

## Article

# Pre-Harvest Corn Grain Moisture Estimation Using Aerial Multispectral Imagery and Machine Learning Techniques

Pius Jjagwe <sup>1,2</sup> , Abhilash K. Chandel <sup>1,2,\*</sup>  and David Langston <sup>1</sup>

<sup>1</sup> Virginia Tech Tidewater Agricultural Research and Extension Center, Suffolk, VA 23437, USA; [pjjagwe@vt.edu](mailto:pjjagwe@vt.edu) (P.J.); [dlangston@vt.edu](mailto:dlangston@vt.edu) (D.L.)

<sup>2</sup> Department of Biological Systems Engineering, Virginia Tech, Blacksburg, VA 24061, USA

\* Correspondence: [abhilashchandel@vt.edu](mailto:abhilashchandel@vt.edu)

**Abstract:** Corn grain moisture (CGM) is critical to estimate grain maturity status and schedule harvest. Traditional methods for determining CGM range from manual scouting, destructive laboratory analyses, and weather-based dry down estimates. Such methods are either time consuming, expensive, spatially inaccurate, or subjective, therefore they are prone to errors or limitations. Realizing that precision harvest management could be critical for extracting the maximum crop value, this study evaluates the estimation of CGM at a pre-harvest stage using high-resolution (1.3 cm/pixel) multispectral imagery and machine learning techniques. Aerial imagery data were collected in the 2022 cropping season over 116 experimental corn planted plots. A total of 24 vegetation indices (VIs) were derived from imagery data along with reflectance (REF) information in the blue, green, red, red-edge, and near-infrared imaging spectrum that was initially evaluated for inter-correlations as well as subject to principal component analysis (PCA). VIs including the Green Normalized Difference Index (GNDVI), Green Chlorophyll Index (GCI), Infrared Percentage Vegetation Index (IPVI), Simple Ratio Index (SR), Normalized Difference Red-Edge Index (NDRE), and Visible Atmospherically Resistant Index (VARI) had the highest correlations with CGM ( $r$ : 0.68–0.80). Next, two state-of-the-art statistical and four machine learning (ML) models (Stepwise Linear Regression (SLR), Partial Least Squares Regression (PLSR), Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF), and K-nearest neighbor (KNN)), and their 120 derivatives (six ML models  $\times$  two input groups (REFs and REFs+VIs)  $\times$  10 train–test data split ratios (starting 50:50)) were formulated and evaluated for CGM estimation. The CGM estimation accuracy was impacted by the ML model and train–test data split ratio. However, the impact was not significant for the input groups. For validation over the train and entire dataset, RF performed the best at a 95:5 split ratio, and REFs+VIs as the input variables ( $r_{\text{train}}$ : 0.97,  $r\text{RMSE}_{\text{train}}$ : 1.17%,  $r_{\text{entire}}$ : 0.95,  $r\text{RMSE}_{\text{entire}}$ : 1.37%). However, when validated for the test dataset, an increase in the train–test split ratio decreased the performances of the other ML models where SVM performed the best at a 50:50 split ratio ( $r$  = 0.70,  $r\text{RMSE}$  = 2.58%) and with REFs+VIs as the input variables. The 95:5 train–test ratio showed the best performance across all the models, which may be a suitable ratio for relatively smaller or medium-sized datasets. RF was identified to be the most stable and consistent ML model ( $r$ : 0.95,  $r\text{RMSE}$ : 1.37%). Findings in the study indicate that the integration of aerial remote sensing and ML-based data-run techniques could be useful for reliably predicting CGM at the pre-harvest stage, and developing precision corn harvest scheduling and management strategies for the growers.



**Citation:** 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. <https://doi.org/10.3390/land12122188>

Academic Editor: Chuanrong Zhang

Received: 23 November 2023

Revised: 14 December 2023

Accepted: 15 December 2023

Published: 18 December 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/>).

**Keywords:** aerial multispectral sensing; corn grain moisture; machine learning; precision harvest

## 1. Introduction

Grain moisture is critical for determining optimum harvest schedules for crops, which has economic implications during harvest and storage. Markets and safe storages require crops to be harvested at a grain moisture content between 13 to 15.5%, depending on the crop, its variety, and storage duration [1]. Harvesting below this range leads to yield

losses due to grain shrinkage, lodging, and grain dropping during harvest as well as bird interference. On the other hand, harvesting at a moisture level above this range risks fungal infection during storage, requires additional costs and infrastructure for artificial drying, and eventually discounted prices at sales points. Under both situations grain yield, quality, and net returns are at risk [2]. Corn grain moisture (CGM) decreases from about 85% during the silking stage to 30% at around maturity through dehydration [3]. This dehydration occurs in two steps in the field: (i) during maturation, and (ii) post-maturity [4]. As the grain approaches physiological maturity (i.e., maturation dehydration), the assimilates of starch and protein displace water molecules within the grain [1,5,6]. During the post-maturity stage, the grain moisture is lost through exchange with the atmosphere, and this dehydration is influenced by air temperature, relative humidity, and husk weight and thickness [6].

Conventionally, corn growers assess grain moisture indirectly by spotting the milk line and black layer around the grain to determine harvest dates. Among direct methods, cup-shaped capacitive units and portable grain analyzers are used in fields [7]. Another traditional technique is oven drying [4]. Researchers have also developed moisture detection techniques based on the electrical and dielectric characteristics of the grains [8,9]. For non-invasive estimation, generalized growing degree days-based models are used to determine grain moisture and dry down periods (GDDs) [1]. However, this approach provides minimum accountability of localized soil factors, crop varieties, crop management practices, and tillage practices that may impact CGM at spatiotemporal scales. Nonetheless, all these methods are either destructive, time consuming, spatially inaccurate, subjective, or expensive, therefore they are prone to errors or limitations [6–8,10]. Given these limitations, there is a great need for techniques that not only determine CGM non-destructively but are high-throughput in nature as well as account for spatial variability.

Remote sensing is a convenient, timely, high-throughput, and precise technique for the non-destructive assessment of crop physiology and health such as for water [11], chlorophyll or nitrogen, disease infection, and pest infestation, among others, for different crops [12]. This makes remote sensing an extremely useful tool for guiding precision agriculture operations [1]. Pertaining to corn or field crops, research has been maximally restricted to the use of remote sensing with vegetation indices (VIs) or machine learning (ML) techniques for yield predictions [13–16]. Whereas very limited explorations have been conducted for estimating CGM using remote sensing. One study so far utilized satellite-based remote sensing imagery for estimating CGM in China using vegetation indices (VIs) as inputs to the crop-physiological model and observed  $R^2$  values of up to 0.9 [16]. However, satellite-based remote sensing is highly restricted due to fixed data acquisitions, spatial resolutions, and cloud cover issues especially in coastal ecosystems [17–19]. On the other hand, small unmanned aircraft system (SUAS) platforms are widely adopted for precision agriculture operations due to the advantages of providing on-demand data at the desired spatial resolution, and avoidance of atmospheric and cloud interferences [12,15,20,21].

The advancement of data-run techniques such as ML has revolutionized precision agriculture operations significantly in recent years by broadening the horizons for crop health estimations as well as yield forecasting [15]. Some of the most widely used ML models include the support vector machine (SVM), random forest (RF), k-nearest neighbor (KNN), and artificial neural network (ANN), among others [15]. These models deploy an approach of supervised learning, which are trained to approximate complexities between the input and output variables. This enhances robustness and generalizability of ML for estimations compared to other conventional statistical or empirical models. ML is also capable of handling overfitting, remains unaffected by collinearity, number, or non-normal distribution of the variables, and does not require scale normalizations [22].

Given the restricted research of using high-resolution remote sensing and ML mostly for yield predictions, this study addresses an important gap of estimating CGM using SUAS-based multispectral imagery and a range of statistical and ML models. This would eventually help determine precision corn harvest schedules for the growers. It is also

important to note that ML techniques have been restrictively evaluated for the number of variables, and a typical range of training–testing data-split ratios. Most of the studies generally consider 70:30 or 80:20 train–test splits, which may or may not serve for the small–medium datasets [16]. Therefore, this study also evaluates the influence of those two factors on CGM estimation accuracies. Specific objectives are (1) evaluating aerial multispectral imagery-derived reflectance (REFs) and VIs for assessing CGM, (2) estimating CGM using a range of statistical as well as state-of-the-art ML models, and (3) evaluating the performance of those models when using only REFs and a combination of REFs and VIs as input variables at multiple train–test data split ratios. These evaluations will be validated over the entire dataset (100%) as well as train and test datasets independently.

## 2. Materials and Methods

### 2.1. Experimental Details

The study was conducted at an experimental farm of the Tidewater Agriculture Research and Extension Center (TAREC) of Virginia Tech ( $36^{\circ}41'7.22''$  N,  $76^{\circ}45'57.232''$  W), located in Suffolk, VA, USA. The corn seeds were planted between 25–28 April 2022, into a total of 116 plots of 4 rows each that were 30-ft long. These plots were applied with 29 distinct rates and compositions of fungicides at a reproductive growth stage for disease control and to achieve variability in crop vigor for CGM estimation modeling. The crop was harvested on 21 September 2022 (79 DAP (days after planting)) using a plot combine harvester that recorded yield and grain moisture contents for two middle rows of each plot. The combine harvester is equipped with a capacitive-type grain moisture sensor to measure grain moisture and a load sensor to measure yield. No irrigation was applied during the course of the trial.

### 2.2. Aerial Image Acquisition

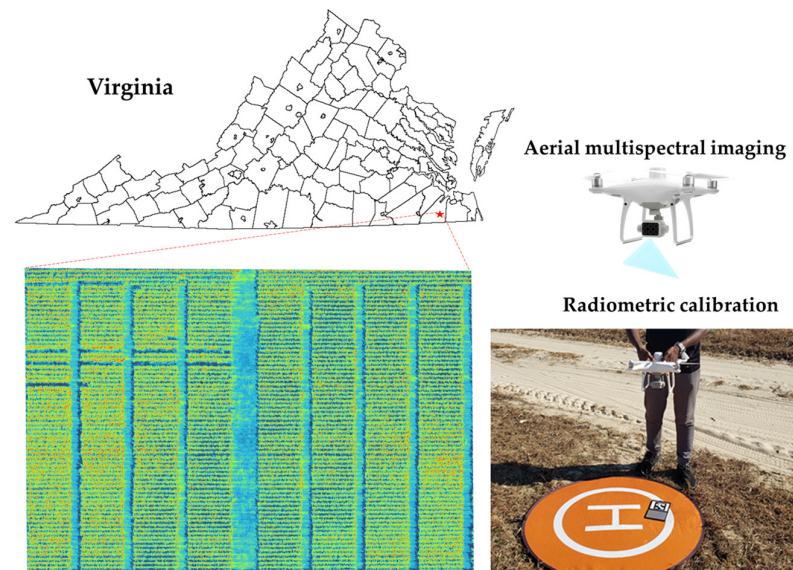
Aerial imagery was acquired at vegetative stage-R5 on August 25, 2022 using a DJI Phantom 4 Multispectral quadcopter drone (SZ DJI Technology Co., Shenzhen, China, Figure 1). Imagery data were acquired earlier than the harvest date (i.e., 21 September 2022) to evaluate the feasibility of CGM estimation before the actual harvest operation was deployed. In addition, this is also the stage after which the crop started senescing. The SUAS was equipped with a five-band multispectral imaging sensor with blue ( $450\text{ nm} \pm 16\text{ nm}$ ), green ( $560\text{ nm} \pm 16\text{ nm}$ ), red ( $650\text{ nm} \pm 16\text{ nm}$ ), red-edge (RE:  $730\text{ nm} \pm 16\text{ nm}$ ), and near-infrared (NIR:  $840\text{ nm} \pm 26\text{ nm}$ ) wavelength sensors of 2.08 megapixels each. DJI Ground Station Pro (DJI GS Pro, version 2.0.17, SZ DJI Technology Co., Shenzhen, China) was used as the ground station control software to set up the SUAS flight mission for an altitude of 25 m above ground level (AGL). This provided multispectral images at a spatial resolution of 1.3 cm/pixel. The multispectral imaging sensor was also configured to capture images at 80% front and 75% side overlaps for seamless orthomosaicing during stitching operations. The SUAS had a real time kinematic (RTK) sensor to receive geolocation corrections for each image as well as a skyward facing downwelling light sensor to record light irradiance during each capture. This light information is used along with the images of a calibrated reflectance panel ( $6\times$ , Sentra, Inc., St. Paul, MN, USA) that were captured after each flight for radiometric calibration of imagery from the mission. This process eliminates any inconsistencies induced within images due to sunlight fluctuations during the flight mission (Figure 1). The imaging flight was conducted near solar ( $\pm 2\text{ h}$ ) noon period for high-quality crop feature retrieval. The SUAS has an SD card for the storage of acquired imagery.

### 2.3. Image Analysis

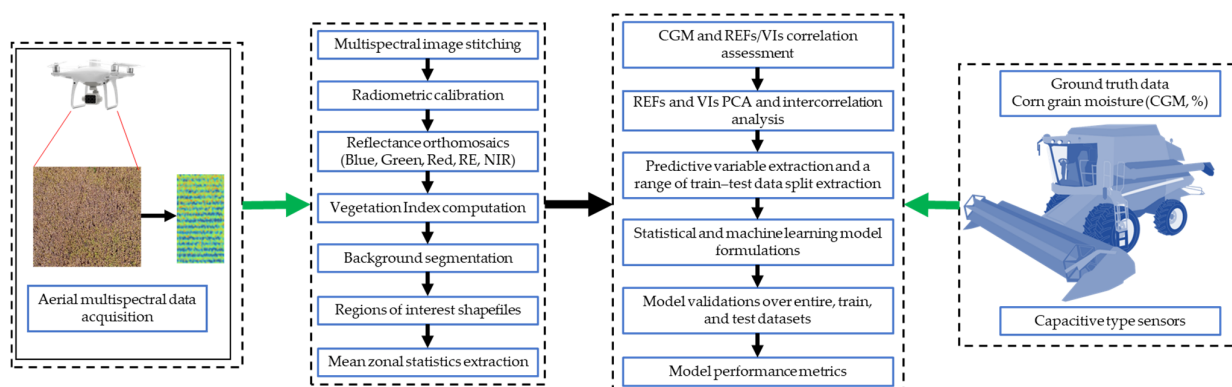
#### Pre-Processing and Feature Extraction

Initially, multispectral snapshots (1125 images: 225 per waveband) were transferred from the SUAS SD card to a photogrammetry and mapping software platform (Pix4D Mapper, Pix4D, Inc., Lausanne, Switzerland). In this platform five seamless multispectral reflectance orthomosaics pertaining to each type of sensor (blue, green, red, RE, NIR) were

obtained as a result of sequential image stitching operations (Figure 2), which include keypoint feature extraction and matching, imagery optimization, georectification, point cloud generation, orthomosaicing, and radiometric calibration.



**Figure 1.** Corn trial plots at Tidewater Agricultural Research and Extension Center in Suffolk, VA, imaged using aerial multispectral platform.



**Figure 2.** Flowchart showing steps of aerial multispectral image analysis and estimation of corn grain moisture using statistical and machine learning model.

The obtained REF orthomosaics were further processed in QGIS using the “Raster Calculator” toolbar (Figure 2) to obtain 24 VIs (Table 1). These VIs were selected for their significance reported in characterizing crop health under a broad range of growth and agroclimatic conditions. The soil background was segmented out from each VI raster using the histogram separation method [23,24]. Next, a shapefile polygon layer was created, where rectangular areas of interest (AOI) of equal dimensions were drawn around the two central rows of each trial plot. Using this shapefile and the “Zonal Statistics” toolbar, mean REF and VI values for each AOI (of all non-zero and not-a-number pixels) were extracted, which were then exported in the “\*.xls” format for further analysis (Figure 2).



**Table 1.** Vegetation indices extracted from aerial multispectral imagery for corn grain moisture assessments.

Vegetation Index	Equation	Reference
Normalized Difference Vegetation Index (NDVI)	$(\text{NIR} - \text{R})/(\text{NIR} + \text{R})$	[25]
Infrared Percentage Vegetation Index (IPVI)	$(\text{NIR})/(\text{NIR} + \text{R})$	[26]
Green Normal Difference Vegetation Index (GNDVI)	$(\text{NIR} - \text{G})/(\text{NIR} + \text{G})$	[27]
Green Difference Vegetation Index (GDVI)	$\text{NIR} - \text{G}$	[28]
Enhanced Vegetation Index (EVI)	$2.5 \times (\text{NIR} - \text{R})/(\text{NIR} + 6 \times \text{R} - 7.5 \times \text{B} + 1)$	[29]
Leaf Area Index (LAI)	$3.618 \times \text{EVI} - 0.118$	[30]
Modified Non-Linear Index (MNLI)	$(\text{NIR}^2 - \text{R}) \times (1 + \text{L})/(\text{NIR}^2 + \text{R} + \text{L})$	[31]
Soil Adjusted Vegetation Index (SAVI)	$1.5 \times (\text{NIR} - \text{R})/(\text{NIR} + \text{R} + 0.5)$	[32]
Optimized Soil Adjusted Vegetation Index (OSAVI)	$(\text{NIR} - \text{R})/(\text{NIR} + \text{R} + 0.16)$	[33]
Green Soil Adjusted Vegetation Index (GSAVI)	$(\text{NIR} - \text{G})/(\text{NIR} + \text{G} + 0.5)$	[32]
Green Optimized Soil Adjusted Vegetation Index (GOSAVI)	$(\text{NIR} - \text{G})/(\text{NIR} + \text{G} + 0.16)$	[32]
Modified Soil Adjusted Vegetation Index (MSAVI2)	$(2 \times \text{NIR} + 1 - \sqrt{(2 \times \text{NIR} + 1)^2 - 8 \times (\text{NIR} - \text{R})})/2$	[34]
Normalized Difference Red-edge Index (NDRE)	$(\text{NIR} - \text{RE})/(\text{NIR} + \text{RE})$	[35]
Green Ratio Vegetation Index (GRVI)	$\text{NIR}/\text{G}$	[28]
Green Chlorophyll Index (GCI)	$(\text{NIR}/\text{G}) - 1$	[36]
Green Leaf Index (GLI)	$((\text{G} - \text{R}) + (\text{G} - \text{B})) / ((2 \times \text{G}) + \text{R} + \text{B})$	[37]
Simple Ratio (SR)	$\text{NIR}/\text{R}$	[38]
Modified Simple Ratio (MSR)	$((\text{NIR}/\text{R}) - 1) / (\sqrt{(\text{NIR}/\text{R})} + 1)$	[39]
Renormalized Difference Vegetation Index (RDVI)	$(\text{NIR} - \text{R})/\sqrt{(\text{NIR} + \text{R})}$	[40]
Transformed Difference Vegetation Index (TDVI)	$1.5 \times ((\text{NIR} - \text{R})/\sqrt{(\text{NIR} + \text{R} + 0.5)})$	[41]
Visible Atmospherically Resistant Index (VARI)	$(\text{G} - \text{R})/(\text{G} + \text{R} - \text{B})$	[42]
Wide Dynamic Range Vegetation Index (WDRVI)	$(a \times \text{NIR} - \text{R})/(a \times \text{NIR} + \text{R})$	[43]

R, G, B, RE, and NIR are pixel values of the spectral responses in red, green, blue, red-edge, and near-infrared images.

#### 2.4. Data Analysis and CGM Estimation

A dataset containing CGM measurements (%) along with five REF and 24 VI features was derived for 116 plots. Firstly, data normality was checked, and all the data followed a normal distribution. Then, a Pearson correlation analysis was conducted to identify the association between the CGM and all the derived REF and VI features.

Next, four ML models and two statistical models were formulated for CGM estimation. These models include stepwise linear regression (SLR), partial least-squares regression (PLSR), random forest (RF), k-nearest neighbor (KNN), support vector machine (SVM), and artificial neural network (ANN). In SLR, the variable with the maximum sum of squares of regression is selected first, and then binary regression is formed by selecting an additional variable from the remaining variables. This process repeats until all non-significant variables are eliminated that could induce confounding effects [44,45]. PLSR combines basic multiple linear regression functions and performs correlation and PCA to eliminate collinearity between variables and maintains relationships with dependent variables, i.e., CGM [46,47]. PLSR also has the capability of avoiding non-normal data. RF is a highly used ML model for agricultural operations that assembles multiple decision trees to estimate a result. The strength of RF is its ability to handle complex datasets and mitigate overfitting for predictive modeling. In this study, the RF model was initially tested with 1000 trees for all dependent variables and optimum trees were identified in the ranges of 300–400 where the prediction accuracy was almost saturated. This hyperparameter tuning was achieved by setting “five variables selected at random” as candidates for each iteration of tuning [48]. KNN performs its function by approximating the association between the independent and dependent variables by averaging the observations in the same neighborhood. In this study, for KNN, repeated cross validation was adopted with three repeats or iterations for up to 30 neighbors. Once the least mean square error was obtained for a particular number of neighbors, that number was used for final model training [49]. SVM identifies a hyperplane in an n-dimensional space that distinctly classifies the data points. This hyperplane is developed iteratively such that the misclassification error is minimal while predicting continuous outputs [50,51]. ANN is a supervised ML model that comprises node layers, namely, an input layer, one or more hidden layers, and an output layer. The structure of ANN is inspired by the brain where each node connects to another with an associated weight and threshold. If the output of any node is above the threshold, that node gets activated and sends the data to the next layer of the network. This process repeats for user-defined iterations until the network’s output error reaches the desired value [50,52]. The

major advantage of ANN over other statistical or linear models is that it flexibly computes the complicated or non-linear relationships between the input and the outputs. In this study, two hidden layers were selected with ten and three nodes, respectively.

Prior to implementing these models, significant variables that would be used as inputs were identified among the derived 29 REF and VI features. This was completed to complement reduced overfitting and enhanced robustness of ML models for CGM estimation. For this, firstly a principal component analysis was conducted to identify the collinear variables. Two primary axes that explained the main variability, intercorrelations, and dominating pattern of VIs in the data matrix were used to generate the PCA biplots for dimensionality reduction. Next, a pair-wise correlation analysis was conducted between all REF and VI features to reduce the number of variables. A threshold of 0.99 was defined in this pair-wise correlation analysis and variables with correlations above this threshold were identified and variables with largest mean absolute correlation were removed.

In the next step, two groups of input variables, (1) REFs and (2) REFs+VIs, as well as ten training–testing datasets were defined. These training–testing datasets were based on ten split ratios starting from 50:50 up until 95:5 at a 5% increment of the training dataset. These sets of train–test splits were developed to identify and evaluate appropriate training data sizes for the best model performance, especially for small- to medium-sized datasets as in this study (i.e., total 116 data points). For evaluating the estimation model performances, the trained models were implemented on the entire dataset, the testing dataset, as well as the training dataset. Metrics of Pearson correlation ( $r$ ) and relative root mean square error (rRMSE, %, Equation (1)) were computed to evaluate the model performance and accuracy of CGM estimation. All the ML and statistical modeling, metrics ( $r$  and rRMSE) computations, and other analyses were performed with the R statistical computing software (version 4.3.1; RStudio, Inc. Boston, MA, USA) with all statistical analyses inferred at 5% significance.

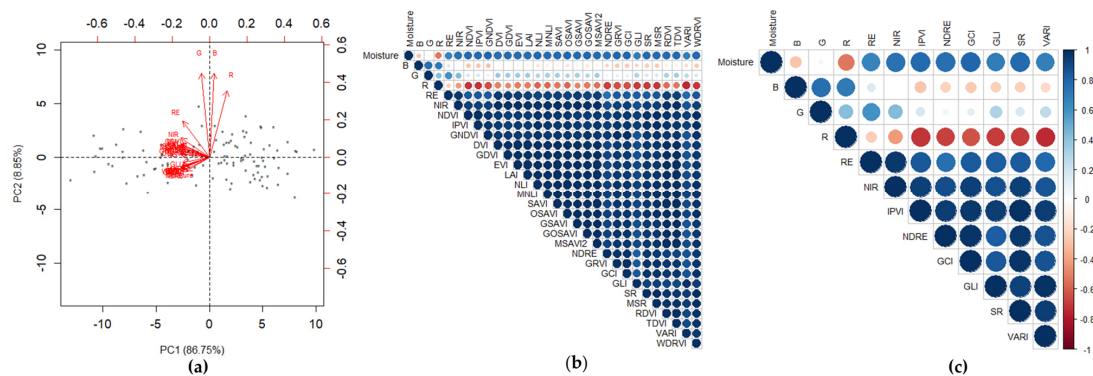
$$\text{rRMSE (\%)} = 100 \times \frac{\sqrt{\frac{\sum_{i=1}^n (\text{CGM}_E - \text{CGM}_m)^2}{n}}}{\text{mean}(\text{CGM}_m)} \quad (1)$$

where  $\text{CGM}_E$  is the estimated CGM and  $\text{CGM}_m$  is the measured CGM.

### 3. Results

#### 3.1. Crop Reflectance and Vegetation Index Feature Evaluation

Pearson's correlation ( $r$ ) analysis (Figure 3 and Table 2) showed that CGM had strong and significant correlations with REFs in RE, and NIR, and the derived 24 VIs (0.68–0.80). The correlation with REFs in the red band was moderate ( $r = -0.52$ ). Among the VIs, the highest correlation was observed for GNDVI ( $r = 0.80$ ) and the lowest for VARI (0.68). Correlations with the REFs in blue and green wavebands were the lowest ( $-0.27$  and  $0.05$ ).



**Figure 3.** (a) Principal component analysis biplot of 24 vegetation indices and five reflectance features accounting for a total of 95.60% of the variability in the data, (b) intercorrelation heat map between spectral features, and (c) final selected input features after dimensionality reduction.

**Table 2.** Correlations of reflectance and vegetation indices with corn grain moisture.

Vegetation Index	Pearson Correlation (r)
Blue	−0.27
Green	0.05
Red	−0.52
Red Edge	0.66
Near Infrared	0.74
Normalized Difference Vegetation Index (NDVI)	0.77
Infrared Percentage Vegetation Index (IPVI)	0.77
Green Normal Difference Vegetation Index (GNDVI)	0.80
Difference Vegetation Index (DVI)	0.76
Green Difference Vegetation Index (GDVI)	0.76
Enhanced Vegetation Index (EVI)	0.77
Leaf Area Index (LAI)	0.77
Non-Linear Index (NLI)	0.78
Modified Non-Linear Index (MNLI)	0.76
Soil Adjusted Vegetation Index (SAVI)	0.77
Optimized Soil Adjusted Vegetation Index (OSAVI)	0.78
Green Soil Adjusted Vegetation Index (GSAVI)	0.78
Green Optimized Soil Adjusted Vegetation Index (GOSAVI)	0.79
Modified Soil Adjusted Vegetation Index (MSAVI2)	0.77
Normalized Difference Red-edge Index (NDRE)	0.76
Green Ratio Vegetation Index (GRVI)	0.79
Green Chlorophyll Index (GCI)	0.79
Green Leaf Index (GLI)	0.69
Simple Ratio (SR)	0.77
Modified Simple Ratio (MSR)	0.78
Renormalized Difference Vegetation Index (RDVI)	0.77
Transformed Difference Vegetation Index (TDVI)	0.78
Visible Atmospherically Resistant Index (VARI)	0.68
Wide Dynamic Range Vegetation Index (WDRVI)	0.78

R, G, B, RE, and NIR are reflectance in red, green, blue, red-edge, and NIR images. Correlation coefficients are significant at  $p < 0.001$ .

### 3.2. Non-Invasive CGM Estimation with ML

#### 3.2.1. Input Feature Selection

In the PCA, two primary PCs comprising 24 VIs and five REFs accounted for the variability of 86.75% and 8.85% (Total = 95.60%, (Figure 3a)). The eigenvectors for the REF in blue, green, and red wavelengths tended towards the top of the biplot (Figure 3a), so they could be inferred to have more influence on PC-2 while the REFs in RE and NIR wavelengths, as well as all other VIs, formed a dense cluster towards the extreme left, top-left, or lower-left region, so they could be inferred to have more influence on PC-1. The PCA could also visualize numerous VIs that completely coincided or were colinear with other VIs. This observation was also supported by Figure 3b that shows complete intercorrelations (i.e.,  $r = 1$ ) between such VIs. Next, using the function “findCorrelation” in RStudio, we were able to identify the groups of VIs that had complete intercorrelations and among them drop VIs that had the largest mean absolute correlation. The function considers the absolute values of pair-wise correlations between variables and removes the variable with the largest mean absolute correlation. This is similar to removing variables that have lower loadings (determined through PCA) or less representation of the variability in data compared to its collinear variable(s). The process determined five REFs and six VIs that were not collinear and included B, G, R, RE, NIR, IPVI, NDRE, GCI, GLI, SR, and VARI. These were finally used for CGM estimation through statistical and ML models (Figure 3c).

#### 3.2.2. Using Reflectance Features as Inputs

The CGM through statistical and ML models was initially estimated using only the REF features as the predictor variables (Table 3). For validation over the test dataset,

SLR performed the best at a 50:50 split ( $r = 0.74$ ,  $rRMSE = 2.43\%$ ), followed by PLSR, SVM (50:50), RF, and KNN, and ANN was the weakest performer at the same split ratio ( $r = 0.61$ ,  $rRMSE = 4.43\%$ ). For validation over the train dataset, RF performed best at a 95:5 split ( $r = 0.96$ ,  $rRMSE = 1.31\%$ ), followed by ANN (70:30), PLSR (75:25), KNN (75:25), and SVM (70:30). SLR was the weakest performer at a 75:25 split ( $r = 0.8$ ,  $rRMSE = 2.37\%$ ). For validation over the entire dataset, RF ( $r = 0.94$ ,  $rRMSE = 1.51\%$ ) performed the best followed by ANN, PLSR, SLR, SVM, while KNN ( $r = 0.79$ ,  $rRMSE = 2.6\%$ ) was the weakest at a 95:5 split.

**Table 3.** Comparison of model analysis using reflectance and a combination of reflectance and VIs at different train–test ratios.

Parameters		Dataset: Entire			Dataset: Test			Dataset: Train		
Train:Test Ratio	Input Group	Best Model	$r$	$rRMSE$ (%)	Best Model	$r$	$rRMSE$ (%)	Best Model	$r$	$rRMSE$ (%)
50:50	REFs	RF	0.86	2.14	SLR	0.74	2.43	RF	0.96	1.59
	REFs+VIs		0.87	2.08	SVM	0.70	2.58		0.97	1.34
55:45	REFs	RF	0.87	2.11	SLR	0.74	2.51	RF	0.96	1.47
	REFs+VIs		0.87	2.08	SVM	0.69	2.68	ANN	0.97	1.26
60:40	REFs	RF	0.88	2.05	SLR	0.70	2.27	RF	0.96	1.47
	REFs+VIs		0.88	2.03	SVM	0.67	2.64		0.97	1.26
65:35	REFs	RF	0.88	2.02	SLR	0.66	2.67	RF	0.96	1.41
	REFs+VIs		0.88	2.02	SVM	0.64	2.78		0.97	1.22
70:30	REFs	RF	0.89	1.95	SLR	0.61	2.76	RF	0.96	1.43
	REFs+VIs		0.89	1.93	SVM	0.60	2.92		0.97	1.21
75:25	REFs	RF	0.89	1.92	ANN	0.62	2.82	RF	0.96	1.35
	REFs+VIs		0.90	1.86	SVM	0.60	3.08		0.97	1.17
80:20	REFs	RF	0.91	1.86	PLSR	0.65	2.70	RF	0.96	1.34
	REFs+VIs		0.92	1.73	SLR	0.71	2.74		0.96	1.21
85:15	REFs	RF	0.93	1.69	PLSR	0.62	2.82	RF	0.96	1.33
	REFs+VIs		0.92	1.65	SLR	0.67	2.69		0.97	1.20
90:10	REFs	RF	0.94	1.55	PLSR	0.43	2.91	RF	0.96	1.32
	REFs+VIs		0.94	1.45	SLR	0.51	2.84		0.97	1.16
95:5	REFs	RF	0.94	1.51	KNN	0.69	3.25	RF	0.96	1.31
	REFs+VIs		0.95	1.37	SLR	0.77	2.59		0.97	1.17

REFs is the reflectance-only input group, REFs+VIs is the reflectance and selected vegetation indices input group.

### 3.2.3. Using Reflectance and Vegetation Index Features as Inputs

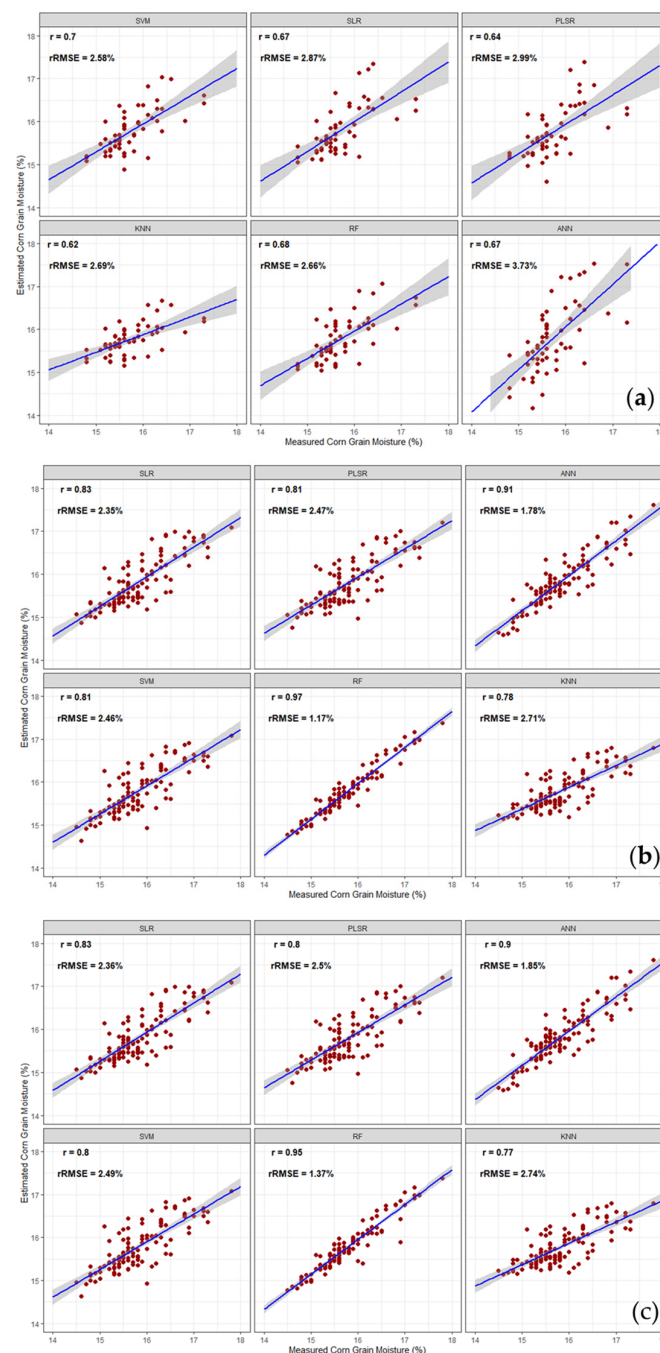
In the second stage of CGM estimation, five selected REFs and six VIs as a result of dimensionality reduction process were used as the inputs (Table 3). For the validation over the test dataset, SVM performed the best at a 50:50 split ( $r = 0.70$ ,  $rRMSE = 2.58\%$ ), followed by RF, SLR, ANN, and PLSR, and KNN was the weakest performer at the same 50:50 split ( $r = 0.62$ ,  $rRMSE = 2.69\%$ ). For the validation over the train dataset, RF performed best ( $r = 0.97$ ,  $rRMSE = 1.17\%$ ), followed by ANN, SLR, PLSR, and SVM at a 95:5 split while PLSR was the weakest performer at that split ( $r = 0.82$ ,  $rRMSE = 2.54\%$ ). For the validation over the entire dataset, RF at a 95:5 split ( $r = 0.95$ ,  $rRMSE = 1.37\%$ ) performed best followed by ANN, SLR, SVM, and PLSR, while KNN ( $r = 0.77$ ,  $rRMSE = 2.74\%$ ) was the weakest performer at a 95:5 split.

### 3.2.4. Impact of Training and Testing Data Split Ratios

As the training dataset size increased for training the models, the CGM estimation accuracy also increased when validated over the train and entire datasets ( $r_{\text{train}}$ : 0.61–0.97,  $rRMSE_{\text{train}}$ : 1.15–2.86%,  $r_{\text{entire}}$ : 0.76–0.95,  $rRMSE_{\text{entire}}$ : 1.37–3.31%) also increased (Figures 4b,c and 5, Table 3) and decreased when validated over the test dataset ( $r_{\text{test}}$ : 0.61–0.77,  $rRMSE_{\text{test}}$ : 2.27–5.59%, Figures 4a and 5, Table 3). For the train–test split ratio of 95:5, the accuracy of CGM estimation was the best and RF was the best performing model when validated over the train and entire datasets ( $r_{\text{train}} = 0.97$ ,  $rRMSE_{\text{train}} = 1.17\%$ ,  $r_{\text{entire}} = 0.95$ ,  $rRMSE_{\text{entire}} = 1.37\%$ , Figure 4b,c and Figure 5, Table 3) with REFs+VIs as the input group. In addition, SLR performed the best at a 95:5 split ratio when validated over



the test dataset with REFs+VIs as the input group. At the same split ratio, even when using REFs as the input group and validation over train and entire datasets, RF performed the best ( $r_{\text{train}} = 0.96$ ,  $r\text{RMSE}_{\text{train}} = 1.31\%$ ,  $r_{\text{entire}} = 0.94$ ,  $r\text{RMSE}_{\text{entire}} = 1.51\%$ ). SLR performed the best ( $r = 0.74$ ,  $r\text{RMSE} = 2.43\%$ ) with REFs as the input group and SVM performed the best ( $r = 0.70$ ,  $r\text{RMSE} = 2.58\%$ ) with REFs+VIs as the input group for the train–test split ratio of 50:50 when those models were validated over the test dataset. ANN also improved its performance at a train–test split ratio of 55:45 when using REFs+VIs as the input group and was validated over the train dataset ( $r_{\text{train}} = 0.97$ ,  $r\text{RMSE}_{\text{train}} = 1.26\%$ , Table 3). When validated over the test dataset, ANN improved its performance for the split ratio of 75:25 and with REFs as inputs ( $r_{\text{test}} = 0.62$ ,  $r\text{RMSE}_{\text{test}} = 2.82\%$ ).



**Figure 4.** Plots showing measured and estimated CGM using REFs+VIs as input group for models validated over (a) the test dataset at 50:50, (b) train dataset at 95:5, and (c) entire dataset at 95:5 splits.



**Figure 5.** Plots of (a) Pearson correlation ( $r$ ), and (b) Relative root mean square error (rRMSE) summarizing the performance of six corn grain moisture estimation models for ten train–test data split ratios, and for two input groups (REFs, REFs+VIs) over three validation datasets (entire, test, train).

#### 4. Discussion

Among the REFs in five wavebands, NIR had the highest correlation with CGM ( $r = 0.74$ ) depicting sensitivity to the chlorophyll light absorption feature of plants [53]. Correlations with REF in blue and green ( $r = -0.27$ ,  $r = 0.05$ ) wavebands were the lowest, possibly due to a low signal-to-noise ratio [44,54]. Among the total 24 derived VIs, GNDVI had the highest correlation with CGM, followed by GCI, GRVI, GOSAVI, and WDRVI, among others ( $r = 0.78$ – $0.80$ ), while VARI had the lowest correlation ( $r = 0.68$ ). GNDVI is derived using NIR ( $840 \pm 20$  nm), which is more sensitive to chlorophyll content, supporting a strong correlation [44]. On the other hand, VARI had low correlation due to its nonlinear mathematical operation as well as its derivation using blue and green wavebands that had low correlations with CGM. VIs such as IPVI and GCI had stronger correlations with CGM as those take into consideration the dynamic variations in the visible–NIR region pertaining to canopy water, chlorophyll, and nitrogen contents [55]. In this study, GCI and GNDVI outperformed VIs that use reflectance in the red band such as IPVI, NDVI, TDVI, RDVI, and others, as the reflectance in the green band ( $560 \pm 10$  nm) is relatively more sensitive to chlorophyll and crop moisture contents [16,42]. This was also corroborated by observations made by Kayad et al. [56] where VIs computed using green band reflectance outperformed others in estimating corn grain yield. These VIs may also perform well for CGM estimation using simple or multiple linear regression or other statistical models (Table 2) as also supported by previous studies [16]. Nonetheless, using REF or VI feature as independent inputs may lack robustness when evaluated under other agroclimatic conditions [57,58]. Therefore, this study advanced research towards the estimation of CGM using statistical and ML models as those have the capability to robustly approximate complex and non-linear relationships between the inputs (VIs or REFs) and outputs (i.e., CGM).

From the process of conducting PCA and eliminating collinear variables, blue, green, red, RE, NIR, IPVI, NDRE, GCI, GLI, SR, and VARI were identified as not to have absolute correlations (i.e.,  $r = 1$ ) among each other. IPVI had collinearity with NDVI but was selected over the latter for its capability to overcome the limitations of NDVI, which can become saturated for higher biomass, and is also subjected to relatively higher noise from atmospheric and soil background conditions [19]. Studies have also reported IPVI

to perform superior to NDVI for estimating crop nitrogen status and grain yield across different growth stages [21,59]. Although GNDVI had a higher correlation with CGM compared to GCI, it was not selected most probably due to it having a higher mean absolute correlation compared to GCI [60]. An only study conducted thus far for CGM estimation reported canopy chlorophyll content representing LAI as a strong input variable [16]. Most of the crop health status estimations such as chlorophyll content, water content, or yield have utilized not only the REFs as inputs to ML or statistical models but also the VIs [16,58]. Studies that have performed predictive modeling using ML have neither evaluated inter-correlations between the input variables (i.e., VIs) nor eliminated the collinear variables before estimating the output [61,62]. This may often lead to model overfittings, compromised robustness, and require extensive computations from a user's practical standpoint [63]. Most of the ML-based prediction studies have by default utilized 70:30 or 80:20 as the train–test data split ratios for model training and validations [5,64]. However, the consideration of the entire data size as well as impact of varying training data proportions to identify the best train–test data split ratio have been minimally assessed. This may also impact model over-fitness and robustness [50].

For these reasons as well as by identifying our dataset to be of medium size, our study not only eliminated the collinear input variables but also identified the best train–test split ratio(s) for statistical or ML models for CGM estimation. It was observed that model performances improved (Figure 5) when validated over the train and entire datasets, for increasing proportions of training data [16,62]. This observation was consistent when using both input groups, REFs or REFs+VIs. Although REFs+VIs improved the model performance compared to REFs as inputs, the impact was not significant ( $p = 0.374$ , Table 4). Apparently, in the maximum cases irrespective of both input groups, ML models outperformed statistical models for estimating CGM (Figure 5) when validated over the training and entire datasets.

**Table 4.** Effect of input parameters on performance of models for corn grain moisture estimation.

Variable	<i>p</i> Value ( <i>r</i> )	<i>p</i> Value (rRMSE)
Model	<0.001	<0.001
Train–test split	<0.001	0.619
Dataset	<0.001	<0.001
Input group	0.374	0.725
Train–test split: Dataset	<0.001	<0.001
Train–test split: Input group	0.189	0.290
Dataset: Input group	0.450	0.002
Train–test split: Dataset: Input group	0.204	0.544

Where Dataset (train, test, entire), Input groups (REF, REF+VIs), and Model (SLR, PLSR, ANN, SVM, RF, KNN).

Interestingly, SVM and SLR at a 50:50 split ratio when validated over test datasets using either of the input groups performed the best as those have been reported for smaller datasets [50]. SVM is computationally expensive to work with large data as the algorithm often fails while determining optimum boundary hyperplanes, making it more accurate and robust for small data sizes, which has also been supported by other studies [65,66]. By removing the collinearity of inputs, the confounding effects on the estimation of CGM was also eliminated, thereby improving performances of statistical models such as SLR and PLSR [67]. This study's data size was relatively small compared to what ANNs generally require, and this was the reason why ANN was the least good performer amongst all other evaluated models in this study [50,62,68]. Overall, RF performed the best compared to all other models as it is capable of withstanding the overfitting problem unlike other statistical linear models, and it was relatively less reliant on dataset size compared to other ML models [53,69]. This is because RF is a decision-tree-based model that employs several

sub-models and bagging techniques, for increased stability and resilience of the prediction outcomes [70].

This study demonstrated the feasibility of SUAS and integrated ML techniques for CGM estimation, which has not been explored thus far. The performance of the model may further be improved by collecting data over multiple cropping seasons as well as agroclimatic conditions. Identifying the earliest stage where accurate CGM could be predicted as well as their translation to satellite imaging platforms are the next goals in our efforts. Those estimates can be later converted into maps (raster or shapefiles) for the corn growers who can develop precision harvest scheduling and management strategies for enhanced crop value.

## 5. Conclusions

This study investigated the use of aerial multispectral imagery for assessing CGM as well as estimating it using state-of-the-art ML and statistical models. To the best of our knowledge, this was the first investigation of its kind to estimate grain moisture contents. Using Pearson correlation, the REFs and VIs derived from the SUAS imagery data were found to have a strong correlation between CGM and GNDVI, GCI, and IPVI, among others with the highest correlations ( $r$ : 0.68–0.80). PCA and pairwise correlation analysis identified REFs in blue, green, red, RE, NIR and VIs such as GCI, IPVI, NDRE, GLL, SR, and VARI as potential inputs to estimate CGM using statistical and ML models.

All four evaluated ML models and two statistical models for estimating CGM improved in performance with the increase in size of training datasets. While most ML models performed well overall, RF was observed to be the most stable ( $r$ : 0.86–0.97,  $rRMSE$ : 2.14–1.17%). It was observed that the input groups (only REFs or REFs+VIs) for CGM estimation did not impact model performances. However, the train–test split ratios did impact the model performances significantly with 50:50, 50:45, 60:40, 80:20, and 95:5 being among the split ratios that yielded strong performances. The 95:5 train–test split ratio was the best when models were validated over the train and entire datasets while the 50:50 split ratio was the best when models were validated over the test dataset. The statistical models i.e., SLR and PLSR, also yielded comparable performances to most of the ML models ( $r$ : 0.61–0.74,  $rRMSE$ : 2.76–2.43%) while ANN could not be the best-performing model of the study except at a 55:45 split ratio and for validation over train and entire datasets.

Overall, our study demonstrated that aerial multispectral imagery when integrated with ML models could suitably estimate CGM even for small–medium dataset sizes. These computations are critical for the corn growers to non-invasively as well as spatially map CGM status for scheduling and managing harvest schedules and resources. We will be further evaluating the models tested in the study over different growth stages to identify the earliest time when CGM near optimum harvest could be estimated. Moreover, these models could be translated in the form of webtools that farmers could utilize for planning and executing precision operations on the ground and for extracting the best economic value of their crop.

**Author Contributions:** Conceptualization, P.J. and A.K.C.; methodology, P.J., A.K.C. and D.L.; software, P.J. and A.K.C.; validation, P.J., A.K.C. and D.L.; formal analysis, P.J. and A.K.C.; investigation, P.J., A.K.C. and D.L.; resources, A.K.C. and D.L.; data curation, P.J., A.K.C. and D.L.; writing—original draft preparation, P.J. and A.K.C.; writing—review and editing, P.J., A.K.C. and D.L.; visualization, A.K.C. and D.L.; supervision, A.K.C.; project administration, A.K.C. and D.L.; funding acquisition, A.K.C. and D.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 Project #VA136412, and Faculty startup.

**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 conflict of interest.

## References

- Martinez-Feria, R.A.; Licht, M.A.; Ordóñez, R.A.; Hatfield, J.L.; Coulter, J.A.; Archontoulis, S.V. Evaluating Maize and Soybean Grain Dry-down in the Field with Predictive Algorithms and Genotype-by-Environment Analysis. *Sci. Rep.* **2019**, *9*, 7167. [\[CrossRef\]](#) [\[PubMed\]](#)
- Agvei, B.; Andresen, J.; Singh, M.P. Evaluation of a Handheld Near-Infrared Spectroscopy Sensor for Rapid Corn Kernel Moisture Estimation. *Crop Forage Turfgrass Manag.* **2023**, *9*, e20235. [\[CrossRef\]](#)
- Pordesimo, L.O.; Sokhansanj, S.; Edens, W.C. Moisture and Yield of Corn Stover Fractions before and after Grain Maturity. *Trans. ASAE* **2004**, *47*, 1597–1603. [\[CrossRef\]](#)
- Fan, L.-F.; Chai, Z.-Q.; Zhao, P.-F.; Tian, Z.-F.; Wen, S.-Q.; Li, S.-M.; Wang, Z.-Y.; Huang, L. Nondestructive Measurement of Husk-Covered Corn Kernel Layer Dynamic Moisture Content in the Field. *Comput. Electron. Agric.* **2021**, *182*, 106034. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Maiorano, A.; Fanchini, D.; Donatelli, M. MIMYCS. Moisture, a Process-Based Model of Moisture Content in Developing Maize Kernels. *Eur. J. Agron.* **2014**, *59*, 86–95. [\[CrossRef\]](#)
- Sadaka, S.; Rosentrater, K.A. Tips on Examining the Accuracy of On-Farm Grain Moisture Meters. In *Agriculture and Natural Resources*; UAEX: Fayetteville, AR, USA, 2019; pp. 1–5.
- Nelson, S.O.; Trabelsi, S. A Century of Grain and Seed Moisture Measurement by Sensing Electrical Properties. *Trans. ASABE* **2012**, *55*, 629–636. [\[CrossRef\]](#)
- Soltani, M.; Alimardani, R. Prediction of Corn and Lentil Moisture Content Using Dielectric Properties. *J. Agric. Technol.* **2011**, *7*, 1223–1232.
- Zhang, H.L.; Ma, Q.; Fan, L.F.; Zhao, P.F.; Wang, J.X.; Zhang, X.D.; Zhu, D.H.; Huang, L.; Zhao, D.J.; Wang, Z.Y. Nondestructive in Situ Measurement Method for Kernel Moisture Content in Corn Ear. *Sensors* **2016**, *16*, 2196. [\[CrossRef\]](#)
- Clevers, J.G.P.W.; Kooistra, L.; Schaepman, M.E. Estimating Canopy Water Content Using Hyperspectral Remote Sensing Data. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, 119–125. [\[CrossRef\]](#)
- Croft, H.; Chen, J.M.; Zhang, Y.; Simic, A. Modelling Leaf Chlorophyll Content in Broadleaf and Needle Leaf Canopies from Ground, CASI, Landsat TM 5 and MERIS Reflectance Data. *Remote Sens. Environ.* **2013**, *133*, 128–140. [\[CrossRef\]](#)
- Khanal, S.; Klopfenstein, A.; Kushal, K.C.; Ramarao, V.; Fulton, J.; Douridas, N.; Shearer, S.A. Assessing the Impact of Agricultural Field Traffic on Corn Grain Yield Using Remote Sensing and Machine Learning. *Soil Tillage Res.* **2021**, *208*, 104880. [\[CrossRef\]](#)
- Shajahan, S.; Cho, J.; Guinness, J.; van Aardt, J.; Czymmek, K.J.; Kettersings, Q.M. Corn Grain Yield Prediction and Mapping from Unmanned Aerial System (Uas) Multispectral Imagery. *Remote Sens.* **2021**, *13*, 3948.
- 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. [\[CrossRef\]](#)
- Xu, J.; Meng, J.; Quackenbush, L.J. Use of Remote Sensing to Predict the Optimal Harvest Date of Corn. *Field Crops Res.* **2019**, *236*, 1–13. [\[CrossRef\]](#)
- Zhang, L.; Zhang, H.; Niu, Y.; Han, W. Mapping Maize Water Stress Based on UAV Multispectral Remote Sensing. *Remote Sens.* **2019**, *11*, 605. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Moeinade, S.; Pham, H.; Han, Y.; Dobbels, A.; Hu, G. An Applied Deep Learning Approach for Estimating Soybean Relative Maturity from UAV Imagery to Aid Plant Breeding Decisions. *Mach. Learn. Appl.* **2022**, *7*, 100233. [\[CrossRef\]](#)
- Qi, H.; Wu, Z.; Zhang, L.; Li, J.; Zhou, J.; Jun, Z.; Zhu, B. Monitoring of Peanut Leaves Chlorophyll Content Based on Drone-Based Multispectral Image Feature Extraction. *Comput. Electron. Agric.* **2021**, *187*, 106292. [\[CrossRef\]](#)
- Ali, I.; Greifeneder, F.; Stamenkovic, J.; Neumann, M.; Notarnicola, C. Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data. *Remote Sens.* **2015**, *7*, 16398–16421. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Great Plains with ERTS. *NASA Spec. Publ.* **1974**, *351*, 309–317.
- Crippen, R.E. Calculating the Vegetation Index Faster. *Remote Sens. Environ.* **1990**, *34*, 71–73. [\[CrossRef\]](#)

27. Gitelson, A.A.; Merzlyak, M.N. Remote Sensing of Chlorophyll Concentration in Higher Plant Leaves. *Adv. Space Res.* **1998**, *22*, 689–692. [[CrossRef](#)]
28. Sripada, R.P.; Heiniger, R.W.; White, J.G.; Meijer, A.D. Aerial Color Infrared Photography for Determining Early In-Season Nitrogen Requirements in Corn. *Agron. J.* **2006**, *98*, 968–977. [[CrossRef](#)]
29. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [[CrossRef](#)]
30. Boegh, E.; Soegaard, H.; Broge, N.; Hasager, C.B.; Jensen, N.O.; Schelde, K.; Thomsen, A. Airborne Multispectral Data for Quantifying Leaf Area Index, Nitrogen Concentration, and Photosynthetic Efficiency in Agriculture. *Remote Sens. Environ.* **2002**, *81*, 179–193. [[CrossRef](#)]
31. Yang, Z.; Willis, P.; Mueller, R. Impact of band-ratio enhanced awifs image to crop classification accuracy. In Proceedings of the Pecora 17—The Future of Land Imaging. . . Going Operational, Denver, CO, USA, 18–20 November 2008.
32. Sripada, R.P. *Determining In-Season Nitrogen Requirements for Corn Using Aerial Color-Infrared Photography*; North Carolina State University: Raleigh, NC, USA, 2005.
33. Rondeaux, G.; Steven, M.; Baret, F. Optimization of Soil-Adjusted Vegetation Indices. *Remote Sens. Environ.* **1996**, *55*, 95–107. [[CrossRef](#)]
34. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A Modified Soil Adjusted Vegetation Index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [[CrossRef](#)]
35. Leprieux, C.; Kerr, Y.H.; Pichon, J.M. Critical Assessment of Vegetation Indices from Avhrr in a Semi-Arid Environment. *Int. J. Remote Sens.* **1996**, *17*, 2549–2563. [[CrossRef](#)]
36. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between Leaf Chlorophyll Content and Spectral Reflectance and Algorithms for Non-Destructive Chlorophyll Assessment in Higher Plant Leaves. *J. Plant Physiol.* **2003**, *160*, 271–282. [[CrossRef](#)] [[PubMed](#)]
37. Louhaichi, M.; Borman, M.M.; Johnson, D.E. Spatially Located Platform and Aerial Photography for Documentation of Grazing Impacts on Wheat. *Geocarto Int.* **2001**, *16*, 65–70. [[CrossRef](#)]
38. Birth, G.S.; McVey, G.R. Measuring the Color of Growing Turf with a Reflectance Spectrophotometer 1. *Agron. J.* **1968**, *60*, 640–643. [[CrossRef](#)]
39. Chen, J.M. Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. *Can. J. Remote Sens.* **1996**, *22*, 229–242. [[CrossRef](#)]
40. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [[CrossRef](#)]
41. Bannari, A.; Asalhi, H.; Teillet, P.M. Transformed Difference Vegetation Index (TDVI) for Vegetation Cover Mapping. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Toronto, ON, Canada, 24–28 June 2002; IEEE: Piscataway, NJ, USA, 2002; Volume 5, pp. 3053–3055.
42. 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](#)]
43. Gitelson, A.A. Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *J. Plant Physiol.* **2004**, *161*, 165–173. [[CrossRef](#)]
44. Chandel, A.K.; Khot, L.R.; Yu, L.-X. Alfalfa (*Medicago sativa* L.) Crop Vigor and Yield Characterization Using High-Resolution Aerial 1 Multispectral and Thermal Infrared Imaging Technique. *Comput. Electron. Agric.* **2021**, *182*, 105999. [[CrossRef](#)]
45. Yu, X.; Liu, Q.; Wang, Y.; Liu, X.; Liu, X. Evaluation of MLR and PLSR for Estimating Soil Element Contents Using Visible/near-Infrared Spectroscopy in Apple Orchards on the Jiaodong Peninsula. *Catena* **2016**, *137*, 340–349. [[CrossRef](#)]
46. Wold, S.; Sjostrom, M.; Eriksson, L. PLS-Regression. A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [[CrossRef](#)]
47. Nijat, K.; Shi, Q.; Wang, J.; Rukeya, S.; Ilyas, N.; Gulnur, I. Estimation of Spring Wheat Chlorophyll Content Based on Hyperspectral Features and PLSR Model. *Trans. Chin. Soc. Agric. Eng.* **2017**, *33*, 208–216.
48. Marques Ramos, A.P.; Prado Osco, L.; Elis Garcia Furuya, D.; Nunes Gonçalves, W.; Cordeiro Santana, D.; Pereira Ribeiro Teodoro, L.; Antonio da Silva Junior, C.; Fernando Capristo-Silva, G.; Li, J.; Henrique Rojo Baio, F.; et al. A Random Forest Ranking Approach to Predict Yield in Maize with Uav-Based Vegetation Spectral Indices. *Comput. Electron. Agric.* **2020**, *178*, 105791. [[CrossRef](#)]
49. 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](#)]
50. 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](#)] [[PubMed](#)]
51. Mountrakis, G.; Im, J.; Ogole, C. Support Vector Machines in Remote Sensing: A Review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [[CrossRef](#)]
52. Zou, J.; Han, Y.; So, S.S. Overview of Artificial Neural Networks. In *Artificial Neural Networks: Methods and Applications*; Humana Press: Totowa, NJ, USA, 2009; pp. 14–22.
53. 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. [[CrossRef](#)]

54. Jiang, Z.; Huete, A.R.; Didan, K.; Miura, T. Development of a Two-Band Enhanced Vegetation Index without a Blue Band. *Remote Sens. Environ.* **2008**, *112*, 3833–3845. [[CrossRef](#)]
55. Xue, J.; Su, B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *J. Sens.* **2017**, *2017*, 1353691. [[CrossRef](#)]
56. 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](#)]
57. 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](#)]
58. 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](#)]
59. 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](#)]
60. Kuhn, M. Caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26.
61. Zhou, Y.; Lao, C.; Yang, Y.; Zhang, Z.; Chen, H.; Chen, Y.; Chen, J.; Ning, J.; Yang, N. Diagnosis of Winter-Wheat Water Stress Based on UAV-Borne Multispectral Image Texture and Vegetation Indices. *Agric. Water Manag.* **2021**, *256*, 107076. [[CrossRef](#)]
62. 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](#)]
63. 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](#)]
64. Gill, W.R.; Asae, M. Influence of Compaction Hardening of Soil on Penetration Resistance. *Trans. ASAE* **1968**, *11*, 741–745. [[CrossRef](#)]
65. 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](#)]
66. Adugna, T.; Xu, W.; Fan, J. Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. *Remote Sens.* **2022**, *14*, 574. [[CrossRef](#)]
67. 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](#)]
68. 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*, 4832864. [[CrossRef](#)]
69. Palmer, D.S.; O’Boyle, N.M.; Glen, R.C.; Mitchell, J.B.O. Random Forest Models to Predict Aqueous Solubility. *J. Chem. Inf. Model.* **2007**, *47*, 150–158. [[CrossRef](#)]
70. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* **2006**, *9*, 181–199. [[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.