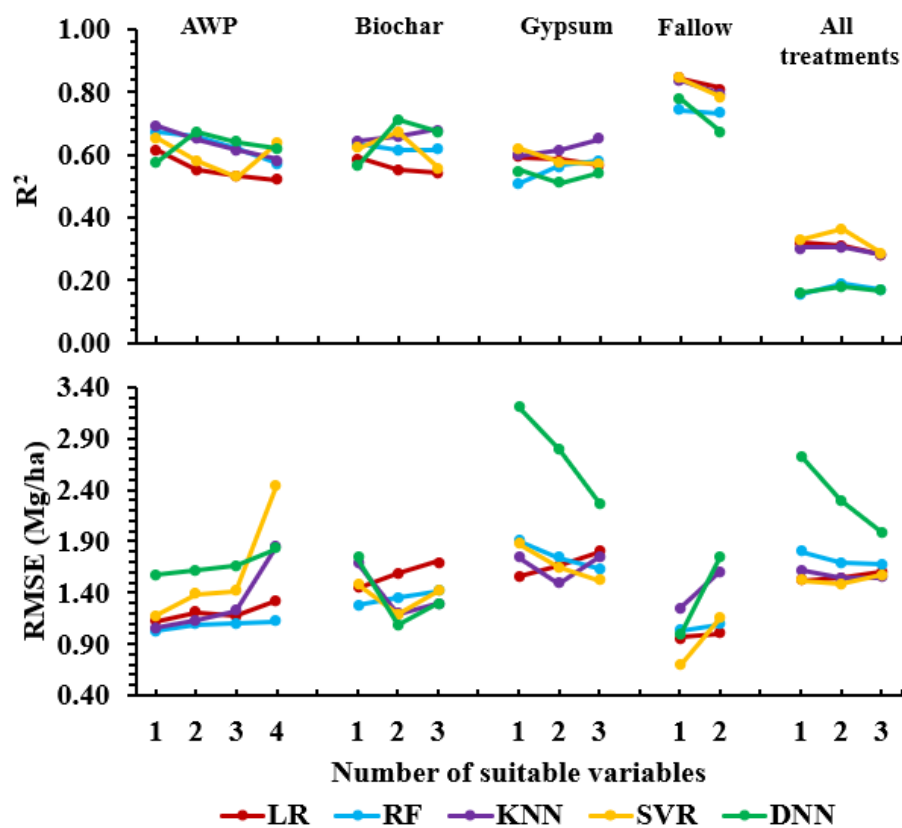


Figure 7. Impact of the number of suitable variables on ML models' performance in yield prediction at the V6 stage.

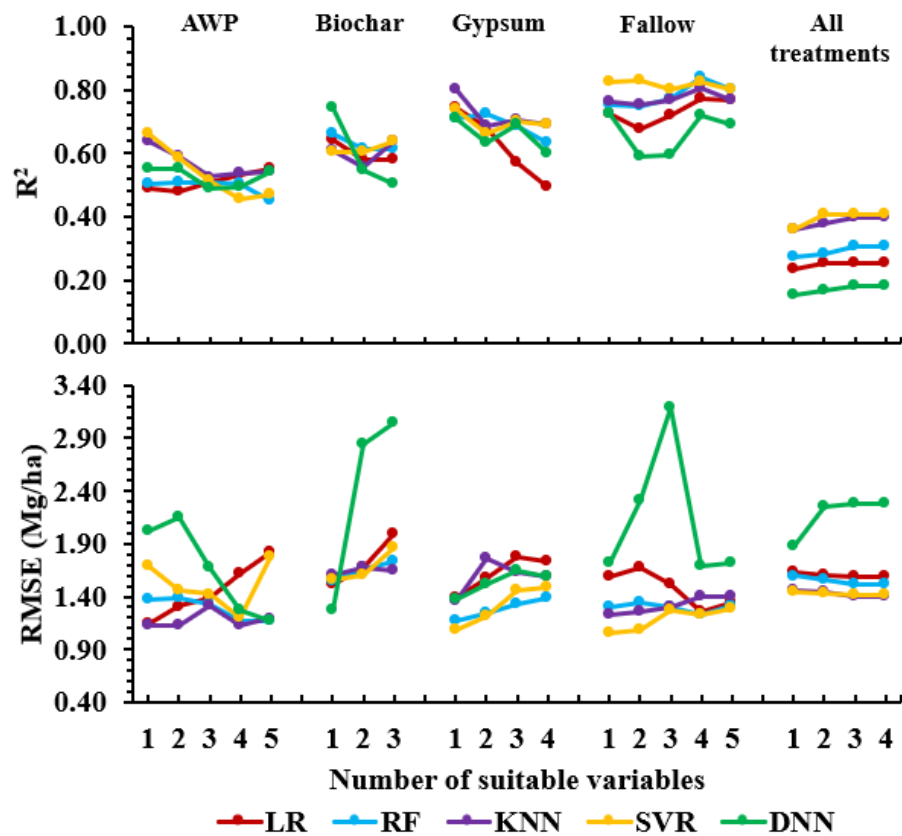


Figure 8. Impact of number of suitable variables on ML models' performance in yield prediction at the R5 stage.

Table 6. Impact of the number of suitable variables on ML models' performance for each treatment at the R5 growth stage. The bold font indicates the best accuracy achieved by the ML models for each treatment.

	Number of Suitable Variables									
	One		Two		Three		Four		Five	
	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE
AWP										
LR	0.49	1.15	0.48	1.31	0.51	1.39	0.53	1.62	0.55	1.83
RF	0.50	1.38	0.51	1.39	0.51	1.33	0.50	1.18	0.45	1.17
KNN	0.64	1.13	0.59	1.13	0.52	1.31	0.54	1.12	0.54	1.19
SVR	0.66	1.68	0.59	1.46	0.51	1.42	0.45	1.20	0.47	1.77
DNN	0.55	2.03	0.55	2.15	0.49	1.68	0.49	1.28	0.54	1.17
Biochar										
LR	0.64	1.52	0.58	1.67	0.58	1.99				
RF	0.66	1.55	0.61	1.67	0.61	1.73				
KNN	0.61	1.60	0.56	1.68	0.64	1.65				
SVR	0.60	1.57	0.60	1.60	0.64	1.87				
DNN	0.74	1.27	0.55	2.84	0.50	3.04				
Gypsum										
LR	0.74	1.39	0.68	1.57	0.57	1.78	0.49	1.74		
RF	0.71	1.17	0.72	1.25	0.68	1.33	0.63	1.39		
KNN	0.80	1.35	0.69	1.76	0.71	1.63	0.69	1.58		
SVR	0.74	1.08	0.66	1.21	0.70	1.45	0.69	1.48		
DNN	0.71	1.37	0.63	1.52	0.69	1.64	0.60	1.58		
Fallow										
LR	0.72	1.59	0.67	1.67	0.72	1.51	0.77	1.26	0.77	1.34
RF	0.75	1.30	0.75	1.34	0.77	1.30	0.84	1.22	0.80	1.31
KNN	0.76	1.23	0.75	1.26	0.77	1.29	0.81	1.41	0.77	1.39
SVR	0.83	1.05	0.83	1.08	0.80	1.28	0.82	1.23	0.80	1.28
DNN	0.72	1.72	0.59	2.31	0.59	3.19	0.72	1.69	0.69	1.72
All treatments										
LR	0.23	1.64	0.25	1.60	0.25	1.58	0.25	1.58		
RF	0.27	1.60	0.28	1.56	0.31	1.52	0.31	1.52		
KNN	0.36	1.45	0.38	1.45	0.40	1.40	0.40	1.40		
SVR	0.36	1.45	0.41	1.43	0.40	1.41	0.40	1.41		
DNN	0.15	1.88	0.17	2.25	0.18	2.28	0.18	2.28		

The best prediction (i.e., $R^2 = 0.67$ and $RMSE = 1.62$ Mg/ha) was achieved by the DNN for the AWP treatment at V6 when developed using two variables (i.e., red edge and NIR). For the biochar treatment at V6, the R^2 of DNN increased from 0.57 to 0.71 and $RMSE$ decreased from 1.74 to 1.08 Mg/ha when the number of variables increased from one to two, but the model's performance was reduced when developed using three variables (i.e., $R^2 = 0.67$ and $RMSE = 1.29$ Mg/ha). Similar inferences can be made at the R5 stage across different treatments (Figure 8 and Table 6). For example, the $RMSE$ of DNN for the AWP treatment decreased from 2.03 to 1.17 Mg/ha when developed using five variables. Similarly, for the AWP treatment, the KNN predicted the yield accurately when developed using the top one variable (i.e., GNDVI) with $R^2 = 0.64$ and $RMSE = 1.13$ Mg/ha.

3.4. ML Performance Evaluation in Yield Prediction

The ML model achieved maximum prediction accuracy using the optimal number of variables in comparing their performance for each treatment at both stages (Figure 9 and Table 7). The optimal values of the hyperparameter of the best-performing models are presented in Table 8. Most of the ML models show noticeable differences in yield prediction accuracy. As compared to other ML models, the KNN and SVR produced the most accurate prediction results. The KNN produced the best prediction accuracy for the AWP ($R^2 = 0.69$ and $RMSE = 1.05$ Mg/ha at V6 and $R^2 = 0.64$ and $RMSE = 1.13$ Mg/ha at R5) and gypsum treatments ($R^2 = 0.65$ and $RMSE = 1.75$ Mg/ha at V6 and $R^2 = 0.80$ and $RMSE = 1.35$ Mg/ha at R5). The SVR achieved the best prediction accuracy for the fallow ($R^2 = 0.84$ and $RMSE = 0.69$ Mg/ha at V6 and $R^2 = 0.83$ and $RMSE = 1.05$ Mg/ha at R5) and all treatments ($R^2 = 0.36$ and $RMSE = 1.48$ Mg/ha at V6 and $R^2 = 0.41$ and $RMSE = 1.41$

Mg/ha at R5). The DNN achieved the best prediction accuracy for the biochar ($R^2 = 0.71$ and $RMSE = 1.08$ Mg/ha at V6 and $R^2 = 0.74$ and $RMSE = 1.27$ Mg/ha at R5).

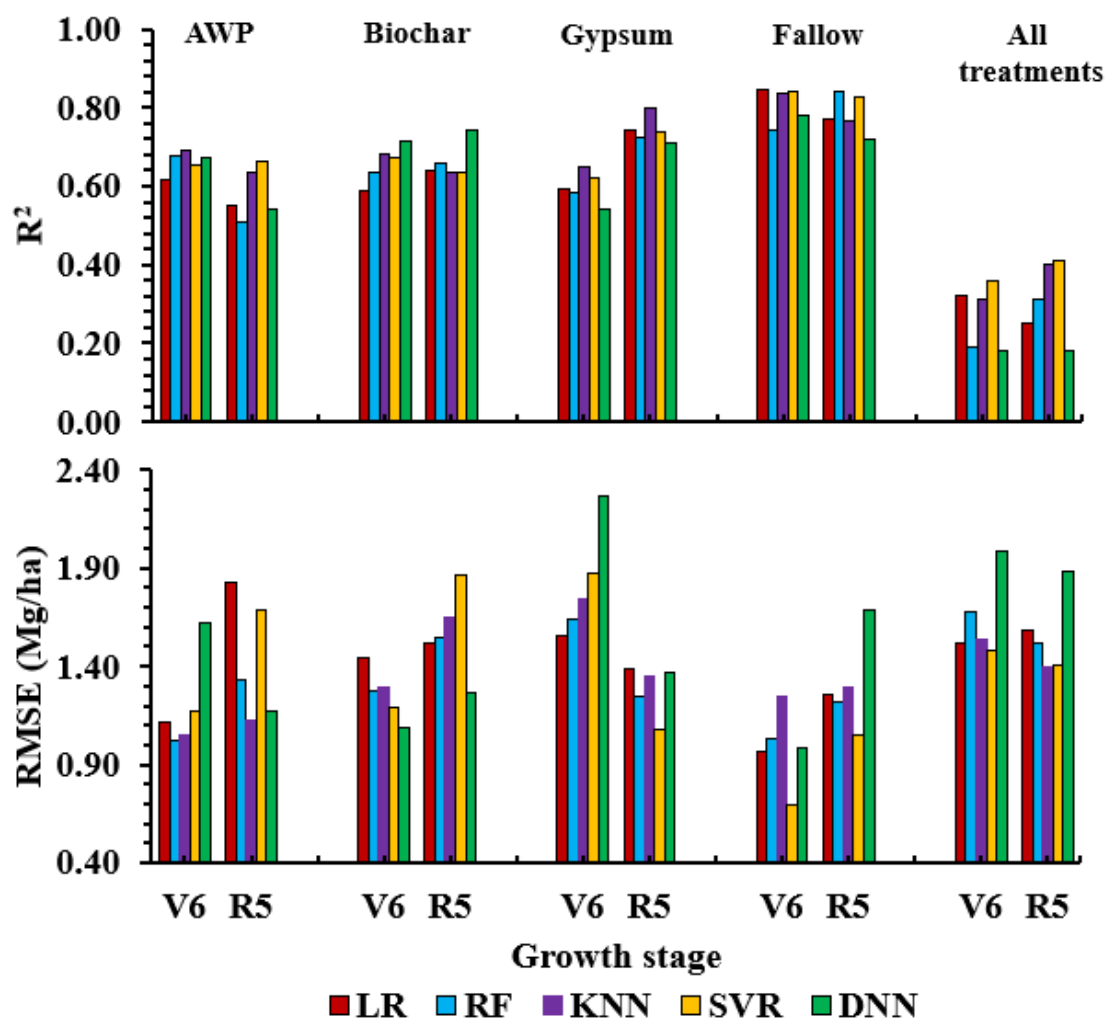


Figure 9. Best-performing ML models for each treatment at the V6 and R5 growth stages.

Table 7. Best-performing ML models for each treatment using optimal variables at the V6 and R5 stages. The bold font indicates the best accuracy achieved by the ML models.

	LR		RF		KNN		SVR		DNN	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Austrian Winter Peas										
V6	0.62	1.12	0.68	1.02	0.69	1.05	0.65	1.17	0.67	1.62
R5	0.55	1.83	0.51	1.33	0.64	1.13	0.66	1.68	0.54	1.17
Biochar										
V6	0.59	1.44	0.63	1.27	0.68	1.29	0.67	1.19	0.71	1.08
R5	0.64	1.52	0.66	1.55	0.64	1.65	0.64	1.87	0.74	1.27
Gypsum										
V6	0.59	1.56	0.58	1.64	0.65	1.75	0.62	1.87	0.54	2.26
R5	0.74	1.39	0.72	1.25	0.80	1.35	0.74	1.08	0.71	1.37
Fallow										
V6	0.84	0.96	0.74	1.03	0.84	1.25	0.84	0.69	0.78	0.99
R5	0.77	1.26	0.84	1.22	0.77	1.29	0.83	1.05	0.72	1.69
All treatments										
V6	0.32	1.52	0.19	1.68	0.31	1.54	0.36	1.48	0.18	1.99
R5	0.25	1.58	0.31	1.52	0.40	1.40	0.41	1.41	0.18	1.88

Table 8. Optimal value of hyperparameters of the best-performing ML models at the V6 and R5 growth stages.

	Best-Performing ML Model (V6 and R5)	Optimal Hyperparameter Value (V6 and R5)		Grid Search Space (V6 and R5)
AWP	KNN	K = 5	K = 13	K = 1, 3, 5, 7, 9, 11, 13, 15
Gypsum	KNN	K = 5	K = 3	
Biochar	DNN	Layer1, layer2, layer3 = 7	Layer1, layer2, layer3 = 11	Layer1, layer2, and layer3 = 1 to 15.
Fallow	SVR	C = 10 Sigma = 0.412	C = 30 Sigma = 0.711	C = 1, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100. Sigma (V6) = 0.412, 0.917, 3.185. Sigma (R5) = 0.189, 0.711, 8.714.
All treatments	SVR	C = 1 Sigma = 0.148	C = 5 Sigma = 0.387	C = 1, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100. Sigma (V6) = 0.148, 1.855, 111.72. Sigma (R5) = 0.120, 0.387, 1.878.

The RF and DNN showed more variability in their performance as compared to SVM and KNN with respect to treatments. The LR produced equally good accuracy compared with the other models on a few occasions. For example, for the fallow treatment at the V6 stage, the LR produced a similar prediction accuracy as obtained by the best-performing model (i.e., SVR). On a few occasions, the performance of RF and DNN was even poorer than that of LR. For the gypsum treatment at the V6 stage, the RF and DNN produced even worse prediction accuracies than LR.

4. Discussion

Agronomic treatments (i.e., AWP cover crop, biochar, and gypsum) applied from November 2020 to April 2021 in this study did not increase corn yields compared to the fallow treatment after only a year of treatment application (Figure 4 and Table 3). This was attributed to the fact that AWP used nitrogen for growth and the subsequent nitrogen release through mineralization to meet the ensuing corn crop demands may not have synchronized with corn needs in the first year of corn production [76]. Reduced yields due to the use of cover crops and resultant lower availability of nitrogen have been reported elsewhere [77]. Biochar's high carbon-to-nitrogen ratio results in nitrogen immobilization, reducing the available inorganic nitrogen for the corn crop [78,79]. Leaving the soil bare during the winter fallow period makes the soil vulnerable to erosion, with potential nutrient losses to the environment through surface runoff. The use of gypsum as a soil amendment was meant to mitigate phosphorus losses through enhanced soil aggregation and water infiltration, thus reducing surface runoff [80] and enhancing phosphorus use efficiency. The use of gypsum showed comparable corn yields to traditional fallow practice in the first year of use.

The first and critical step for in-season crop decision-making is yield prediction [81]. VIs are sensitive to photosynthesis and overall crop health, which makes them suitable input variables in crop yield prediction [17,24,31,34,82–84]. Historically, sensor-based decision systems have relied on simple mechanistic frameworks using single variable information, limiting their application [85]. However, we used thirty variables consisting of four spectral bands and twenty-six VIs to explore their effectiveness in explaining corn yield variability for each treatment individually and combined at the V6 and R5 growth stages. Our results indicated that different VIs were suitable for explaining yield variability for each treatment at both stages (Table 4). Like AWP, all treatment data (i.e., the combination of each treatment) also indicated a lower correlation with yield. The spatial yield variability was well explained by VIs for fallow treatment as compared to other treatments at both growth stages. Most of the previous studies have noted a strong relationship between VIs incorporating red edge wavelength, and crop yield as growth stages progress [86]. The red

edge, CCCI, RECI, CARI, GNDVI, green, and CVI are among the most suitable variables in explaining the corn yield variability in this study. The suitable variables found in this study were also found to be suitable in explaining crop yield variability in previous studies [87]. For example, Li et al. [61] found the red-edge-based Vis, particularly CCCI, to be effective in estimating summer corn plant N concentration and uptake across different growth stages.

The influence of the number of variables on ML performance is vital in obtaining an accurate and less complex model [44]. A model developed using relatively fewer variables offers a better interpretability of the decision made and is likely to reduce the risk of overfitting. In this study, we assessed the impact of the number of selected variables on model performance (Figures 7 and 8) and developed the best possible yield prediction ML model for each treatment at both growth stages (Figure 9 and Table 7). Most of the ML models achieved best-performing results using the top one or two variables that indicated a higher correlation with yield.

The performance of any ML model highly depends on its input variables. ML models indicated different levels of prediction accuracy based on different treatments applied during the non-growing corn season (Figures 7 and 8). The differences in ML performance of yield prediction for different treatments are linked with their input variables and correlation with yield. For example, the yield was most accurately predicted for fallow using all ML models compared to the other treatments as its variables indicated a higher correlation with yield. The yield prediction model that grouped AWP, biochar, and gypsum treatments, and the long-term fallow practice, resulted in lower prediction accuracy with $R^2 = 0.36$ and $RMSE = 1.48$ Mg/ha at V6 and $R^2 = 0.41$ and $RMSE = 1.41$ Mg/ha at R5. However, yield prediction models that handled the different treatments in isolation significantly improved the predictive accuracy with maximum R^2 values ranging from 0.64 to 0.83 and RMSE values from 1.05 to 1.75 Mg/ha for both the V6 and R5 growth stages. The relatively lower R^2 values for AWP and biochar treatments were attributed to the treatment's effects on the use of inorganic nitrogen by AWP for growth and nitrogen immobilization by biochar lowering the available nitrogen for the corn crop, bringing random variation since this was the first year of treatment. This research showed that corn yields could be predicted with similar precision at both vegetative (V6) and reproductive (R5) stages and that farmer practice should be considered when implementing predictive models for reliable outcomes.

The comparative analysis of ML models with their optimal configuration demonstrated that KNN and SVR outperform other models at both growth stages (Figure 9 and Table 6). Unlike this study, Mupangwa et al. [25] found poor corn yield prediction from the SVR but achieved accurate results from KNN. Matsumura et al. [51] achieved the best prediction accuracy using neural networks but our study found a higher RMSE from neural networks, indicating overfitting, possibly due to the limited number of samples. The prediction accuracy of RF in this study was not good when obtained from SVR and KNN, except on a few occasions, such as for AWP at V6 and fallow at R5. The LR produced a similar prediction accuracy where input variables indicated a higher linear correlation with yield, as expected. For example, the prediction accuracy of LR for the fallow treatment is approximately as good as the best-performing ML models (i.e., KNN and SVR), but performance decreases where the input variables do not indicate a higher linear correlation with yield as seen in AWP and all treatments at both growth stages. On the other hand, other ML models produce a higher prediction accuracy than LR even if there is a low to moderate correlation with yield due to their better ability to learn the complex relationship between dependent and independent variables. Furthermore, the experimental results demonstrated that selected VIs can be efficiently used in corn yield prediction across different growth stages (i.e., V6 and R5) with reasonably good accuracy (Figure 9 and Table 7).

This study demonstrated the utility of UAV-based VIs and ML models for predicting corn yield with reasonably good accuracy under the constraint of a limited number of training data and a lack of knowledge of other variables like soil and weather. Obtaining larger training data is practically challenging for small agricultural plots. Studies should also focus on assessing how early corn yield can be predicted with adequate accuracy

to support effective crop management practices and decision-making. Furthermore, the integration of UAV-based optical and thermal sensors along with ground proximal sensing data can potentially improve the prediction results and facilitate a more thorough ground-based validation of crop yield prediction. Future studies can also explore data augmentation approaches to derive synthesis training data and their subsequent utilization in developing different ML and deep learning models for accurate yield prediction and forecasting.

5. Conclusions

This study demonstrated the potential application of UAV-derived Vegetation Index (VIs) coupled with state-of-the-art Machine Learning (ML) models in corn yield prediction with a limited number of training samples at the farm scale. Spectral bands and VIs including red edge, canopy chlorophyll content index, red edge chlorophyll index, chlorophyll absorption ratio index, green normalized difference vegetation index, green spectral band, and chlorophyll vegetation index indicated moderate to good correlation with the yield at the V6 and R5 growth stages, indicating their suitability for use in predicting yield using ML models.

Support Vector Regression (SVR) and k-Nearest Neighbor (KNN) outperformed the other models such as the Linear Regression (LR), Random Forest (RF), and Deep Neural Network (DNN) models. The performance of DNN was found to be more sensitive to the number of variables used in its development and often produced a higher RMSE than other models, mainly due to the limited amount of training data. This study suggests the utilization of simpler models like LR or KNN, particularly when input variables indicate a strong linear relationship with the dependent variable as these models also offer a better model interpretability than the black-box nature of ML.

Corn yield prediction models under different agronomic practices showed higher accuracy when they were management specific rather than combined. This study highlighted the importance of considering agronomic practices and farming history to enhance models' predictive performance. This research also showed that corn yields could be predicted with adequate accuracy at both vegetative and reproductive stages, which is vital in time-sensitive decision-making. Thus, this study confirmed that UAV-derived VIs in conjunction with ML models can produce an adequate corn yield prediction even with a limited number of training samples and could be effectively used in better-informed crop management.

Author Contributions: Conceptualization, C.K. and Y.H.; methodology, C.K.; software, C.K.; validation, C.K.; formal analysis, C.K.; investigation, C.K., P.M. and Y.H.; resources, J.D. and K.R.; data curation, Y.H. and P.M.; writing—original draft preparation, C.K.; writing—review and editing, C.K., P.M., Y.H., J.D. and K.R.; visualization, C.K.; supervision, Y.H., J.D. and K.R.; project administration, K.R.; funding acquisition, K.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the U. S. Department of Agriculture-Agriculture Research Service (USDA-ARS), Crop Production Systems Research Unit through project number 6066-22000-089-000D and through USDA-ARS and Mississippi State University non-assistance cooperative agreement number 58-6066-1-020.

Data Availability Statement: The data used in this manuscript are available upon reasonable request to the corresponding author. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

Acknowledgments: We would like to express our sincere thanks to the anonymous reviewers for their constructive comments and suggestions to improve the original version of the manuscript.

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

References

- Shiferaw, B.; Prasanna, B.M.; Hellin, J.; Bänziger, M. Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Secur.* **2011**, *3*, 307–327. [\[CrossRef\]](#)
- McConnell, M. *Feedgrains Sector at a Glance*; USDA Economic Research Service US Department of Agriculture: Washington, DC, USA, 2022.
- World Health Organization. *The State of Food Security and Nutrition in the World 2021: Transforming Food Systems for Food Security, Improved Nutrition and Affordable Healthy Diets for All*; Food & Agriculture Organization: Rome, Italy, 2021; Volume 2021.
- Sishodia, R.P.; Ray, R.L.; Singh, S.K. Applications of remote sensing in precision agriculture: A review. *Remote Sens.* **2020**, *12*, 3136. [\[CrossRef\]](#)
- Kamath, P.; Patil, P.; Shrilatha, S.; Sowmya, S. Crop yield forecasting using data mining. *Glob. Transit. Proc.* **2021**, *2*, 402–407. [\[CrossRef\]](#)
- Surya, P.; Aroquiaraj, I.L. Crop yield prediction in agriculture using data mining predictive analytic techniques. *Int. J. Res. Anal. Rev.* **2018**, *5*, 783–787.
- Bala, A. Machine Learning Approaches for Crop Yield Prediction-Review. *Int. J. Comput. Eng. Technol.* **2020**, *11*, 23–27.
- Sagar, B.; Cauvery, N. Agriculture data analytics in crop yield estimation: A critical review. *Indones. J. Electr. Eng. Comput. Sci.* **2018**, *12*, 1087–1093. [\[CrossRef\]](#)
- Wójtowicz, M.; Wójtowicz, A.; Piekarczyk, J. Application of remote sensing methods in agriculture. *Commun. Biometry Crop Sci.* **2016**, *11*, 31–50.
- Ali, A.M.; Abouelghar, M.; Belal, A.A.; Saleh, N.; Yones, M.; Selim, A.I.; Amin, M.E.S.; Elwesemy, A.; Kucher, D.E.; Maginan, S.; et al. Crop Yield Prediction Using Multi Sensors Remote Sensing (Review Article). *Egypt. J. Remote Sens. Space Sci.* **2022**, *25*, 711–716. [\[CrossRef\]](#)
- Yang, C. A high-resolution airborne four-camera imaging system for agricultural remote sensing. *Comput. Electron. Agric.* **2012**, *88*, 13–24. [\[CrossRef\]](#)
- Tsouros, D.C.; Bibi, S.; Sarigiannidis, P.G. A review on UAV-based applications for precision agriculture. *Information* **2019**, *10*, 349. [\[CrossRef\]](#)
- Wang, T.; Xu, X.; Wang, C.; Li, Z.; Li, D. From smart farming towards unmanned farms: A new mode of agricultural production. *Agriculture* **2021**, *11*, 145. [\[CrossRef\]](#)
- Ji, Z.; Pan, Y.; Zhu, X.; Zhang, D.; Dai, J. Prediction of Corn Yield in the USA Corn Belt Using Satellite Data and Machine Learning: From an Evapotranspiration Perspective. *Agriculture* **2022**, *12*, 1263. [\[CrossRef\]](#)
- Maes, W.H.; Steppe, K. Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. *Trends Plant Sci.* **2019**, *24*, 152–164. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rokhmana, C.A. The potential of UAV-based remote sensing for supporting precision agriculture in Indonesia. *Procedia Environ. Sci.* **2015**, *24*, 245–253. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Hussain, N.; Sarfraz, S.; Javed, S. A Systematic Review on Crop-Yield Prediction through Unmanned Aerial Vehicles. In Proceedings of the 2021 16th International Conference on Emerging Technologies (ICET), Islamabad, Pakistan, 22–23 December 2021; pp. 1–9.
- Huang, Y.; Thomson, S.J.; Hoffmann, W.C.; Lan, Y.; Fritz, B.K. Development and prospect of unmanned aerial vehicle technologies for agricultural production management. *Int. J. Agric. Biol. Eng.* **2013**, *6*, 1–10.
- Yang, G.; Liu, J.; Zhao, C.; Li, Z.; Huang, Y.; Yu, H.; Xu, B.; Yang, X.; Zhu, D.; Zhang, X. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Front. Plant Sci.* **2017**, *8*, 1111. [\[CrossRef\]](#)
- Ma, Y.; Zhang, Z.; Kang, Y.; Özdoğan, M. Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. *Remote Sens. Environ.* **2021**, *259*, 112408. [\[CrossRef\]](#)
- Shahhosseini, M.; Hu, G.; Khaki, S.; Archontoulis, S.V. Corn yield prediction with ensemble CNN-DNN. *Front. Plant Sci.* **2021**, *12*, 709008. [\[CrossRef\]](#)
- Buthelezi, S.; Mutanga, O.; Sibanda, M.; Odindi, J.; Clulow, A.D.; Chimonyo, V.G.; Mabhaudhi, T. Assessing the prospects of remote sensing maize leaf area index using UAV-derived multi-spectral data in smallholder farms across the growing season. *Remote Sens.* **2023**, *15*, 1597. [\[CrossRef\]](#)
- Pipatsitee, P.; Tisarum, R.; Taota, K.; Samphumphuang, T.; Eiumnoh, A.; Singh, H.P.; Cha-Um, S. Effectiveness of vegetation indices and UAV-multispectral imageries in assessing the response of hybrid maize (*Zea mays* L.) to water deficit stress under field environment. *Environ. Monit Assess* **2022**, *195*, 128. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mupangwa, W.; Chipindu, L.; Nyagumbo, I.; Mkuhlani, S.; Sisito, G. Evaluating machine learning algorithms for predicting maize yield under conservation agriculture in Eastern and Southern Africa. *SN Appl. Sci.* **2020**, *2*, 952. [\[CrossRef\]](#)
- Shahhosseini, M.; Martinez-Feria, R.A.; Hu, G.; Archontoulis, S.V. Maize yield and nitrate loss prediction with machine learning algorithms. *Environ. Res. Lett.* **2019**, *14*, 124026. [\[CrossRef\]](#)
- Danilevich, M.F.; Bayer, P.E.; Boussaid, F.; Bennamoun, M.; Edwards, D. Maize yield prediction at an early developmental stage using multispectral images and genotype data for preliminary hybrid selection. *Remote Sens.* **2021**, *13*, 3976. [\[CrossRef\]](#)

28. Dhaka, V.; Vikas, L. Wheat yield prediction using artificial neural network and crop prediction techniques (a survey). *IJRASET* **2014**, *2*, 330–341.
29. Tanabe, R.; Matsui, T.; Tanaka, T.S.T. Winter wheat yield prediction using convolutional neural networks and UAV-based multispectral imagery. *Field Crops Res.* **2023**, *291*, 108786. [[CrossRef](#)]
30. Bascon, M.V.; Nakata, T.; Shibata, S.; Takata, I.; Kobayashi, N.; Kato, Y.; Inoue, S.; Doi, K.; Murase, J.; Nishiuchi, S. Estimating Yield-Related Traits Using UAV-Derived Multispectral Images to Improve Rice Grain Yield Prediction. *Agriculture* **2022**, *12*, 1141. [[CrossRef](#)]
31. 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](#)]
32. Zhou, X.; Zheng, H.; Xu, X.; He, J.; Ge, X.; Yao, X.; Cheng, T.; Zhu, Y.; Cao, W.; Tian, Y. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 246–255. [[CrossRef](#)]
33. Tian, L.; Wang, C.; Li, H.; Sun, H. Yield prediction model of rice and wheat crops based on ecological distance algorithm. *Environ. Technol. Innov.* **2020**, *20*, 101132. [[CrossRef](#)]
34. 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](#)]
35. Siegfried, J.; Adams, C.B.; Rajan, N.; Hague, S.; Schnell, R.; Hardin, R. Combining a cotton ‘Boll Area Index’ with in-season unmanned aerial multispectral and thermal imagery for yield estimation. *Field Crops Res.* **2023**, *291*, 108765. [[CrossRef](#)]
36. Ashapure, A.; Jung, J.; Chang, A.; Oh, S.; Yeom, J.; Maeda, M.; Maeda, A.; Dube, N.; Landivar, J.; Hague, S.; et al. Developing a machine learning based cotton yield estimation framework using multi-temporal UAS data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 180–194. [[CrossRef](#)]
37. He, L.; Fang, W.; Zhao, G.; Wu, Z.; Fu, L.; Li, R.; Majeed, Y.; Dhupia, J. Fruit yield prediction and estimation in orchards: A state-of-the-art comprehensive review for both direct and indirect methods. *Comput. Electron. Agric.* **2022**, *195*, 106812. [[CrossRef](#)]
38. de Oliveira, R.P.; Barbosa Júnior, M.R.; Pinto, A.A.; Oliveira, J.L.P.; Zerbato, C.; Furlani, C.E.A. Predicting Sugarcane Biometric Parameters by UAV Multispectral Images and Machine Learning. *Agronomy* **2022**, *12*, 1992. [[CrossRef](#)]
39. Akbarian, S.; Xu, C.; Wang, W.; Ginns, S.; Lim, S. Sugarcane yields prediction at the row level using a novel cross-validation approach to multi-year multispectral images. *Comput. Electron. Agric.* **2022**, *198*, 107024. [[CrossRef](#)]
40. Mitchell, T.M. *Machine Learning*; McGraw-Hill: New York, NY, USA, 2007; Volume 1.
41. Mjolsness, E.; DeCoste, D. Machine learning for science: State of the art and future prospects. *Science* **2001**, *293*, 2051–2055. [[CrossRef](#)] [[PubMed](#)]
42. Kumar, C.; Chatterjee, S.; Oommen, T.; Guha, A. Automated lithological mapping by integrating spectral enhancement techniques and machine learning algorithms using AVIRIS-NG hyperspectral data in Gold-bearing granite-greenstone rocks in Hutti, India. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *86*, 102006. [[CrossRef](#)]
43. Kumar, C.; Chatterjee, S.; Oommen, T.; Guha, A.; Mukherjee, A. Multi-sensor datasets-based optimal integration of spectral, textural, and morphological characteristics of rocks for lithological classification using machine learning models. *Geocarto Int.* **2022**, *37*, 6004–6032. [[CrossRef](#)]
44. Kumar, C.; Walton, G.; Santi, P.; Luza, C. An Ensemble Approach of Feature Selection and Machine Learning Models for Regional Landslide Susceptibility Mapping in the Arid Mountainous Terrain of Southern Peru. *Remote Sens.* **2023**, *15*, 1376. [[CrossRef](#)]
45. Bhatt, P.; Maclean, A.; Dickinson, Y.; Kumar, C. Fine-Scale Mapping of Natural Ecological Communities Using Machine Learning Approaches. *Remote Sens.* **2022**, *14*, 563. [[CrossRef](#)]
46. Santana, D.C.; Teodoro, L.P.R.; Baio, F.H.R.; Santos, R.G.d.; Coradi, P.C.; Biduski, B.; da Silva Junior, C.A.; Teodoro, P.E.; Shiratsuchi, L.S. Classification of soybean genotypes for industrial traits using UAV multispectral imagery and machine learning. *Remote Sens. Appl. Soc. Environ.* **2023**, *29*, 100919. [[CrossRef](#)]
47. Rashid, M.; Bari, B.S.; Yusup, Y.; Kamaruddin, M.A.; Khan, N. A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. *IEEE Access* **2021**, *9*, 63406–63439. [[CrossRef](#)]
48. Shahhosseini, M.; Hu, G.; Huber, I.; Archontoulis, S.V. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Sci. Rep.* **2021**, *11*, 1606. [[CrossRef](#)]
49. Van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* **2020**, *177*, 105709. [[CrossRef](#)]
50. Croci, M.; Impollonia, G.; Meroni, M.; Amaducci, S. Dynamic Maize Yield Predictions Using Machine Learning on Multi-Source Data. *Remote Sens.* **2022**, *15*, 100. [[CrossRef](#)]
51. Matsumura, K.; Gaitan, C.F.; Sugimoto, K.; Cannon, A.J.; Hsieh, W.W. Maize yield forecasting by linear regression and artificial neural networks in Jilin, China. *J. Agric. Sci.* **2015**, *153*, 399–410. [[CrossRef](#)]
52. Kim, N.; Lee, Y.-W. Machine learning approaches to corn yield estimation using satellite images and climate data: A case of Iowa State. *J. Korean Soc. Surv. Geod. Photogramm. Cartogr.* **2016**, *34*, 383–390. [[CrossRef](#)]
53. Shen, Y.; Mercatoris, B.; Cao, Z.; Kwan, P.; Guo, L.; Yao, H.; Cheng, Q. Improving Wheat Yield Prediction Accuracy Using LSTM-RF Framework Based on UAV Thermal Infrared and Multispectral Imagery. *Agriculture* **2022**, *12*, 892. [[CrossRef](#)]

54. Liebman, A.M.; Grossman, J.; Brown, M.; Wells, M.S.; Reberg-Horton, S.; Shi, W. Legume cover crops and tillage impact nitrogen dynamics in organic corn production. *Agron. J.* **2018**, *110*, 1046–1057. [\[CrossRef\]](#)
55. DeLaune, P.; Mubvumba, P. Winter cover crop production and water use in Southern Great Plains cotton. *Agron. J.* **2020**, *112*, 1943–1951. [\[CrossRef\]](#)
56. Syuhada, A.B.; Shamshuddin, J.; Fauziah, C.; Rosenani, A.; Arifin, A. Biochar as soil amendment: Impact on chemical properties and corn nutrient uptake in a Podzol. *Can. J. Soil Sci.* **2016**, *96*, 400–412. [\[CrossRef\]](#)
57. Rogovska, N.; Laird, D.A.; Karlen, D.L. Corn and soil response to biochar application and stover harvest. *Field Crops Res.* **2016**, *187*, 96–106. [\[CrossRef\]](#)
58. Prakash, N.B.; Dhumgond, P.; Shruthi; Chikkaramappa, T.; Ashrit, S. Performance of slag-based gypsum on maize yield and available soil nutrients over commercial gypsum under acidic and neutral soil. *Commun. Soil Sci. Plant Anal.* **2020**, *51*, 1780–1798. [\[CrossRef\]](#)
59. Bossolani, J.W.; Crusciol, C.A.C.; Merloti, L.F.; Moretti, L.G.; Costa, N.R.; Tsai, S.M.; Kuramae, E.E. Long-term lime and gypsum amendment increase nitrogen fixation and decrease nitrification and denitrification gene abundances in the rhizosphere and soil in a tropical no-till intercropping system. *Geoderma* **2020**, *375*, 114476. [\[CrossRef\]](#)
60. 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. [\[CrossRef\]](#)
61. 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. [\[CrossRef\]](#)
62. 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. [\[CrossRef\]](#)
63. 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. [\[CrossRef\]](#)
64. San Bautista, A.; Fita, D.; Franch, B.; Castiñeira-Ibáñez, S.; Arizo, P.; Sánchez-Torres, M.J.; Becker-Reshef, I.; Uris, A.; Rubio, C. Crop Monitoring Strategy Based on Remote Sensing Data (Sentinel-2 and Planet), Study Case in a Rice Field after Applying Glycinebetaine. *Agriculture* **2022**, *12*, 708. [\[CrossRef\]](#)
65. Hassan, M.A.; Yang, M.; Rasheed, A.; Jin, X.; Xia, X.; Xiao, Y.; He, Z. Time-series multispectral indices from unmanned aerial vehicle imagery reveal senescence rate in bread wheat. *Remote Sens.* **2018**, *10*, 809. [\[CrossRef\]](#)
66. Qiao, L.; Tang, W.; Gao, D.; Zhao, R.; An, L.; Li, M.; Sun, H.; Song, D. UAV-based chlorophyll content estimation by evaluating vegetation index responses under different crop coverages. *Comput. Electron. Agric.* **2022**, *196*, 106775. [\[CrossRef\]](#)
67. Kuhn, M.; Wing, J.; Weston, S.; Williams, A.; Keefer, C.; Engelhardt, A.; Cooper, T.; Mayer, Z.; Kenkel, B.; R Core Team. Package ‘caret’. *R J.* **2020**, *223*, 1–224.
68. Ihaka, R.; Gentleman, R. R: A language for data analysis and graphics. *J. Comput. Graph. Stat.* **1996**, *5*, 299–314.
69. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2021.
70. Alpaydin, E. *Introduction to Machine Learning*; MIT Press: Cambridge, MA, USA, 2020.
71. Wong, T.-T.; Yeh, P.-Y. Reliable accuracy estimates from k-fold cross validation. *IEEE Trans. Knowl. Data Eng.* **2019**, *32*, 1586–1594. [\[CrossRef\]](#)
72. Fushiki, T. Estimation of prediction error by using K-fold cross-validation. *Stat. Comput.* **2011**, *21*, 137–146. [\[CrossRef\]](#)
73. Wong, T.-T.; Yang, N.-Y. Dependency analysis of accuracy estimates in k-fold cross validation. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 2417–2427. [\[CrossRef\]](#)
74. Shrestha, N. Detecting multicollinearity in regression analysis. *Am. J. Appl. Math. Stat.* **2020**, *8*, 39–42.
75. Kim, J.H. Multicollinearity and misleading statistical results. *Korean J. Anesthesiol.* **2019**, *72*, 558–569. [\[CrossRef\]](#)
76. Jani, A.D.; Grossman, J.; Smyth, T.J.; Hu, S. Winter legume cover-crop root decomposition and N release dynamics under disking and roller-crimping termination approaches. *Renew. Agric. Food Syst.* **2016**, *31*, 214–229. [\[CrossRef\]](#)
77. Parr, M.; Grossman, J.; Reberg-Horton, S.; Brinton, C.; Crozier, C. Nitrogen delivery from legume cover crops in no-till organic corn production. *Agron. J.* **2011**, *103*, 1578–1590. [\[CrossRef\]](#)
78. Bruun, E.W.; Ambus, P.; Egsgaard, H.; Hauggaard-Nielsen, H. Effects of slow and fast pyrolysis biochar on soil C and N turnover dynamics. *Soil Biol. Biochem.* **2012**, *46*, 73–79. [\[CrossRef\]](#)
79. Nelissen, V.; Rütting, T.; Huygens, D.; Staelens, J.; Ruyschaert, G.; Boeckx, P. Maize biochars accelerate short-term soil nitrogen dynamics in a loamy sand soil. *Soil Biol. Biochem.* **2012**, *55*, 20–27. [\[CrossRef\]](#)
80. Kaur, H.; Williard, K.W.; Schoonover, J.E.; Singh, G. Impact of Flue Gas Desulfurization Gypsum Applications to Corn-Soybean Plots on Surface Runoff Water Quality. *Water Air Soil Pollut.* **2022**, *233*, 72. [\[CrossRef\]](#)
81. Dhillon, J.; Aula, L.; Eickhoff, E.; Raun, W. Predicting in-season maize (*Zea mays* L.) yield potential using crop sensors and climatological data. *Sci. Rep.* **2020**, *10*, 11479. [\[CrossRef\]](#)
82. Mwinuka, P.R.; Mourice, S.K.; Mbungu, W.B.; Mbilinyi, B.P.; Tumbo, S.D.; Schmitter, P. UAV-based multispectral vegetation indices for assessing the interactive effects of water and nitrogen in irrigated horticultural crops production under tropical sub-humid conditions: A case of African eggplant. *Agric. Water Manag.* **2022**, *266*, 107516. [\[CrossRef\]](#)
83. Nevavuori, P.; Narra, N.; Linna, P.; Lipping, T. Crop yield prediction using multitemporal UAV data and spatio-temporal deep learning models. *Remote Sens.* **2020**, *12*, 4000. [\[CrossRef\]](#)

84. Zhou, X.; Kono, Y.; Win, A.; Matsui, T.; Tanaka, T.S.T. Predicting within-field variability in grain yield and protein content of winter wheat using UAV-based multispectral imagery and machine learning approaches. *Plant Prod. Sci.* **2020**, *24*, 137–151. [[CrossRef](#)]
85. Colaço, A.; Richetti, J.; Bramley, R.; Lawes, R. How will the next-generation of sensor-based decision systems look in the context of intelligent agriculture? A case-study. *Field Crops Res.* **2021**, *270*, 108205. [[CrossRef](#)]
86. Oglesby, C.; Fox, A.A.; Singh, G.; Dhillon, J. Predicting In-Season Corn Grain Yield Using Optical Sensors. *Agronomy* **2022**, *12*, 2402. [[CrossRef](#)]
87. Sumner, Z.; Varco, J.J.; Dhillon, J.S.; Fox, A.A.; Czarnecki, J.; Henry, W.B. Ground versus aerial canopy reflectance of corn: Red-edge and non-red edge vegetation indices. *Agron. J.* **2021**, *113*, 2782–2797. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.