

Article

Impact Assessment of Nematode Infestation on Soybean Crop Production Using Aerial Multispectral Imagery and Machine Learning

Pius Jjagwe ^{1,2}, Abhilash K. Chandel ^{1,2,*} and David B. Langston ¹

¹ Virginia Tech Tidewater Agricultural Research and Extension Center, Suffolk, VA 23437, USA; pjjagwe@vt.edu (P.J.); dlangston@vt.edu (D.B.L.)

² Department of Biological Systems Engineering, Virginia Tech, Blacksburg, VA 24061, USA

* Correspondence: abhilashchandel@vt.edu

Abstract: Accurate and prompt estimation of geospatial soybean yield (SY) is critical for the producers to determine key factors influencing crop growth for improved precision management decisions. This study aims to quantify the impacts of soybean cyst nematode (SCN) infestation on soybean production and the yield of susceptible and resistant seed varieties. Susceptible varieties showed lower yield and crop vigor recovery, and high SCN population (20 to 1080) compared to resistant varieties (SCN populations: 0 to 340). High-resolution (1.3 cm/pixel) aerial multispectral imagery showed the blue band reflectance ($r = 0.58$) and Green Normalized Difference Vegetation Index (GNDVI, $r = -0.6$) have the best correlation with the SCN populations. While GDNVI, Green Chlorophyll Index (GCI), and Normalized Difference Red Edge Index (NDRE) were the best differentiators of plant vigor and had the highest correlation with SY ($r = 0.59$ – 0.75). Reflectance (REF) and VIs were then used for SY estimation using two statistical and four machine learning (ML) models at 10 different train–test data split ratios (50:50–95:5). The ML models and train–test data split ratio had significant impacts on SY estimation accuracy. Random forest (RF) was the best and consistently performing model ($r: 0.84$ – 0.97 , $rRMSE: 8.72$ – 20%), while a higher train–test split ratio lowered the performances of the ML models. The 95:5 train–test ratio showed the best performance across all the models, which may be a suitable ratio for modeling over smaller or medium-sized datasets. Such insights derived using high spatial resolution data can be utilized to implement precision crop protective operations for enhanced soybean yield and productivity.

Keywords: SUAS multispectral sensing; machine learning; soybean yield; vegetation index; precision management



Citation: Jjagwe, P.; Chandel, A.K.; Langston, D.B. Impact Assessment of Nematode Infestation on Soybean Crop Production Using Aerial Multispectral Imagery and Machine Learning. *Appl. Sci.* **2024**, *14*, 5482. <https://doi.org/10.3390/app14135482>

Academic Editors: Samuel Adelabu, Romano Lottering and Kabir Peerbhaya

Received: 26 April 2024

Revised: 6 June 2024

Accepted: 21 June 2024

Published: 24 June 2024



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

1. Introduction

Soybean is one of the most economically significant crops grown worldwide for food, feed, and industrial products. In order to meet the demands of the enormously growing population, improving soybean yield (SY) is one of the primary goals of the breeding programs, globally [1]. Understanding soybean's in-season growth and production is crucial for macro (government and corporate) and micro (farmer) level decision-making which directly impacts crop health and harvest management, crop insurances, food security, supply regulation, the financial market, and strategic planning concerning social, environmental, and economic policies [2,3]. Soybean production can be significantly reduced by various stressors, in particular the soybean cyst nematode (SCN, *Heterodera glycines*) which can yield losses up to 90% [4–6]. Juvenile nematodes initiate the infection by feeding on the cells in the vascular system of soybean roots. After feeding, a male juvenile will move into the soil, and thus does not spread infection as much as a female because females grow in their current location [5]. SCN attacks are characterized by circular patches on the leaves that frequently change color coupled with yellowing of the crop cover [5]. Nematode

infestation is conventionally assessed from visible features of crop response by visual scouting and depends on human experience. This process is laborious, time-consuming, expensive, and most prone to misdiagnosis [6]. In addition, the visible symptoms may often be confused by the observers with other biotic or abiotic stressors including nutrient, water, and other fungal or pest infestations. Destructive laboratory analysis of soil and root samples is also conducted to ascertain the degree of SCN infestation [5]. Pertaining to the soil analysis method, soil samples are collected from the field and brought to the taxonomic/pathology laboratories. The samples are cleaned thoroughly, and specimen slides of the cleaned samples are prepared to be observed under the microscope. The nematodes are then visually-manually counted by trained nematologists to identify the degree of infestation [7]. Although accurate, this method is again laborious, time taking and most importantly, lacks geospatial sampling accuracy. During peak seasons and shortage of taxonomic experts, the lab results to the growers can be delayed which leads into missed opportunities of prompt and precise control of SCN infestations in the field.

Given the challenges associated with conventional impact assessment techniques and rapid phenotyping, remote sensing (using aerial multispectral imagery) is a possible solution for quantifying crop stressors [8,9]. Remote sensing offers a quick and economical way to identify issues that arise during a cropping season to deploy appropriate and timely management practices. For example, Bajwa et al. [10] was able to distinguish plants that were healthy, and the ones infected with nematodes and a soil-borne fungal pathogen using spectral reflectance and vegetation indices (VIs). A correlation between disease rating and VI of (>0.8) was observed using stepwise linear discriminant analysis (LDA) and logistic discriminant analysis (LgDA). Based on the plant spectra, a two-class discriminant model was able to identify 58% of the infested plants and 97% of the healthy plants with a disease. A study by Hillnhutter et al. [6] showed that hyperspectral imaging is useful in detecting and discriminating root disease development (*Heterodera schachtii* and *Rhizoctonia solani*) in sugar beet. The AISA and HyMap data showed a classification accuracy of 72% and 64%, respectively, from a supervised classification using a Spectral Angle Mapper of the organism-induced leaf symptoms.

In addition to SCN management, estimation of crop yield prior to season end is a challenging task that requires significant time investments involving the collection of complicated information of crop genotype, crop phenotypes, environmental factors (weather and soil) as well as management practices, among others [2]. Pre-harvest yield estimation much ahead enables growers to deploy potential management practices such as fertilizers, irrigation, pesticides, and herbicides to achieve a better yield at the season end. Season-end yield is typically recorded using yield monitoring systems on grain combines which have been in use for more than 25 years [11]. Such yield monitoring devices are invasive, expensive to own and operate, demand frequent field calibration, and require data cleaning for retrieving useful information [11,12]. Furthermore, SY is not only influenced by inherent physiological and structural traits but also by environmental factors like soil patterns, climate, and hydrology, making manual monitoring labor-intensive and typically ineffective [13]. Numerous methods for non-invasive estimation of crop yield prior to season end have also been developed which include but are not limited to statistical, agro-meteorological, as well as crop growth models [14]. Typically, crop growth models are built on a regional scale for growth monitoring and yield estimation, feeding on regional weather data or low-resolution satellite remote sensing inputs [15]. These models fail to account for the temporal and spatial variations in meteorological data, crop parameters (LAI, biomass), and soil characteristics. For example, Ma et al. [15] estimated wheat yields using Sentinel-2 remote sensing data, crop growth model of SAFY and the results showed an R^2 of 0.73, 0.83 and 0.49 as well as RMSE of 0.72, 1.13 and 1.14 t/ha for correlation between vegetation indices, leaf area index (LAI), biomass, and yield. The MODIS satellite imagery-derived LAI was integrated into the BEPS crop growth model, which improved the accuracy of corn yield estimation [15]. Specific to soybeans, CROPGRO and SOYGRO models have been used to estimate biomass and SY using LAI and VIs and inputs [16]. Ade-

boye et al. [17] used the AQUACROP model to estimate soybean seed yield with the model providing a satisfactory performance ($R^2 = 0.99$ and low RMSE of 0.10 t ha^{-1}). Inputs for the AQUACROP model were environmental conditions (air temperature, relative humidity, rainfall among others) and crop parameters such as dry aboveground biomass, root length and canopy cover. Most crop growth models are designed for regional scale estimation, limiting spatiotemporal variability assessments [15]. As discussed earlier, techniques that can quantify crop growth and yield at high-throughput scale but are also able to provide spatiotemporal variations at high resolution are needed [15].

Agricultural productivity has been estimated using a range of remote sensing data inputs in various data-driven models [1,6]. These inputs come from the ground, small unmanned aircraft systems (SUAS), and high-altitude (satellite, aerospace) platforms. High altitude remote sensing, like satellites, are appropriate for large scale operations but are prone to be affected by weather and offer low resolution. Ground-based remote sensing (analytical spectral devices) offers high resolution and are easy to use but are labor-intensive, time-consuming, and have a low operation efficiency for monitoring a larger region [18]. SUAS-based remote sensing applications have rapidly advanced for agricultural systems due to their ability to overcome the limitations of satellites and ground remote sensing. SUAS platforms can be custom integrated with a range of sensors, including digital, multispectral, hyperspectral, thermal infrared, and Lidar sensing and imaging devices for crop characterization at a very broad spectrum [2]. SUAS imaging systems have been successfully used for numerous mapping and phenotyping applications [13] drought resistance [19], crop growth [2], yield [12], and grain moisture estimation [20], among others. Such applications have demonstrated significant improvement in yield estimations for crops including corn [12], wheat [21], and cotton [22], among others. For example, studies by Maimaitijiang et al. [23], Yosefzadeh-Najafabadi et al. [1], and Wu et al. [15] noted relative accuracies in the form of $R^2 = 0.72\text{--}0.77$ and $\text{rRMSE}\% = 15.9\%$, using aerial multi-modal spectral data through deep neural networks and machine learning. Recently, with the advancements in precision agriculture, large aerial imagery datasets have been processed to model crop traits [2] using supervised machine learning (ML) models [24–26] such as random forest (RF), support vector machine (SVM), k-nearest neighbor (KNN), and artificial neural network (ANN), among others [1,2,26–28]. These ML models could often be affected in their performance due to dataset sizes, requirement of substantial computational resources and time to handle large datasets [23]. For example, Pinto et al. [27] combined satellite remote sensing and ML models to forecast corn grain yield. Similarly, Fei et al. [28] utilized remote sensing data and ensemble learning algorithms to estimate wheat yield using VIs. Sakamoto [29] fused MODIS imagery-derived VIs and environmental variables using RF to estimate SY and obtained errors of 0.206 t/ha [30]. ML models are capable of capturing key complexities and distributions of data to reduce the effects of an unknown variability, and they thereby tend to become robust.

While studies have utilized SUAS remote sensing and ML to estimate crop traits and yield [31], limited studies exist which have quantified the impact of SCN on SY and crop physiology. Moreover, evaluating the impact of data size on such modeling is a critical factor to study. This gap is being addressed in this study by evaluating the impact of SCN on soybean yield and physiology followed by evaluating ML models for SY estimation under varied dataset sizes. The specific objectives are to (1) evaluate the impact of SCN infestation on crop vigor and yield, (2) quantify the impact of SCN infestation on soybean physiology through aerial multispectral imagery, and (3) estimate SY using aerial imagery-derived reflectance (REFs) and VIs inputs with statistical and ML models at various training data sizes.

2. Materials and Methods

2.1. Experimental Details

The study was conducted in the 2022 growing season at the Tidewater Agriculture Research and Extension Center (TAREC) experimental farm in Suffolk, Virginia ($36^\circ 41' 03.4'' \text{ N}$,

76°46′05.6″ W). Two varieties of soybean seeds i.e., resistant and susceptible to parasitic nematodes were planted in six replicate blocks of nine fungicide treatments, including the control (2 seed varieties × 6 replicates × 9 fungicide treatments = 108 plots). Each plot consisted of two crop rows spaced at 0.91 m and had lengths of 9.1 m. The fungicide treatments ranged in concentrations and combinations of Fluopyram [32], *Bacillus amyloliquefaciens* strains [33], Pydiflumetofen [34] and heat-killed *Burkholderia rinojenses* [35], hereafter termed as treatments A to I (The specific composition of any treatment is proprietary to the fungicide manufacturer, so is not discussed in this paper).

2.2. Aerial Spectral Imaging Campaign

High-resolution aerial imagery in visible and near-infrared wavelength ranges (blue [450 nm ± 16 nm], green [560 nm ± 16 nm], red [650 nm ± 16 nm], red-edge [RE: 730 nm ± 16 nm], and near-infrared [NIR: 840 nm ± 26 nm], ~1884 images) were acquired at 1.3 cm/pixel of resolution using a DJI Phantom 4 Multispectral quadcopter (SZ DJI Technology Co., Shenzhen, China, Figure 1) near late season on 18 October 2022. These images were acquired near solar noon at an altitude of 25 m above ground level (AGL) with 80% front and 75% side overlaps between them. A downwelling light sensor onboard the SUAS faced skyward to record solar irradiance during the flight mission for correcting spectral inconsistencies due to the ambient light variations during flight. Following the completion of the mission, images of a calibrated reflectance panel (6×, Sentera, Inc., St. Paul, MN, USA) were collected to perform radiometric calibration of the spectral images.

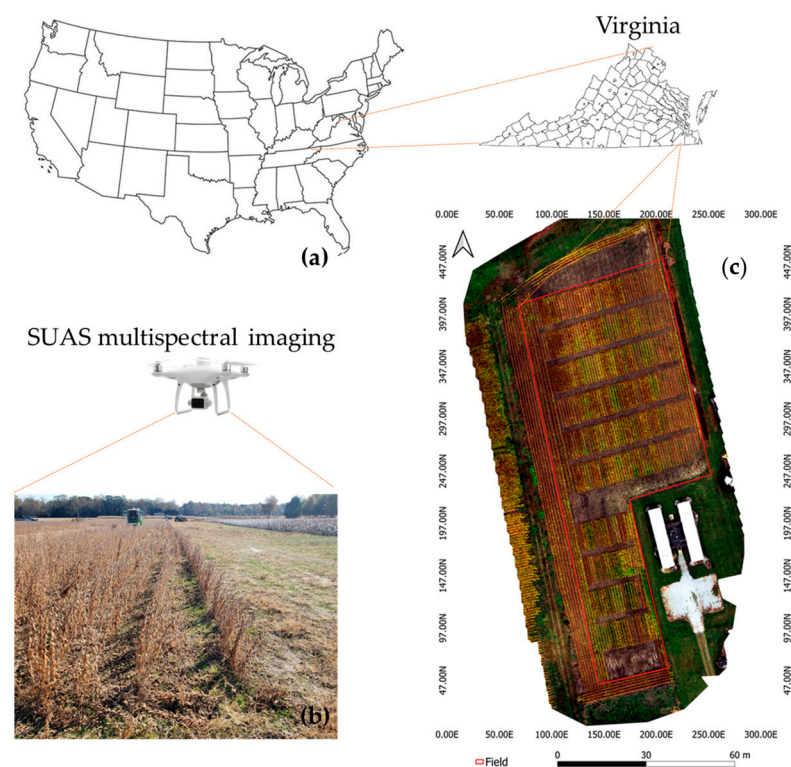


Figure 1. (a) Characterization of the experimental area at Tidewater Agricultural Research and Extension Center in Suffolk, VA. (b) Soybean experimental plots imaged using the aerial multispectral platform. (c) RGB natural color composition with study area bordered in red.

2.3. Imagery Processing for Deriving Vegetation Index Features

Pix4D Mapper (Pix4D, Inc., Lausanne, Switzerland) was used to obtain five spectral orthomosaics (corresponding to blue, green, red, red edge, and NIR sensors on the imaging sensor) through a series of image stitching operations on acquired imagery snapshots. These operations include key point feature extraction and matching, image model opti-

mization, georectification, point cloud densification, radiometric calibration, meshing, and orthomosaic generation (Figure 2). Generated orthomosaics were then exported to QGIS for further processing to obtain 24 VIs listed in Table 1 through the “Raster Calculator” toolbar (Figure 2). These listed VIs were selected for their importance in highlighting crop health under a variety of agroclimatic conditions. Soil background was then removed from each VI layer using the histogram separation method [36,37]. Next, rectangular regions of interest (ROI) were created over two central rows of each plot and mean REF and VI features were extracted for further analysis (Figure 2).

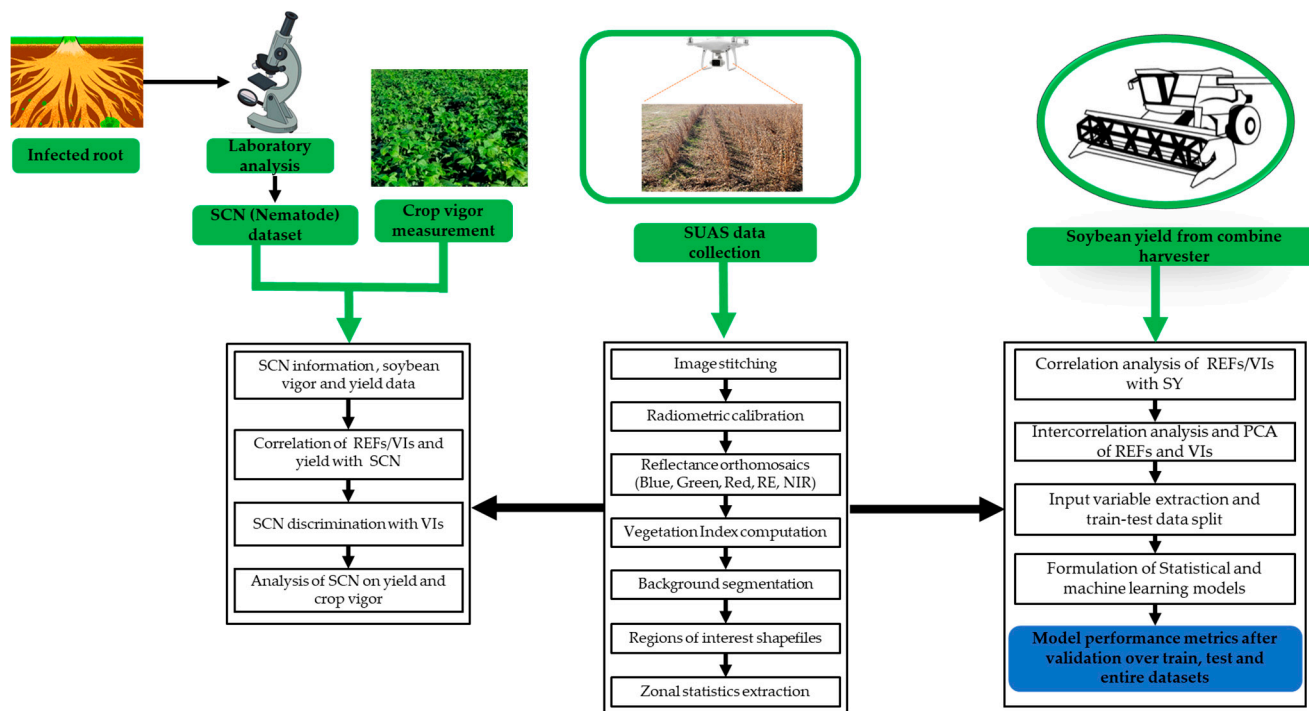


Figure 2. Flowchart showing soybean cyst nematode and crop vigor evaluation; and estimation of soybean yield using aerial multispectral imagery and statistical and machine learning models.

Table 1. Multispectral imagery-derived vegetation indices for soybean yield evaluations.

Vegetation Index	Equation	Reference
Green Difference Vegetation Index (GDVI)	$NIR - G$	[27]
Enhanced Vegetation Index (EVI)	$2.5 \times (NIR - R) / (NIR + 6 \times R - 7.5 \times B + 1)$	[6]
Infrared Percentage Vegetation Index (IPVI)	$(NIR) / (NIR + R)$	[26]
Green Normal Difference Vegetation Index (GNDVI)	$(NIR - G) / (NIR + G)$	[18]
Normalized Difference Vegetation Index (NDVI)	$(NIR - R) / (NIR + R)$	[18]
Wide Dynamic Range Vegetation Index (WDRVI)	$(a \times NIR - R) / (a \times NIR + R)$	[26]
Simple Ratio (SR)	NIR / R	[6]
Modified Non-Linear Index (MNLI)	$(NIR^2 - R) \times (1 + L) / (NIR^2 + R + L)$	[27]
Soil Adjusted Vegetation Index (SAVI)	$1.5 \times (NIR - R) / (NIR + R + 0.5)$	[13]
Optimized Soil Adjusted Vegetation Index (OSAVI)	$(NIR - R) / (NIR + R + 0.16)$	[15]
Green Soil Adjusted Vegetation Index (GSAVI)	$(NIR - G) / (NIR + G + 0.5)$	[27]
Green Optimized Soil Adjusted Vegetation Index (GOSAVI)	$(NIR - G) / (NIR + G + 0.16)$	[26]
Modified Soil Adjusted Vegetation Index (MSAVI2)	$(2 \times NIR + 1 - \sqrt{(2 \times NIR + 1)^2 - 8 \times (NIR - R)}) / 2$	[26]
Normalized Difference Red Edge Index (NDRE)	$(NIR - RE) / (NIR + RE)$	[6]

Table 1. Cont.

Vegetation Index	Equation	Reference
Green Ratio Vegetation Index (GRVI)	NIR/G	[23]
Green Chlorophyll Index (GCI)	$(NIR/G) - 1$	[21]
Green Leaf Index (GLI)	$((G - R) + (G - B))/((2 \times G) + R + B)$	[27]
Modified Simple Ratio (MSR)	$((NIR/R) - 1)/(\sqrt{(NIR/R) + 1})$	[15]
Renormalized Difference Vegetation Index (RDVI)	$(NIR - R)/\sqrt{(NIR + R)}$	[25]
Transformed Difference Vegetation Index (TDVI)	$1.5 \times ((NIR - R)/\sqrt{(NIR + R + 0.5)})$	[26]
Visible Atmospherically Resistant Index (VARI)	$(G - R)/(G + R - B)$	[13]
Leaf Area Index (LAI)	$3.618 \times EVI - 0.118$	[26]

The spectral reflectance in red, green, blue, red edge, and near-infrared wavelength ranges are represented by R, G, B, RE, and NIR, respectively.

2.4. Ground Data

The ground data was acquired at three crop stages: crop vigor in early season (7 and 21 July 2022), SCN counts in late season (21 October 2022), and yield at season end (5 December 2022). The soil sampling for SCN was conducted on the same date as the aerial imaging. The SY data was collected using a combine harvester equipped with a load sensor type yield monitor. Crop vigor was rated on a scale of 1–10 by the plant pathology specialist. Soil samples from each treatment plot were collected using sampling probes and transported to the taxonomic laboratory. These samples were cleaned to remove soil and debris through a semi-automatic elutriator apparatus. Next, the resultant samples were treated with glycerol solution to separate nematodes floating in the water to the top. The separated water-nematode layers were then used to prepare slide specimens for visualization under digital compound microscope. The nematode populations were then counted and recorded for each treatment plot by the taxonomic expert.

2.5. Data Analysis

2.5.1. Impact of Nematode Infestation

A dataset comprising crop vigor ratings, SCN populations, 24 VI features, five REF, and SY measurements (lbs/plot) was generated for 108 plots and checked to ensure that all the data followed a normal distribution (histogram and Shapiro–Wilk test). In the first analyses, the vigor readings, vigor progression, and SCN populations (females and juveniles) for both susceptible and resistant varieties under different treatments were evaluated and compared using two sample *t*-tests. Remotely sensed crop vigor (VI) was then examined for differences between the susceptible and resistant varieties using the two sample *t*-test. Next, the Pearson correlations between SCN and VIs as well as SCN and yield were evaluated.

2.5.2. Soybean Grain Yield Estimation Using Machine Learning

For yield estimation, methods similar to those published by Jjagwe et al. [26] were adopted. Initially, correlation (*r*) analysis was performed between SY and all the derived spectral (REF and VI) features. Next, from these features, variables to be used as inputs for the ML and statistical modeling of SY were identified as part of dimensionality reduction to minimize the model's over-fitness and enhance robustness. This was accomplished by first identifying the collinear variables through a principal component analysis (PCA) around two main axes that explained maximum variability in data. Secondly, intercorrelations between the REFs and VIs were evaluated through a pair-wise correlation and variables with strong correlations (>0.95) were identified. The variables depicting largest mean absolute correlation were eliminated and the remaining ones were selected for SY modeling. Following dimensionality reduction, SY was modeled using two statistical models: stepwise linear regression (SLR), partial least-square regression (PLSR); and four ML models: k-

nearest neighbor (KNN), support vector machine (SVM), random forest (RF), and artificial neural network (ANN).

In SLR, the variable with the highest sum of squares of regression is chosen. Next, an additional variable is selected from the remaining variables to develop binary relationships between dependent and independent variables [18,38]. In this study, forward and backward variable selection and Akaike Information Criterion (AIC) was used for implementing SLR. PLSR offers a multivariate method for utilizing highly correlated (collinear) variables to predict a single response variable utilizing the cross-validation hyperparameter. PLSR functions by preserving relationships between independent and dependent variables (SY) while eliminating multiple linearities between independent variables.

RF is a widely adopted ML model for agricultural studies that predicts an output through integrated multiple decision trees. RF is known to efficiently processes large datasets and reduce noises [39,40]. In this study, RF was initiated with 1000 trees following which best-fit trees were identified in the ranges of 300–400 through a five-fold cross validation i.e., five variables chosen at “random” to serve as candidates for every tuning iteration [41]. KNN approximates the relationship between input variables and an output based on the average of observations in the same neighborhood. KNN assigns a contribution weight to each neighbor, to have the closer neighbors contribute more to the average than the farther neighbors do. In this study, KNN was trained through repeated cross validation with three iterations for 30 neighbors, and an optimum tuning length of 20 [42]. SVM iteratively finds a hyperplane in an n-dimensional space that relates the input and output data points. Iteratively developing this hyperplane allows for the lowest possible misclassification error while forecasting continuous outputs [18,43,44]. SVM only considers training samples that are nearest to the ideal class boundary in feature space. In this study, Sequential Minimal Optimization (SMO) algorithm was used for SVM optimization. ANN repeatedly modifies its neural weights and thresholds to adapt to the computing environment. The network’s training is considered successful when the output error reaches the expected value. An input layer, one or more hidden layers, and an output layer, each having nodes (or neurons), make up the layers–node combinations of ANN. The neurons or nodes receive inputs from the nodes of previous layers and provide output to the nodes in the next layer [42,45,46]. With this configuration, ANN has the ability to robustly compute complex relationships between inputs and outputs. This study utilized a Fate of 0.01, backward propagation and two hidden layers with ten and three nodes, respectively, as deemed suitable for optimization problems.

For all the ML and statistical models, ten splits of training–testing datasets and two groups of input variables, (1) REFs and (2) REFs+VIs, were adopted. Using a 5% increment of the training dataset, these training–testing datasets ranged from 50:50 to 95:5 for ten split ratios. The train–test splits were developed to determine and assess the ideal training–testing data size for optimal model performance, particularly for small- to medium-sized datasets like the 108 data points used in this study. The trained models were applied to the training dataset, the testing dataset, and the entire dataset to assess the model estimation performances.

$$\text{rRMSE (\%)} = 100 \times \frac{\sqrt{\frac{\sum_{i=1}^n (\text{SY}_E - \text{SY}_m)^2}{n}}}{\text{mean}(\text{SY}_m)} \quad (1)$$

where SY_m is the measured and SY_E is the estimated SY. The R statistical computing software (version 4.3.1; RStudio, Inc. Boston, MA, USA) was used for all ML and statistical modeling, metrics (r and rRMSE) computations, and other analyses at 5% significance. The model’s performance and the accuracy of the SY estimation were assessed using the Pearson correlation (r) and relative root mean square error (rRMSE, %, Equation (1)).

3. Results

3.1. Impact of Nematode Infestation

Figure 3 shows high resolution true color composite and sample VI maps, NDVI and GNDVI for the study site showing contrasting variations in the canopy vigor. Using the Pearson correlation (r), VIs such as GNDVI, GRVI, and GCI ($r = -0.6$) showed maximum correlations with female nematodes (FNEMS) and juvenile nematodes (JNEMS) (Table 2). VARI had the weakest correlation for both FNEMS ($r = 0.24$) and JNEMS ($r = 0.05$). Amongst REFs, a strong correlation was observed between the blue band and SCN (JNEMS+FNEMS) ($r = 0.58$). From the Pearson correlation with SCN, NDRE appeared particularly suitable for differentiating susceptible and resistant varieties as NDRE values for resistant varieties (0.152–0.187) were higher than those for susceptible varieties (0.132–0.159) (Figure 4). Canopy vigor ratings for resistant varieties ranged from 5 to 9 for 7 July and 6 to 10 for 21 July. Similarly, for susceptible varieties, vigor ratings ranged from 5 to 8 for 7 July and 6 to 10 for 21 July. In addition, the SCN population (JNEMS+FNEMS) for resistant varieties ranged from 0 to 340 (Mean = 95, standard deviation [SD] = 81) and 20 to 1080 for susceptible varieties (Mean = 336, SD = 242). The SCN population had a moderate correlation ($r = 0.41$) with crop vigor on 7 July, a weak correlation ($r = 0.04$) on 21 July and a moderate correlation ($r = -0.39$) with yield (Figure 4). On average, SY was significantly higher for the resistant variety compared to the susceptible ones (on average +15 lbs/plot) (Figure 4). Significant differences were observed between canopy vigor ratings collected on July 7 ($t_{\text{stat}} = -11.126$, $p < 0.001$) and July 21 ($t_{\text{stat}} = -3.146$, $p < 0.025$), vigor recovery ($t_{\text{stat}} = 4.174$, $p < 0.001$), yield ($t_{\text{stat}} = 11.719$, $p < 0.001$), SCN populations ($t_{\text{stat}} = -6.881$, $p < 0.001$) and NDRE ($t_{\text{stat}} = 19.79$, $p < 0.001$) for resistant and susceptible varieties (two-sample t -test, $p < 0.01$).

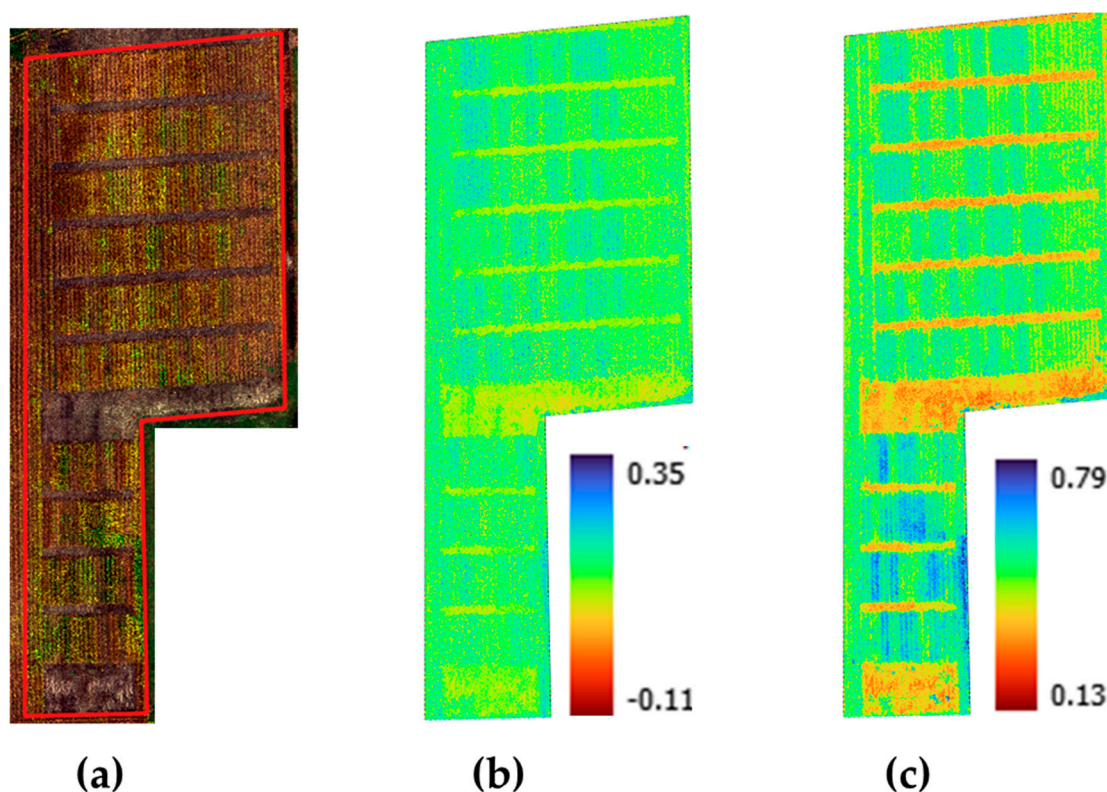


Figure 3. Aerial imagery-derived (a) RGB true color composite; and sample vegetation index maps for (b) Normalized Difference Vegetation Index and (c) Green Normalized Difference Vegetation Index.

Table 2. Correlations of juvenile and female nematodes with crop vigor and soybean yield.

Parameter	Juvenile Nematode (r)	Female Nematode (r)	FNEMS+JNEMS (r)
July 7 vigor	0.33	0.39	0.41
July 21 vigor	0.01	0.07	0.04
Vigor recovery	-0.29	0.28	-0.33
Yield	-0.30	-0.40	-0.39

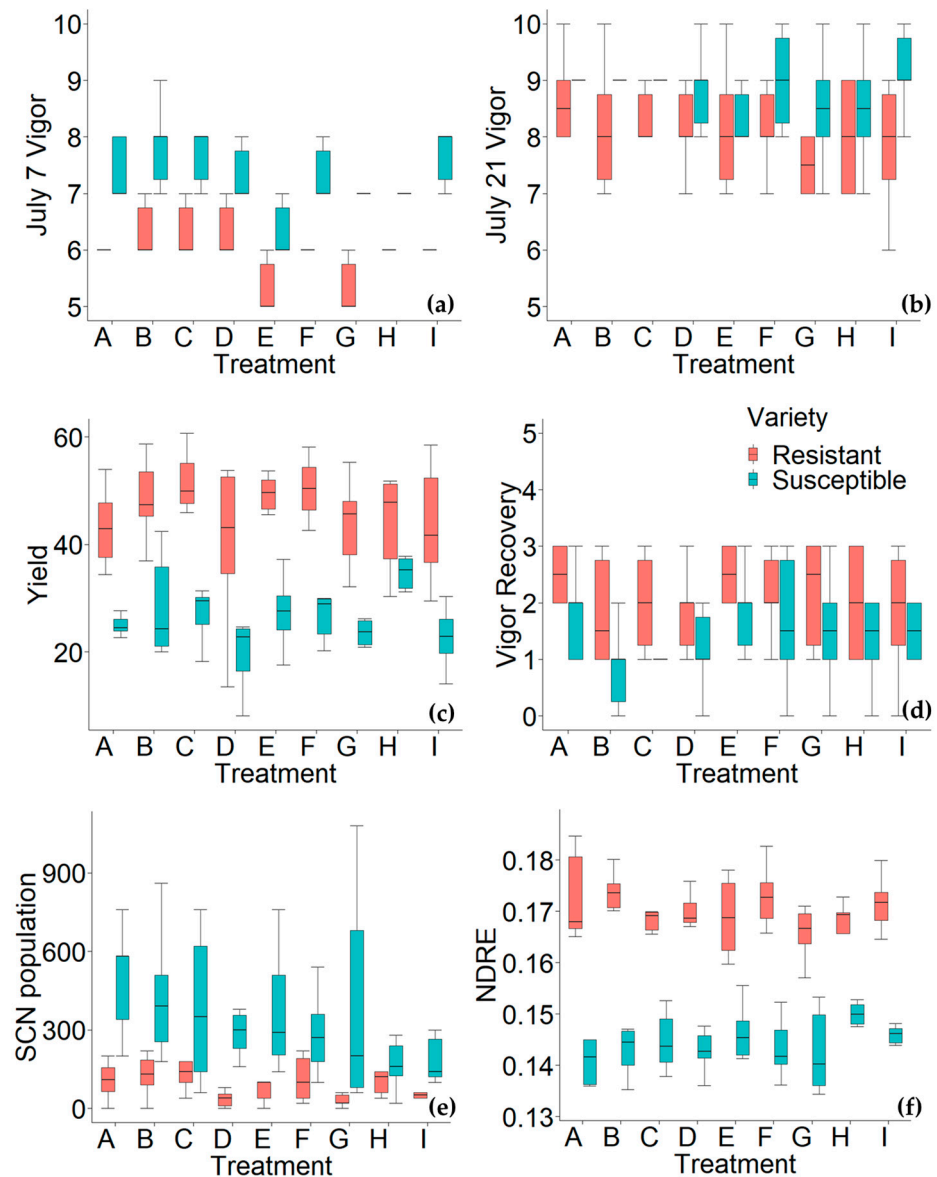


Figure 4. Presentation of contrasting differences in (a) 7 July vigor; (b) 21 July vigor; (c) yield; (d) vigor recovery; (e) soybean cyst nematode population; and (f) aerial imagery-derived normalized difference red edge index for resistant and susceptible soybean varieties under different fungicide treatments (A to I).

3.2. Correlations of Reflectance and Vegetation Indices with Soybean Yield

The Pearson correlation (r) between REFs with SY was the strongest for NDRE (r = 0.75) and weakest for REF in the red wavelength range (R, r = -0.0046) (Figure 5b,c and Table 3). A stronger correlation with the SY was observed for 6 VIs (r = 0.36–0.75) and 18 VIs showed

moderate-to-weaker correlations (0.15–0.26). Among the VIs, NDRE ($r = 0.75$) had the highest and GDVI (0.36) had the lowest correlation with SY.

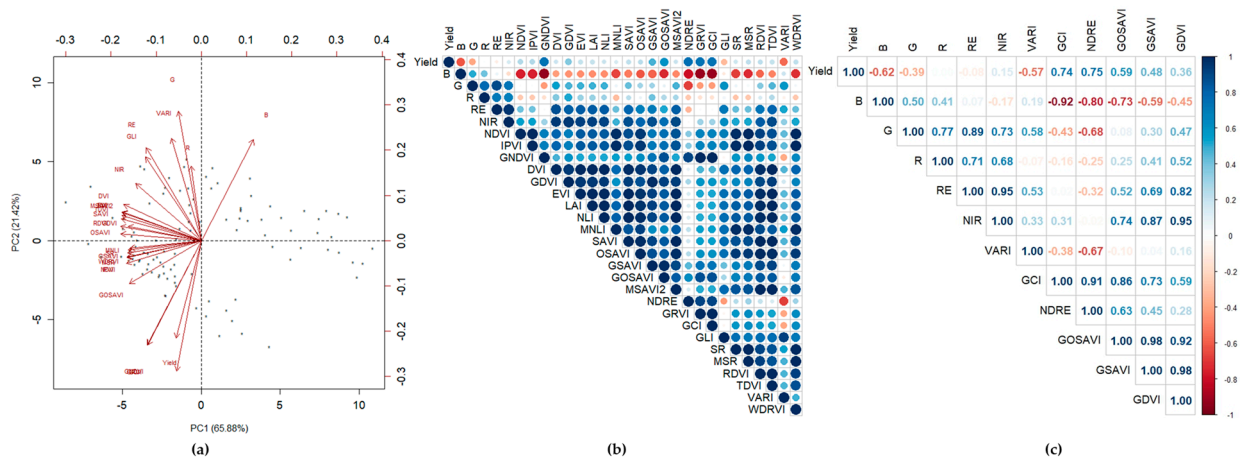


Figure 5. (a) PCA biplot analysis to observe variability and collinearity of the 24 vegetation indices and five reflectance features; (b) correlation of vegetation indices derived from aerial multispectral imagery with yield, and (c) final features selected as inputs after dimensionality reduction.

Table 3. Correlations of reflectance and vegetation indices with nematode populations and soybean yield.

Vegetation Index/Reflectance Feature	Yield (r)	Juvenile Nematode (r)	Female Nematode (r)	FNEMS+JNEMS (r)
Blue	−0.62 *	0.50 **	0.50 **	0.58 **
Green	−0.39	0.14	0.19	0.18
Red	-4.64×10^{-3}	0.09	0.02	0.08
Red Edge	−0.08	−0.09	−0.04	−0.08
Near Infrared	0.15	−0.21	−0.20	−0.24
Normalized Difference Vegetation Index (NDVI)	0.26	−0.42	−0.32	−0.45 **
Infrared Percentage Vegetation Index (IPVI)	0.26	−0.42	−0.32	−0.45 **
Green Normal Difference Vegetation Index (GNDVI)	0.73 *	−0.48	−0.53 **	−0.58 **
Difference Vegetation Index (DVI)	0.18	−0.32	−0.26	−0.35
Green Difference Vegetation Index (GDVI)	0.36	−0.34	−0.34	−0.40
Enhanced Vegetation Index (EVI)	0.21	−0.34	−0.28	−0.37
Leaf Area Index (LAI)	0.21	−0.34	−0.28	−0.37
Non-Linear Index (NLI)	0.20	−0.34	−0.27	−0.37
Modified Non-Linear Index (MNLi)	0.15	−0.36	−0.25	−0.38
Soil Adjusted Vegetation Index (SAVI)	0.21	−0.36	−0.29	−0.39
Optimized Soil Adjusted Vegetation Index (OSAVI)	0.24	−0.39	−0.31	−0.43 **
Green Soil Adjusted Vegetation Index (GSAVI)	0.48	−0.40	−0.41 *	−0.47 **
Green Optimized Soil Adjusted Vegetation Index (GOSAVI)	0.59	−0.45 **	−0.48 *	−0.54 **
Modified Soil Adjusted Vegetation Index (MSAVI2)	0.19	−0.34	−0.27	−0.37
Normalized Difference Red edge Index (NDRE)	0.75 *	−0.40	−0.49 *	−0.50 **
Green Ratio Vegetation Index (GRVI)	0.74 *	−0.48 **	−0.53 *	−0.58 **
Green Chlorophyll Index (GCI)	0.74 *	−0.48 **	−0.53 *	−0.58 **
Green Leaf Index (GLI)	−0.31	−0.13	0.04	−0.08
Simple Ratio (SR)	0.17	−0.37	−0.27	−0.39

Table 3. Cont.

Vegetation Index/Reflectance Feature	Yield (r)	Juvenile Nematode (r)	Female Nematode (r)	FNEMS+JNEMS (r)
Modified Simple Ratio (MSR)	0.22	−0.40	−0.30	−0.43 **
Renormalized Difference Vegetation Index (RDVI)	0.22	−0.38	−0.30	−0.41 **
Transformed Difference Vegetation Index (TDVI)	0.20	−0.35	−0.28	−0.38
Visible Atmospherically Resistant Index (VARI)	−0.57	0.05	0.24	0.14
Wide Dynamic Range Vegetation Index (WDRVI)	0.22	−0.40	−0.30	−0.43 **

The reflectance in red, green, blue, red edge, and NIR images is denoted by R, G, B, RE, and NIR, respectively. FNEMS and JNEMS represent female and juvenile nematodes, respectively. At $p < 0.001$, correlation coefficients are considered significant. * Highest correlations for yield. ** Highest correlations with nematode populations.

3.3. Soybean Yield Estimation with ML

3.3.1. Spectral Feature Assessment and Variable Selection

Based on PCA biplots assessments, two major principal components for 24VIs and five REFs accounted for yield variability of 65.88% (PC-1) and 21.42% (PC-2, Figure 4a), respectively. Eigenvectors for the REFs in green, red, and blue wavebands tended towards the top of the biplot (Figure 4a), showing strong influence on PC-2. While REFs in RE and NIR wavebands, and derived VIs, densely formed a cluster towards the top-left, extreme left, or lower-left region, showing stronger influence on PC-1. Eigenvectors for GNDVI and GRVI tended toward the lower-left region while VARI in the top-left region. Six VIs, including GCI, NDRE, GOSAVI, GSAVI, GDVI, and VARI, had strong correlations with SY when evaluated using the Pearson correlation (r). Following these observations, a total of five REF and six VI features were identified as inputs for the SY estimation (Figure 5).

3.3.2. Using Spectral Reflectance Features as Predictor Variables

SY was modeled using two input groups (1) using REF features, and (2) a combination of REF and VI features in the selected statistical and ML models (Table 4). Pertaining to using only REF as inputs, RF yielded the best performance at a 95:5 train–test data split ($r = 0.96$, $rRMSE = 11.60\%$), followed by ANN, KNN, SLR, and PLSR. SVM yielded the least-best performance at a 95:5 data split ($r = 0.83$, $rRMSE = 20.42\%$) for validation over the train dataset. PLSR was the best performer ($r = 0.83$, $rRMSE = 20.30\%$) at a 50:50 train–test split ratio, followed by SVM, ANN, SLR, RF, while KNN was the least-best performer ($r = 0.70$, $rRMSE = 26.70\%$) when validated over the test dataset. When the validation was conducted over the entire dataset, RF yielded the best performance ($r = 0.93$, $rRMSE = 14.60\%$) at a 95:5 split followed by ANN, SLR, KNN, and PLSR, and SVM ($r = 0.80$, $rRMSE = 21.95\%$) yielded the least-best performance at a 95:5 split ratio.

Table 4. Machine learning model performances at various train–test data split ratios using reflectance and a combination of reflectance and vegetation indices as inputs.

Parameters		Dataset: Train			Dataset: Test			Dataset: Entire		
Train:Test Ratio	Input Group	Best Model	r	rRMSE (%)	Best Model	r	rRMSE (%)	Best Model	r	rRMSE (%)
50:50	REFs	RF	0.95	13.10	PLSR	0.83	20.30	RF	0.84	20.60
	REFs+VIs		0.96	11.10	SLR	0.83	19.98		0.87	18.00
55:45	REFs	RF	0.96	12.00	PLSR	0.81	20.80	RF	0.84	19.90
	REFs+VIs		0.97	10.30	SLR	0.81	21.30		0.88	17.40

Table 4. Cont.

Parameters		Dataset: Train			Dataset: Test			Dataset: Entire		
Train:Test Ratio	Input Group	Best Model	r	rRMSE (%)	Best Model	r	rRMSE (%)	Best Model	r	rRMSE (%)
60:40	REFs	RF	0.96	12.00	PLSR	0.80	21.40	RF	0.84	20.00
	REFs+VIs		0.97	10.10	RF	0.84	20.00		0.88	16.90
65:35	REFs	RF	0.96	11.90	PLSR	0.81	22.20	RF	0.85	19.30
	REFs+VIs		0.97	10.00	SLR	0.81	23.00		0.90	16.10
70:30	REFs	RF	0.96	11.80	SLR	0.80	21.80	RF	0.85	19.10
	REFs+VIs		0.97	9.96	SLR	0.81	21.60		0.90	15.90
75:25	REFs	RF	0.96	11.40	SLR	0.79	21.00	RF	0.85	19.30
	REFs+VIs		0.97	9.62	SLR	0.77	22.00		0.90	15.50
80:20	REFs	RF	0.96	11.60	SLR	0.76	22.30	RF	0.88	17.60
	REFs+VIs		0.97	9.45	SLR	0.73	23.5		0.91	15.20
85:15	REFs	RF	0.96	11.50	SLR	0.61	23.50	RF	0.88	17.20
	REFs+VIs		0.97	9.20	SLR	0.6	24.80		0.91	14.90
90:10	REFs	RF	0.96	11.50	SLR	0.65	27.30	RF	0.91	15.80
	REFs+VIs		0.97	9.08	SLR	0.66	27.40		0.93	13.60
95:5	REFs	RF	0.96	11.60	SLR	0.71	38.00	RF	0.93	14.60
	REFs+VIs		0.97	8.72	ANN	0.79	39.60		0.93	13.32

The input groups are designated as REFs (reflectance only) and REFs+VIs (reflectance and selected vegetation indices).

3.3.3. Using Spectral Reflectance and Vegetation Index Features as Predictor Variables

The dimensionality reduction process yielded six VIs and five REF spectral features which were later used together as predictor variables for SY estimation (Table 4). RF ($r = 0.97$, $rRMSE = 8.72\%$) yielded the best performance at a 95:5 split when validated over the train dataset; reduced performances were obtained for ANN, KNN, SLR, and SVM, while PLSR yielded the least-best performance at that split ($r = 0.83$, $rRMSE = 19.79\%$). For validation over test dataset, RF outperformed at a 60:40 split ($r = 0.84$, $rRMSE = 20.00\%$), whose performance was followed by SLR, SVM, ANN, SLR and PLSR, while KNN was the least-best performer at the same 60:40 split ($r = 0.78$, $rRMSE = 22.7\%$). When validated over the entire dataset, RF at a 95:5 split ($r = 0.93$, $rRMSE = 13.32\%$) yielded the best performance followed by ANN, KNN, SLR, and PLSR, while SVM ($r = 0.82$, $rRMSE = 21.39\%$) was the least performer at a 95:5 split.

3.3.4. Impact of Dataset Sizes on Soybean Yield Modeling

SY estimation accuracy diminished with increase in data split ratios (r_{test} : 0.83–0.79, $rRMSE_{test}$: 20.30–39.60%, Figures 6b and 7, Table 4) when models were validated for performance over the test dataset. On the contrary, the SY estimation accuracy enhanced with the increase in training dataset size (r_{train} : 0.95–0.97, $rRMSE_{train}$: 13.10–8.72%, r_{entire} : 0.84–0.93, $rRMSE_{entire}$: 20.60–13.32%, Figures 6a,c and 7, and Table 4) for the models when validated over train and entire datasets. When validated over those datasets at a 95:5 split, RF with REFs+VIs as the inputs, yielded the best SY estimation performance ($r_{train} = 0.97$, $rRMSE_{train} = 8.72\%$, $r_{entire} = 0.93$, $rRMSE_{entire} = 13.32\%$, Figures 6a,c and 7, and Table 4). RF with REFs+VIs as the input variables also was the best performer when it was validated over the test dataset, at the 60:40 split ratio.

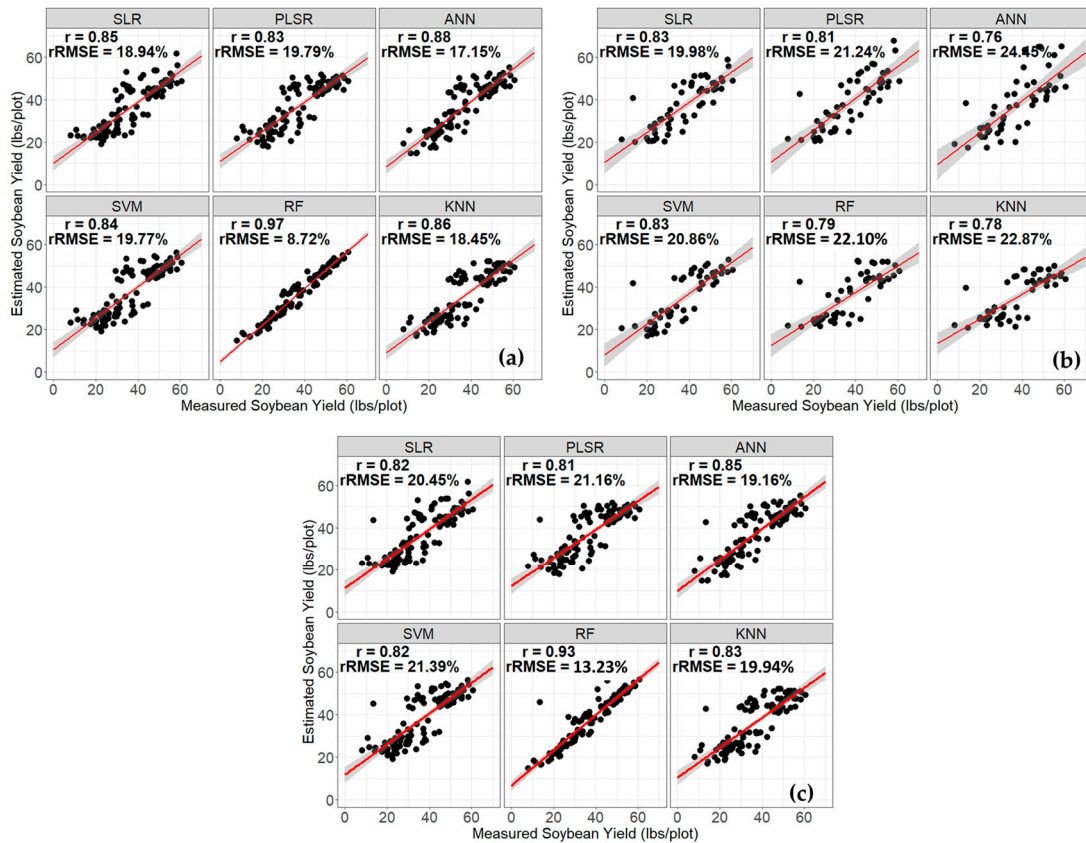


Figure 6. The Pearson correlation (r) and relative root mean square error (rRMSE) for measured and estimated soybean yields with REFs+VIs as an input group for models validated over (a) the train dataset at 95:5, (b) test dataset at 50:50, and (c) entire dataset at 95:5 splits.

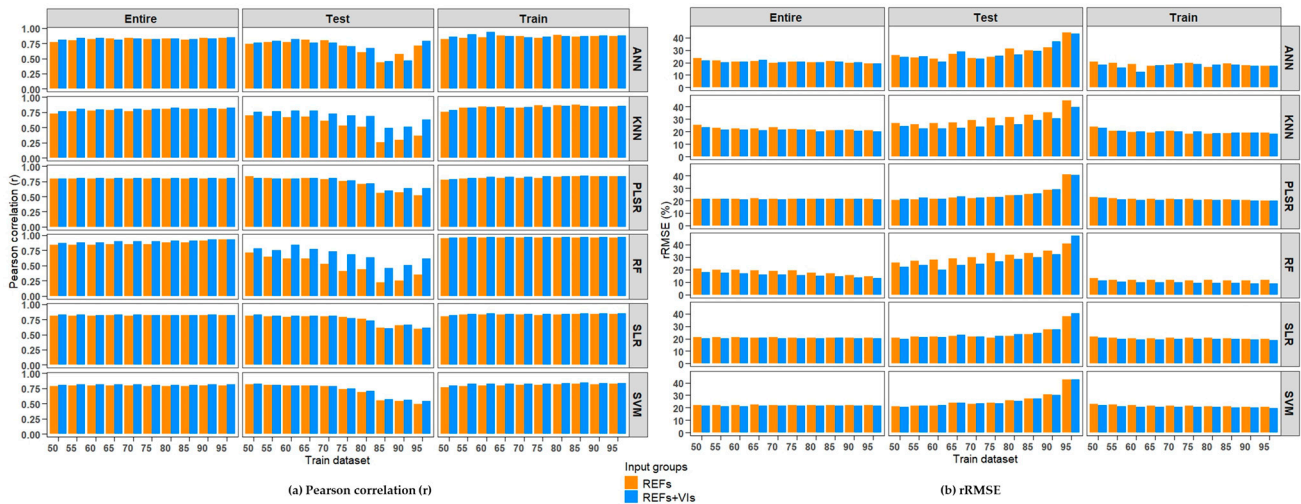


Figure 7. Plots showing performances of six soybean yield estimation models through the (a) Pearson correlation (r) and (b) relative root mean square error (rRMSE) at ten train–test data split ratios and two input groups (REFs, REFs+VIs) when validated over entire, test and train datasets.

4. Discussion

The impact of SCN on soybean yield and physiology was evaluated using ground-based information and aerial spectral imagery. Further, ML models were formulated and tested for SY estimation under varied dataset sizes. The blue band ($r = -0.58$) and VIs such as GNDVI, GRVI, and GCI ($r = -0.58$) were identified to be the best discriminants of

SCN populations for resistant and susceptible varieties. GNDVI has also been identified as suitable for disease identification in other studies that pertained to SCN and sudden death syndrome in soybean [10] and yellow leaf disease of arecanut [47]. The SCN and vigor correlation on 21 July ($r = 0.04$) was lower than that on 7 July ($r = 0.41$) (Table 2) which could be attributed to the development of visible mild symptoms on crop canopy not only due to SCN but also the varied impacts of fungicide treatments, and possibly due to responses to other unmeasured biotic and abiotic stressors. This observation could also be due to the variety response to SCN populations with resistant variety showing a higher vigor and vigor progression. This is consistent with findings by Joalland et al. [48] where the beet cyst nematode population multiplication was higher in the susceptible cultivar compared to the tolerant one. Furthermore, the ability of nematode tolerant cultivar to endure nematode damage in the second half of the growing season was shown by no difference in the final shoot biomass [48]. This makes visual scouting of canopies for SCN infestation very challenging [10]. Nonetheless, the positive vigor progression between the two measurement dates as well as relatively lesser SCN populations indicated sufficient availability of water and minimal level of other stressors on the canopy. This aspect is well supported by a study that observed more SCN on the root cortex instead of vascular tissues under adequate soil moisture conditions [10].

SCN population showed a moderate correlation with NDRE ($r = -0.5$) and yield ($r = -0.4$) which likely reflects SCN damage. This could be due to two primary reasons, firstly, the variation in crop vigor not only due to SCN but also the varied impacts of fungicide treatments, and possible responses to other unmeasured biotic and abiotic stressors. Secondly, the SCN assessment using aerial imaging of crop canopy forms an indirect mode of assessment where SCN, in reality, are observed from soil samples. Nonetheless, yield, vigor, and NDRE did quantify significant differences for resistant and susceptible varieties. This observation is well supported by Joalland et al. [48] who evaluated beet cyst nematode infestation in sugar beets (susceptible and tolerant cultivars) using visible light imaging, thermography and spectrometry. Spectral vegetation indices (NDVI, MCARI2) and CHLG) showed a correlation with yield ($R = 0.69$) and nematode population in the soil ($R = 0.78$) for the susceptible cultivar. Several studies have reported that SCN infection reduces photosynthesis rate, chlorophyll content, and plant growth with chlorosis forming the aboveground symptom on the canopy [6]. NDRE related to leaf chlorophyll showed strong correlation with the SY for the resistant and susceptible varieties and was consistent with observation that increased SCN damage to the roots results in more visible symptoms on the shoots, and lower yields [8]. This is strongly supported by a negative and significant correlation between the SCN population and yield [49]. In this study, GNDVI, GCI, and NDRE demonstrated the ability to discriminate the soybean vigor, yield, and SCN conditions which is similar to the findings by Santos et al. [6]. These results demonstrate the use of aerial multispectral imagery data as a quick, non-destructive, and an economical way for SCN infestation detection and characterization.

Out of the 24 initially derived VIs, six VIs and five REFs that showed strong correlation with SY were selected as inputs for ML and statistical modeling to avoid overfitting and improve accuracy of the estimates. The reflectance in blue ($r = -0.62$) followed by the reflectance in green ($r = -0.39$) wavebands had the highest correlations with SY, while NIR had low correlation with SY ($r = 0.15$) possibly due to lowered chlorophyll levels at the later stage [50]. The reflectance in the red waveband had the lowest correlation due to its lower sensitivity to chlorophyll variations ($r = -0.0046$) [38]. NDRE had the highest correlation with SY out of all 24 derived VIs, followed by GCI, GOSAVI, GSAVI, GDVI, and VARI, among others ($r = -0.57$ – 0.75). The lowest (in magnitude) correlation ($r = 0.36$) was found with GDVI. The strong correlation is supported by the composition of NDRE and GCI, with NIR (840 ± 20 nm), red edge (RE: 730 nm ± 16 nm), and green (560 nm ± 16 nm) wavebands that are relatively more sensitive to the chlorophyll content [38]. However, due to the mathematical nature of GDVI and the use of low-correlated NIR waveband in its computation, it could have exhibited lower correlation of GDVI with SY. Stronger

correlations between SY and VIs like NDRE and GCI are possibly because those indices account for dynamic variations in the visible-NIR region related to nitrogen and chlorophyll content [51]. It is also the reason for higher correlations of NDRE (which uses the red-edge band in its computation) and GCI (which requires the green band) compared to IPVI, and NDVI [52–54]. Although using either the broadband spectral bands (REFs) or VIs as standalone inputs in linear models may be feasible for SY estimation, these may not stand robust when crops are subjected to high variations in agroclimatic conditions [24,28]. On the contrary, non-linear combination of spectral indices, REFs, and texture features could improve the prediction accuracy of the models [24,54]. Following that, this study went a step further to estimate SY (output) using only REFs as well as their non-linear combination with VIs as inputs in statistical and ML models for more accurate, robust, and intricate predictions.

In this study, REF in blue, green, red, RE, and NIR wavebands, and Vis, NDRE, GCI, GOSAVI, GSAVI, GDVI and VARI, were identified as not to have absolute correlations (i.e., $r = 1$) among each other using PCA. This aided dimensionality reduction and eliminating collinear variables prior to SY modeling. NDRE is able to overcome saturation limitation [55,56] when the crop canopy reached 100% [55,57] due to linear correlation with biomass and yield [56]. Studies that have carried out crop yield and nitrogen estimations using spectral wavebands or a combination of wavebands and VIs through ML [28,53] have not assessed or removed inter-correlations among the input variables (VIs) prior to output estimation [58,59]. This may highly result into overfitting and reduced the robustness of the developed and tested models [26]. For model training and validation, the majority of machine learning-based prediction studies have by default used train–test data split ratios of 70:30 or 80:20 [42,60–62]. However, there has not been much research conducted on determining the ideal train–test data splits while considering total dataset size. This could form another factor impacting robustness and over fitness of the models [26]. All these factors were considered in this study especially due to the moderate size of the involved dataset. As the proportion of the training data increased, it was found that the model's performances improved (Figure 7) when validated over the train and entire datasets [26,53,58]. This observation held true for both input groups where REFs+VIs inputs significantly outperformed REFs inputs in the models ($p = 0.01$, Table 5). When validated over the training and entire datasets, it was observed that ML models performed better than statistical models for SY estimation, irrespective of the input groups (Figure 7).

Table 5. Impact of input parameters on soybean yield estimation models' performance.

Variable	<i>p</i> Value (<i>r</i>)	<i>p</i> Value (rRMSE)
Model	<0.001	<0.001
Train–test split	<0.001	<0.001
Dataset	<0.001	<0.001
Input group	<0.001	<0.001
Train–test split: Dataset	<0.001	<0.001
Train–test split: Input group	0.217	0.330
Dataset: Input group	<0.001	0.820
Train–test split: Dataset: Input group	0.002	0.964

Input groups (REF, REF+VIs), the model (SLR, PLSR, ANN, SVM, RF, KNN), and the dataset (train, test, entire).

The performance of statistical models like SLR and PLSR was enhanced and was comparable to ML models as the confounding effects were already eliminated by eliminating the collinearity of input variables [63]. When compared to ML models, SLR and PLSR were the least-best performers irrespective of the input groups possibly due to their nature of

limitedly exploring non-linear complexities between the input and output variables [63]. The relatively smaller size of the data and the need for a larger size of training data for building optimal neural networks affected the performance of ANN [26,58] which was the best performer only at a 95:5 split ratio. Overall, RF outperformed all other models due to its dependency on decision trees that can solve overfitting issues irrespective of dataset size [26]. This is completed by having numerous random subsamples of the original dataset during the training of decision trees. It is also because of this that RF can handle numerous model parameters, reduce estimate bias, and produce more accurate and reliable predictions for novel instances that are not always present in the training dataset [64].

The viability of integrating SUAS and ML techniques using different data splits for SY estimation was shown in this study. Incorporating agroclimatic conditions like soil and weather parameters as well as data collection over several growth stages and cropping seasons may be a potential avenue to improve the model's performance.

It is also worth noting that the cost of implementing this technology such as initial investment in drones, sensors and data processing software, especially for small scale farmers still poses a challenge. Furthermore, handling large volumes of agricultural data, including aerial imagery and crop information, raises concerns about data privacy and security. Safeguarding sensitive information and complying with data protection regulations may present challenges in the implementation of the study. Our next goals are to incorporate these findings into satellite imaging platforms and determine the optimum stage for SCN infestation detection and SY prediction. This will consider providing open-source platforms such as web tools that are affordable and easily accessible to farmers and consider data privacy policies. Based on these estimations, farmers would be able to employ precision management strategies to enhance yield and profitability.

5. Conclusions

This study evaluated SCN impact on crop vigor and yield as well as the viability of using SUAS aerial multispectral imagery integrated with statistical and ML models for SY estimation. Crop vigor and yield showed significant differences between the resistant and susceptible varieties. Lower yield and crop vigor recovery as well as higher SCN population (20 to 1080) was observed for the susceptible varieties. Similarly, resistant varieties had higher yield, higher crop vigor recovery, and lower SCN population (0 to 340). The blue band ($r = 0.58$) and GNDVI ($r = -0.6$) showed the best correlation with the SCN rating. GDNVI, GCI and NDRE were able to differentiate between soybean plants with and without SCN symptoms.

SUAS imagery derived VIs such as NDRE, GCI, and GOSAVI, among others, had the strongest correlation ($r = 0.59$ – 0.75) with SY. REFs in blue, green, red, RE, NIR, and VIs like NDRE, GCI, GOSAVI, GSAVI, GDVI and VARI were identified as key inputs for statistical and ML modeling of SY. The statistical models and ML model performance for estimating SY increased as training dataset sizes increased. RF showed the best overall performance ($r: 0.84$ – 0.97 , $rRMSE: 20.60$ – 8.72%). The input groups (REFs or REFs+VIs) had a significant impact on model performance. When models were validated over the train and entire datasets, the 95:5 train–test split ratio performed the best. While the 60:40 split ratio performed best when models were validated over the test dataset.

This study showed SCN impact on crop vigor and yield and SY estimation using a combination of SUAS aerial multispectral imagery and ML models. This would be vital for implementing timely and precision management practices by farmers to maximize crop yield and profitability.

Author Contributions: Conceptualization, P.J. and A.K.C.; methodology, P.J., A.K.C. and D.B.L.; software, P.J. and A.K.C.; validation, P.J., A.K.C. and D.B.L.; formal analysis, P.J. and A.K.C.; investigation, P.J., A.K.C. and D.B.L.; resources, A.K.C. and D.B.L.; data curation, P.J., A.K.C. and D.B.L.; writing—original draft preparation, P.J. and A.K.C.; writing—review and editing, P.J., A.K.C. and D.B.L.; visualization, A.K.C. and D.B.L.; supervision, A.K.C.; project administration, A.K.C. and

D.B.L.; funding acquisition, A.K.C. and D.B.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by USDA NIFA Project # 420110, Hatch Project # VA160181, Multistate Hatch Projects # VA136412 and VA136438, and Faculty startup.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All collected data and pertaining analysis has been included in the manuscript.

Acknowledgments: We would like to thank the technicians of the plant pathology laboratory as well as Tidewater Agricultural Research and Extension Center, Suffolk, VA for their help in managing trials and collecting ground truth data.

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

References

1. Yoosefzadeh-Najafabadi, M.; Tulpan, D.; Eskandari, M. Using Hybrid Artificial Intelligence and Evolutionary Optimization Algorithms for Estimating Soybean Yield and Fresh Biomass Using Hyperspectral Vegetation Indices. *Remote Sens.* **2021**, *13*, 2555. [[CrossRef](#)]
2. Ren, P.; Li, H.; Han, S.; Chen, R.; Yang, G.; Yang, H.; Feng, H.; Zhao, C. Estimation of Soybean Yield by Combining Maturity Group Information and Unmanned Aerial Vehicle Multi-Sensor Data Using Machine Learning. *Remote Sens.* **2023**, *15*, 4286. [[CrossRef](#)]
3. Sun, J.; Di, L.; Sun, Z.; Shen, Y.; Lai, Z. County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model. *Sensors* **2019**, *19*, 4363. [[CrossRef](#)] [[PubMed](#)]
4. Arantes, B.H.T.; Moraes, V.H.; Geraldine, A.M.; Alves, T.M.; Albert, A.M.; da Silva, G.J.; Castoldi, G. Spectral Detection of Nematodes in Soybean at Flowering Growth Stage Using Unmanned Aerial Vehicles. *Cienc. Rural.* **2021**, *51*, e20200283. [[CrossRef](#)]
5. Arjoune, Y.; Sugunraj, N.; Peri, S.; Nair, S.V.; Skurdal, A.; Ranganathan, P.; Johnson, B. Soybean Cyst Nematode Detection and Management: A Review. *Plant Methods* **2022**, *18*, 1–39. [[CrossRef](#)] [[PubMed](#)]
6. Santos, L.B.; Bastos, L.M.; de Oliveira, M.F.; Soares, P.L.M.; Ciampitti, I.A.; da Silva, R.P. Identifying Nematode Damage on Soybean through Remote Sensing and Machine Learning Techniques. *Agronomy* **2022**, *12*, 2404. [[CrossRef](#)]
7. Agrios, G.N. *Plant Pathology*, 5th ed.; Elsevier Academic Press: Amsterdam, The Netherlands, 2005.
8. Joalland, S.; Screpanti, C.; Varella, H.V.; Reuther, M.; Schwind, M.; Lang, C.; Walter, A.; Liebisch, F. Aerial and Ground Based Sensing of Tolerance to Beet Cyst Nematode in Sugar Beet. *Remote Sens.* **2018**, *10*, 787. [[CrossRef](#)]
9. Martins, G.D.; de Lourdes Bueno Trindade Galo, M.; Vieira, B.S. Detecting and Mapping Root-Knot Nematode Infection in Coffee Crop Using Remote Sensing Measurements. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 5395–5403. [[CrossRef](#)]
10. Bajwa, S.G.; Rupe, J.C.; Mason, J. Soybean Disease Monitoring with Leaf Reflectance. *Remote Sens.* **2017**, *9*, 127. [[CrossRef](#)]
11. Kharel, T.P.; Maresma, A.; Czymmek, K.J.; Oware, E.K.; Ketterings, Q.M. Combining Spatial and Temporal Corn Silage Yield Variability for Management Zone Development. *Agron. J.* **2019**, *111*, 2703–2711. [[CrossRef](#)]
12. Sunoj, S.; Cho, J.; Guinness, J.; van Aardt, J.; Czymmek, K.J.; Ketterings, Q.M. Corn Grain Yield Prediction and Mapping from Unmanned Aerial System (UAS) Multispectral Imagery. *Remote Sens.* **2021**, *13*, 3948. [[CrossRef](#)]
13. Bai, D.; Li, D.; Zhao, C.; Wang, Z.; Shao, M.; Guo, B.; Liu, Y.; Wang, Q.; Li, J.; Guo, S.; et al. Estimation of Soybean Yield Parameters under Lodging Conditions Using RGB Information from Unmanned Aerial Vehicles. *Front. Plant Sci.* **2022**, *13*, 1012293. [[CrossRef](#)] [[PubMed](#)]
14. Terliksiz, A.S.; Altıylar, D.T. Use of Deep Neural Networks for Crop Yield Prediction: A Case Study of Soybean Yield in Lauderdale County, Alabama, USA. In Proceedings of the Eighth International Conference on Agro-Geoinformatics, Istanbul, Turkey, 16–19 July 2019.
15. Ma, C.; Liu, M.; Ding, F.; Li, C.; Cui, Y.; Chen, W.; Wang, Y. Wheat Growth Monitoring and Yield Estimation Based on Remote Sensing Data Assimilation into the SAFY Crop Growth Model. *Sci. Rep.* **2022**, *12*, 5473. [[CrossRef](#)] [[PubMed](#)]
16. Shawon, A.R.; Ko, J.; Ha, B.; Jeong, S.; Kim, D.K.; Kim, H.-Y. Assessment of a Proximal Sensing-integrated Crop Model for Simulation of Soybean Growth and Yield. *Remote Sens.* **2020**, *12*, 410. [[CrossRef](#)]
17. Adeboye, O.B.; Schultz, B.; Adekalu, K.O.; Prasad, K. Modelling of Response of the Growth and Yield of Soybean to Full and Deficit Irrigation by Using Aquacrop. *Irrig. Drain.* **2017**, *66*, 192–205. [[CrossRef](#)]
18. Fu, Z.; Jiang, J.; Gao, Y.; Krienke, B.; Wang, M.; Zhong, K.; Cao, Q.; Tian, Y.; Zhu, Y.; Cao, W.; et al. Wheat Growth Monitoring and Yield Estimation Based on Multi-Rotor Unmanned Aerial Vehicle. *Remote Sens.* **2020**, *12*, 508. [[CrossRef](#)]
19. Yousfi, S.; Marín, J.; Parra, L.; Lloret, J.; Mauri, P.V. Remote Sensing Devices as Key Methods in the Advanced Turfgrass Phenotyping under Different Water Regimes. *Agric. Water Manag.* **2022**, *266*, 107581. [[CrossRef](#)]
20. Zhang, M.; Zhou, J.; Sudduth, K.A.; Kitchen, N.R. Estimation of Maize Yield and Effects of Variable-Rate Nitrogen Application Using UAV-Based RGB Imagery. *Biosyst. Eng.* **2019**, *189*, 24–35. [[CrossRef](#)]

21. 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](#)]
22. Xu, W.; Chen, P.; Zhan, Y.; Chen, S.; Zhang, L.; Lan, Y. Cotton Yield Estimation Model Based on Machine Learning Using Time Series UAV Remote Sensing Data. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *104*, 102511. [[CrossRef](#)]
23. 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.* **2019**, *237*, 111599. [[CrossRef](#)]
24. Habibi, L.N.; Watanabe, T.; Matsui, T.; Tanaka, T.S.T. Machine Learning Techniques to Predict Soybean Plant Density Using UAV and Satellite-Based Remote Sensing. *Remote Sens.* **2021**, *13*, 2548. [[CrossRef](#)]
25. Herrero-Huerta, M.; Rodriguez-Gonzalvez, P.; Rainey, K.M. Yield Prediction by Machine Learning from UAS-Based Multi-Sensor Data Fusion in Soybean. *Plant Methods* **2020**, *16*, 78. [[CrossRef](#)]
26. Jjagwe, P.; Chandel, A.K.; Langston, D. Pre-Harvest Corn Grain Moisture Estimation Using Aerial Multispectral Imagery and Machine Learning Techniques. *Land* **2023**, *12*, 2188. [[CrossRef](#)]
27. Pinto, A.A.; Zerbato, C.; de Souza Rolim, G.; Barbosa Júnior, M.R.; da Silva, L.F.V.; de Oliveira, R.P. Corn Grain Yield Forecasting by Satellite Remote Sensing and Machine-Learning Models. *Agron. J.* **2022**, *114*, 2956–2968. [[CrossRef](#)]
28. Fei, S.; Hassan, M.A.; He, Z.; Chen, Z.; Shu, M.; Wang, J.; Li, C.; Xiao, Y. Assessment of Ensemble Learning to Predict Wheat Grain Yield Based on UAV-Multispectral Reflectance. *Remote Sens.* **2021**, *13*, 2338. [[CrossRef](#)]
29. Sakamoto, T. Incorporating Environmental Variables into a MODIS-Based Crop Yield Estimation Method for United States Corn and Soybeans through the Use of a Random Forest Regression Algorithm. *ISPRS J. Photogramm. Remote Sens.* **2019**, *160*, 208–228. [[CrossRef](#)]
30. Jin, X.; Zarco-Tejada, P.J.; Schmidhalter, U.; Reynolds, M.P.; Hawkesford, M.J.; Varshney, R.K.; Yang, T.; Nie, C.; Li, Z.; Ming, B.; et al. High-Throughput Estimation of Crop Traits: A Review of Ground and Aerial Phenotyping Platforms. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 200–231. [[CrossRef](#)]
31. Li, D.; Miao, Y.; Gupta, S.K.; Rosen, C.J.; Yuan, F.; Wang, C.; Wang, L.; Huang, Y. Improving Potato Yield Prediction by Combining Cultivar Information and UAV Remote Sensing Data Using Machine Learning. *Remote Sens.* **2021**, *13*, 3322. [[CrossRef](#)]
32. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem 2023 Update. *Nucleic Acids Res.* **2023**, *51*, D1373–D1380. [[CrossRef](#)]
33. Messa, V.; Nunes, J.; Mattei, D. Seed Treatment with *Bacillus Amyloliquefaciens* for the Control of *Meloidogyne Javanica* “In Vivo” Bean Culture and Its Direct Effect on the Motility, Mortality and Hatching of *M. Javanica* “In Vitro”. *Agron. Sci. Biotechnol.* **2019**, *5*, 59. [[CrossRef](#)]
34. Lewis, K.A.; Tzilivakis, J.; Warner, D.J.; Green, A. An International Database for Pesticide Risk Assessments and Management. *Hum. Ecol. Risk Assess. Int. J.* **2016**, *22*, 1050–1064. [[CrossRef](#)]
35. Cordova-Kreylos, A.L.; Fernandez, L.E.; Koivunen, M.; Yang, A.; Flor-Weiler, L.; Marrone, P.G. Isolation and Characterization of *Burkholderia rinojensis* sp. nov., a Non-*Burkholderia cepacia* Complex Soil Bacterium with Insecticidal and Miticidal Activities. *Appl. Environ. Microbiol.* **2013**, *79*, 7669–7678. [[CrossRef](#)] [[PubMed](#)]
36. Cazenave, A.B.; Shah, K.; Trammell, T.; Komp, M.; Hoffman, J.; Motes, C.M.; Monteros, M.J. High-Throughput Approaches for Phenotyping Alfalfa Germplasm under Abiotic Stress in the Field. *Plant Phenome J.* **2019**, *2*, 1–13. [[CrossRef](#)]
37. Montandon, L.M.; Small, E.E. The Impact of Soil Reflectance on the Quantification of the Green Vegetation Fraction from NDVI. *Remote Sens. Environ.* **2008**, *112*, 1835–1845. [[CrossRef](#)]
38. Chandel, A.K.; Khot, L.R.; Yu, L.-X. Alfalfa (*Medicago sativa* L.) Crop Vigor and Yield Characterization Using High-Resolution Aerial Multispectral and Thermal Infrared Imaging Technique. *Comput. Electron. Agric.* **2021**, *182*, 105999. [[CrossRef](#)]
39. Kasim, N.; Shi, Q.; Wang, J.; Sawut, R.; Nurmemet, I.; Isak, G. Estimation of Spring Wheat Chlorophyll Content Based on Hyperspectral Features and PLSR Model. *Nongye Gongcheng Xuebao/Trans. Chin. Soc. Agric. Eng.* **2017**, *33*, 208–216.
40. Ramos, A.P.M.; Osco, L.P.; Furuya, D.E.G.; Gonçalves, W.N.; Santana, D.C.; Teodoro, L.P.R.; da Silva Junior, C.A.; Capristo-Silva, G.F.; Li, J.; Baio, F.H.R.; et al. A Random Forest Ranking Approach to Predict Yield in Maize with Uav-Based Vegetation Spectral Indices. *Comput. Electron. Agric.* **2020**, *178*, 105791. [[CrossRef](#)]
41. Abdulridha, J.; Batuman, O.; Ampatzidis, Y. UAV-Based Remote Sensing Technique to Detect Citrus Canker Disease Utilizing Hyperspectral Imaging and Machine Learning. *Remote Sens.* **2019**, *11*, 1373. [[CrossRef](#)]
42. Sharma, P.; Leigh, L.; Chang, J.; Maimaitijiang, M.; Caffé, M. Above-Ground Biomass Estimation in Oats Using UAV Remote Sensing and Machine Learning. *Sensors* **2022**, *22*, 601. [[CrossRef](#)]
43. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [[CrossRef](#)]
44. Zou, J.; Yan, H.; So, S.-S. Overview of Artificial Neural Networks. In *Artificial Neural Networks: Methods and Applications*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 14–22.
45. Mountrakis, G.; Im, J.; Ogole, C. Support Vector Machines in Remote Sensing: A Review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [[CrossRef](#)]
46. Ngie, A.; Ahmed, F. Estimation of Maize Grain Yield Using Multispectral Satellite Data Sets (SPOT 5) and the Random Forest Algorithm. *S. Afr. J. Geomat.* **2018**, *7*, 11–30. [[CrossRef](#)]

47. Lei, S.; Luo, J.; Tao, X.; Qiu, Z. Remote Sensing Detecting of Yellow Leaf Disease of Arecanut Based on UAV Multisource Sensors. *Remote Sens.* **2021**, *13*, 4562. [[CrossRef](#)]
48. Joalland, S.; Screpanti, C.; Liebisch, F.; Varella, H.V.; Gaume, A.; Walter, A. Comparison of Visible Imaging, Thermography and Spectrometry Methods to Evaluate the Effect of *Heterodera schachtii* Inoculation on Sugar Beets. *Plant Methods* **2017**, *13*, 1–14. [[CrossRef](#)]
49. Ranjan, R.; Chandel, A.K.; Khot, L.R.; Bahlol, H.Y.; Zhou, J.; Boydston, R.A.; Miklas, P.N. Irrigated Pinto Bean Crop Stress and Yield Assessment Using Ground Based Low Altitude Remote Sensing Technology. *Inf. Process. Agric.* **2019**, *6*, 502–514. [[CrossRef](#)]
50. Yu, N.; Li, L.; Schmitz, N.; Tian, L.F.; Greenberg, J.A.; Diers, B.W. Development of Methods to Improve Soybean Yield Estimation and Predict Plant Maturity with an Unmanned Aerial Vehicle Based Platform. *Remote Sens. Environ.* **2016**, *187*, 91–101. [[CrossRef](#)]
51. Kayad, A.; Sozzi, M.; Gatto, S.; Marinello, F.; Pirotti, F. Monitoring Within-Field Variability of Corn Yield Using Sentinel-2 and Machine Learning Techniques. *Remote Sens.* **2019**, *11*, 2873. [[CrossRef](#)]
52. Gitelson, A.A.; Stark, R.; Grits, U.; Rundquist, D.; Kaufman, Y.; Derry, D. Vegetation and Soil Lines in Visible Spectral Space: A Concept and Technique for Remote Estimation of Vegetation Fraction. *Int. J. Remote Sens.* **2002**, *23*, 2537–2562. [[CrossRef](#)]
53. Xu, J.; Meng, J.; Quackenbush, L.J. Use of Remote Sensing to Predict the Optimal Harvest Date of Corn. *Field Crop. Res.* **2019**, *236*, 1–13. [[CrossRef](#)]
54. Zhang, Y.; Ta, N.; Guo, S.; Chen, Q.; Zhao, L.; Li, F.; Chang, Q. Combining Spectral and Textural Information from UAV RGB Images for Leaf Area Index Monitoring in Kiwifruit Orchard. *Remote Sens.* **2022**, *14*, 1063. [[CrossRef](#)]
55. Zhang, K.; Ge, X.; Shen, P.; Li, W.; Liu, X.; Cao, Q.; Zhu, Y.; Cao, W.; Tian, Y. Predicting Rice Grain Yield Based on Dynamic Changes in Vegetation Indexes during Early to Mid-Growth Stages. *Remote Sens.* **2019**, *11*, 387. [[CrossRef](#)]
56. Kurihara, J.; Nagata, T.; Tomiyama, H. Rice Yield Prediction in Different Growth Environments Using Unmanned Aerial Vehicle-Based Hyperspectral Imaging. *Remote Sens.* **2023**, *15*, 2004. [[CrossRef](#)]
57. Kanke, Y.; Tubaña, B.; Dalen, M.; Harrell, D. Evaluation of Red and Red-Edge Reflectance-Based Vegetation Indices for Rice Biomass and Grain Yield Prediction Models in Paddy Fields. *Precis. Agric.* **2016**, *17*, 507–530. [[CrossRef](#)]
58. Yue, J.; Feng, H.; Yang, G.; Li, Z. A Comparison of Regression Techniques for Estimation of Above-Ground Winter Wheat Biomass Using Near-Surface Spectroscopy. *Remote Sens.* **2018**, *10*, 66. [[CrossRef](#)]
59. Yue, J.; Yang, G.; Tian, Q.; Feng, H.; Xu, K.; Zhou, C. Estimate of Winter-Wheat above-Ground Biomass Based on UAV Ultrahigh-Ground-Resolution Image Textures and Vegetation Indices. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 226–244. [[CrossRef](#)]
60. Hota, S.; Tewari, V.K.; Chandel, A.K. Workload Assessment of Tractor Operations with Ergonomic Transducers and Machine Learning Techniques. *Sensors* **2023**, *23*, 1408. [[CrossRef](#)]
61. Pham, B.T.; Son, L.H.; Hoang, T.-A.; Nguyen, D.-M.; Tien Bui, D. Prediction of Shear Strength of Soft Soil Using Machine Learning Methods. *CATENA* **2018**, *166*, 181–191. [[CrossRef](#)]
62. Richetti, J.; Judge, J.; Boote, K.J.; Johann, J.A.; Uribe-Opazo, M.A.; Becker, W.R.; Paludo, A.; de Albuquerque Silva, L.C. Using Phenology-Based Enhanced Vegetation Index and Machine Learning for Soybean Yield Estimation in Paraná State, Brazil. *J. Appl. Remote Sens.* **2018**, *12*, 1. [[CrossRef](#)]
63. Nguyen, Q.H.; Ly, H.-B.; Ho, L.S.; Al-Ansari, N.; Van Le, H.; Tran, V.Q.; Prakash, I.; Pham, B.T. Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Math. Probl. Eng.* **2021**, *2021*, 1–15. [[CrossRef](#)]
64. Teodoro, P.E.; Teodoro, L.P.R.; Baio, F.H.R.; da Silva Junior, C.A.; dos Santos, R.G.; Ramos, A.P.M.; Pinheiro, M.M.F.; Osco, L.P.; Gonçalves, W.N.; Carneiro, A.M.; et al. Predicting Days to Maturity, Plant Height, and Grain Yield in Soybean: A Machine and Deep Learning Approach Using Multispectral Data. *Remote Sens.* **2021**, *13*, 4632. [[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.