Smart Agricultural Technology 10 (2025) 100719

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Estimating corn leaf chlorophyll content using airborne multispectral imagery and machine learning

Fengkai Tian [a], Jianfeng Zhou [b,*], Curtis J. Ransom [c], Noel Aloysius [a], Kenneth A. Sudduth [c]

[a] Department of Chemical and Biomedical Engineering, University of Missouri, Columbia, MO 65211, USA
[b] Division of Plant Science and Technology, University of Missouri, Columbia, MO 65211, USA
[c] Cropping Systems & Water Quality Research Unit, USDA-ARS, Columbia, MO 65211, USA

ARTICLE INFO

Keywords:
Corn
Nitrogen application
Leaf chlorophyll content
Remote sensing
Machine learning

ABSTRACT

Chlorophyll is crucial for photosynthesis and impacts plant growth and yield in crops. Accurate estimation of plant health and fertilizer status is essential for effective nitrogen (N) management in corn. However, crop chlorophyll is primarily quantified using handheld sensors, which is time-consuming, labor-intensive, and of low spatial resolution. This study aimed to evaluate an airborne multispectral imaging system in estimating the chlorophyll content of corn leaves at four vegetative growth stages. Three replicates of 12 nitrogen rates (between 0 and 285 kg ha^{-1}) were applied to corn at the V4 vegetative stage (i.e., with four established leaves). Soil apparent electrical conductivity (EC_a) of all test plots was measured before planting and corn leaf chlorophyll content was measured using a commercial handheld chlorophyll meter at four vegetative stages (V8, V9, V11, and V12). A UAV-based multispectral camera collected imagery at the same time as manual readings. Machine learning models developed based on image features derived from UAV images were used to predict leaf chlorophyll content. Results showed that an epsilon support vector regression model built on imagery data across imagery data collected over four growth stages with a sequential forward feature selection achieved the best performance ($R^2 = 0.87$, MAE = 1.80, and RMSE = 2.26 SPAD units). There was no significant difference in the performance of models across the four growth stages. By utilizing the developed model, researchers and growers can effectively map the chlorophyll content of corn leaves at different growth stages, enabling them to make timely and informed management decisions.

1. Introduction

Chlorophyll content in corn (Zea mays L.) is crucial for efficient photosynthesis and serves as an indicator of overall plant health [1]. However, the leaf chlorophyll content can decrease due to disease [2], water stress [3], nutrient deficiencies [4], and unfavorable weather conditions [5]. Therefore, monitoring leaf chlorophyll content is important for timely and accurate field management to maintain plant health. Conventional techniques for quantifying leaf chlorophyll levels involve wet chemistry analysis in a lab, where chlorophyll is extracted using solvents and then measured with a spectrophotometer [6]. However, this method is time-consuming and expensive, making it impractical for quick assessment in the field. Alternatively, handheld optical sensors can be used for non-destructive measurements. These sensors, such as the SPAD-502 chlorophyll meter (Minolta Digital, Osaka, Japan) (SPAD hereafter), use a two-band spectral sensor (650 & 940 nm) to estimate chlorophyll content [7]. This method has been widely used in agricultural research to measure the level of chlorophyll in leaves, monitor plant nutrition status, and create N management recommendations [8,9]. However, the data collection procedure using chlorophyll meters is labor intensive, lacks spatial resolution, and is not suitable for large-scale applications.

Remote sensing technology using satellites, airplanes, or unmanned aerial vehicles (UAV) equipped with optical sensors allows for high-throughput data collection on a large scale to measure the biophysical characteristics of crops [10]. UAV-based remote sensing has become increasingly popular in precision agriculture due to its flexibility, high efficiency, and desirable spatial resolution [11]. Optical sensors (cameras) mounted on the UAVs are used to capture spectral reflectance information of plant leaves and canopy [12]. Cameras commonly used on UAVs for agriculture include red-green-blue (RGB), multispectral (3 - 10 spectral bands), hyperspectral (hundreds of bands) and thermal

* Corresponding author.
E-mail address: zhoujianf@missouri.edu (J. Zhou).

https://doi.org/10.1016/j.atech.2024.100719
Received 13 October 2024; Received in revised form 11 December 2024; Accepted 14 December 2024
Available online 15 December 2024

cameras. RGB cameras are low-cost and high-resolution, often used to assess the morphological and visible color characteristics of crops, such as plant height [13,14], canopy size [15,10], stand count [16,17], and the level of greenness [18,19]. Multispectral and hyperspectral cameras can capture the reflectance spectrum of visible and near-infrared (VNIR) and shortwave infrared (SWIR), which enables them to detect more subtle changes in plant health or biochemical processes, such as water stress [7,10], nutrient content [20,21], herbicide damage [22,23], and seed protein [24,10].

UAV-based multispectral cameras have been used to estimate leaf chlorophyll content in various crops, including corn [25], sugarcane [26], and wheat [27]. Previous studies have explored different spectral vegetation indices, machine learning algorithms, and image pre-processing methods to estimate leaf chlorophyll content. For example, removing image backgrounds, including shadows and soil from the region of interest (ROI), can improve the accuracy of SPAD readings estimation [28]. Integrating environmental factors into imagery data could statistically improve the performance of regression models in estimating SPAD readings [29]. Moreover, Guo et al. [30] concluded that the support vector machine (SVM) outperformed the random forest regression (RF) under the circumstance of limited samples of leaf SPAD readings. Qiao et al. [31] indicated that vegetation indices (VIs) obtained from aerial imagery showed a linear correlation with leaf chlorophyll content at low and medium canopy coverage but a non-linear correlation at high canopy coverage.

Despite the feasibility of estimating corn leaf chlorophyll levels, many studies have focused on only a single time point, usually during the reproductive stages. (i.e., tasseling or silking). However, this limits the ability to detect early season stress. Tracking the spatial-temporal patterns of chlorophyll estimates allows farmers to identify emerging issues and take timely actions to protect against yield loss. Models that can estimate leaf chlorophyll content at any growth stage would be more practical for agricultural production. Additionally, researchers often rely on the mean value to represent the whole dataset, which can introduce bias and overlook important information, resulting in poor performance of machine learning models. Therefore, it is necessary to develop models to estimate leaf chlorophyll content at multiple growth stages using additional data descriptive methods (i.e., other than mean values).

This study aimed to investigate the feasibility of using an aerial multispectral imaging system to estimate the leaf chlorophyll content of corn across multiple vegetation stages. The two primary objectives were: (1) to identify suitable machine learning methods in estimating corn leaf chlorophyll content at different vegetative growth stages, and (2) to assess the performance of different statistical values on the model performance. These objectives aimed to provide a basis for monitoring the dynamic growth of corn using aerial multispectral imagery to support field management strategies.

2. Materials and methods

2.1. Field experiment

This study was conducted on a 0.2 hectare subfield of a 14 hectare corn field near Columbia, MO, USA (38°56′45.7″N, 92°07′59.4″W). The research field is in the humid continental region with four distinct seasons and at an elevation of approximately 232 m above mean sea level. The dimensions of the study area are 115 m (N - S direction) in length by 60 m in width. Corn was planted using a four-row planter equipped with John Deere MaxEmerge XP row units (Deere & Co., IL, USA) at a seeding

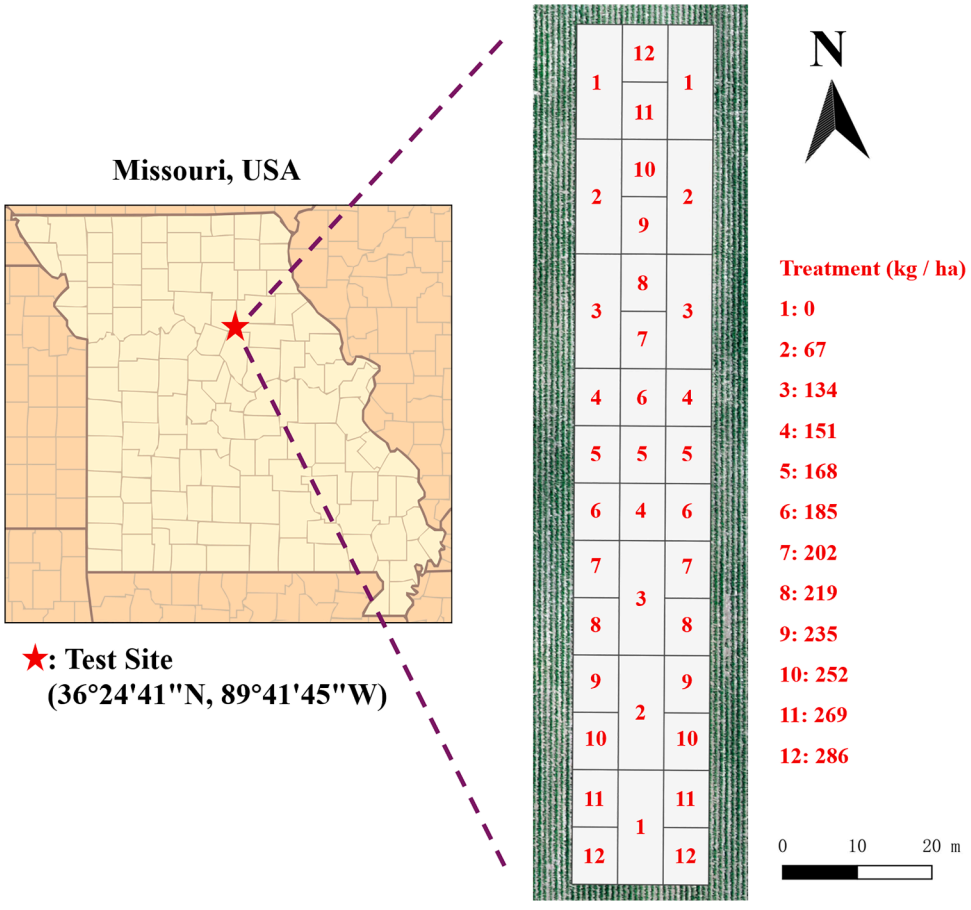

Fig. 1. Overview of the field experimental design using an N rate ramp approach. note: Numerical values (1 - 12) represent nitrogen application rates.

rate of 81,510 seeds ha^{-1} with a 0.76-meter row spacing. The experiment involved three replications of 12 incremental nitrogen rates ranging from 0 kg ha^{-1} to 285 kg ha^{-1} (i.e., 0, 67, 134, 151, 168, 185, 202, 219, 235, 252, 269, and 285 kg ha^{-1}). All plots were designed with a plot size of 7.6 m × 6.1 m, except for the plots with treatments of 0, 67, and 134 kg ha^{-1} (Fig. 1), which had a larger size of 15.2 m × 6.1 m due to the higher N rate increase between these plots. Nitrogen treatments were SuperU (46–0-0) that was broadcast applied at V4 as described in Abendroth et al. [32].

2.2. Data collection

The ground truth of leaf chlorophyll content was collected at four different vegetative growth stages at V8, V9, V11, and V12 (Vn indicates the growth stage when the nth leaf collar is present, Abendroth et al. [32]) from June to July 2021 (Table 1). Leaf nitrogen content was determined using a SPAD-502 chlorophyll meter. The average SPAD value for each plot was determined by taking five to six measurements from five randomly chosen plants in the middle two rows. Measurements were taken on the most recently collared leaf, between 1/3 and 2/3 of the length of the leaf from the central vein to the edge of the leaf. Data were collected from nitrogen treatments of 0, 67, 134, 168, 202, 235, and 286 kg ha^{-1} nitrogen rates in all three replications during all data collection events.

Soil apparent electrical conductivity (EC_a) was collected using a Veris 3100 instrument (Veris Technologies, Salina, KS, USA). Approximately 266 EC_a measurements, including shallow EC_a (0 - 30 cm) and deep EC_a (0 - 100 cm), were collected in the study area, which were used as indirect indicators of soil texture and quality [33]. To improve spatial accuracy in imaging processing, a real-time kinematic (RTK) GNSS system Reach RS+ (EMLID Ltd., Budapest, Hungary) was used to obtain the coordinates of the edge of each plot (four corners) and the ground control points at the edge of the field.

Imagery data were acquired using a UAV platform (Matrice 600 Pro, DJI Technology, Guangdong, China) equipped with a multispectral camera (RedEdge-M, AgEagle Aerial Systems, Wichita, KS, USA). The camera consists of five spectral narrow bands, i.e., Red (R) at 668 nm ± 10 nm, Green (G) at 560 nm ± 20 nm, Blue (B) at 475 nm ± 20 nm, Red Edge (RE) at 717 nm ± 10 nm, and Near Infrared (NIR) at 840 nm ± 40 nm. The camera captures images at a resolution of 1260 × 960 pixels. A flight planning app Autopilot (Hangar Technology, TX, USA) on an iPad Mini 4 (Apple, CA, USA) was used to set the flight path and configure flight parameters (i.e., frame rate, speed, altitude). The UAV flight paths ensured a sufficient image overlap (≥ 70%) for orthomosaic images. The ground sampling distance (GSD) was set to 1.7 cm per pixel for all the flights. Prior to each flight, an image of a manufacturer-provided calibration reflectance panel was captured by holding the camera at approximately 1.0 m above the panel to avoid any shadow. All data collection events were conducted between 10 am to 2 pm, coinciding with minimum variations in the solar zenith angle.

2.3. Image processing and feature extraction

All UAV images were processed using Pix4D (Pix4D, Lausanne, Switzerland) to stitch UAV collected images and build orthomosaic images of the study area. All orthomosaic images were exported in "tiff" (Tag Image File Format) format and segmented using the geographic information system software QGIS (version 3.28.2). The coordinate reference system was set as WGS 84 / UTM zone 15 N (EPSG: 32,615) to avoid the imaging raster distortion. Using the coordinates of the plot corners, a polygon file was generated and then rescaled to focus on the central area of each plot, avoiding boundaries. The rescaled polygons were used to isolate the corresponding plot areas from each orthomosaic image. The resulting images of individual plots were saved with file names that included the plot ID and growth stage.

The background information of segmented images (e.g., soil and shadows) was removed using the Otsu method [34] based on the excess greenness (ExG) index, which is calculated using Eq. 1.

$$ExG = 2 \times Green - Red - Blue \quad (1)$$

where, *Green, Red* and *Blue* are the pixel values of the three channels of an RGB image. The Otsu + ExG method has been applied in other studies to eliminate background information surrounding the crop in different agricultural scenarios [35,36].

Vegetative indices and other image features calculated from the processed multispectral images are listed in Table 2. They have been widely used in other research to quantify the physiological and chemical characteristics of crops [26,31,37].

When reducing the dimensionality of vegetation indices data from multispectral imagery (i.e., from 2D to 1D), previous studies commonly use the mean value for each ROI as an independent variable. However, solely depending on mean values can cause a significant loss of information. In this study, multiple summary statistics were selected, including 25, 50, and 75 percentiles (Q1, Q2, and Q3), Mean, Standard Deviation, Total Absolute Difference, Interquartile Range, Upper Inner Fence, and Lower Inner Fence, which are described in Table 3. They

Table 1
Field collection event dates and corresponding vegetative growth stage.

Growth Stage (Date)	Aerial Data Collection	Ground Data Collection
V8 (06/18)	✓	✓
V9 (06/25)	✓	✓
V11 (07/02)	✓	✓
V12 (07/08)	✓	✓

Table 2
Equations of the extracted image features from multispectral imagery.

Extracted image feature	Equation*	Reference
NDVI	(NIR − R) / (NIR + R)	Carlson and Ripley [38]
GNDVI	(NIR − G) / (NIR + G)	Shanahan et al. [39]
NDRE	(NIR − RE) / (NIR + RE)	Li et al. [40]
CCCI	NDRE / NDVI	El-Shikha et al. [41]
CI_Green	(NIR / G) − 1	Viña et al. [42]
CI_Red_Edge	(NIR / RE) − 1	Viña et al. [42]
MSR	$[(NIR / R) - 1]/\sqrt{(NIR / R) + 1}$	Salas and Henebry [43]
RDVI	$(NIR - R)/\sqrt{NIR + R}$	Chen [44]
RVI	NIR / R	Gupta et al. [45]
EVI	2.5 × ((NIR − R) / (NIR + 6 × R − 7.5 × B + 1))	Matsushita et al. [46]
TVI	$\sqrt{(NIR - R) / (NIR + R) + 0.5}$	Bannari et al. [47]
IPVI	NIR / (NIR + R)	Payero et al. [48]
MTCI	(NIR - RE) / (RE - R)	Dash and Curran [49]
BNDVI	(NIR − B) / (NIR + B)	Wang et al. [50]
WDRVI	(α × NIR - R) / (α ×NIR + R); α = 0.2	Gitelson [51]
TCARI	3 × [(RE - R) - 0.2 × (RE - G) × (RE / R)]	L. Zhang et al. [7]
ExG	2 × G - R - B	Meyer and Neto [52]
GLI	[(G - R) + (G - B)] / [(2 × G) + R + B]	Barbosa et al. [53]
RMS Contrast	Population standard deviation of the grayed plot image	Kukkonen et al. [54]
Canopy Coverage (CC)	Pixel number of the canopy in the ROI / Total pixel number in the ROI	Walton et al. [55]

* R = red; G = green; B = blue; RE = red edge; NIR = near infrared.

Table 3
Methods for extracting statistically descriptive features from the raster data.

Summary Statistics of Each Feature	Definition
Q1	The first quartile (25th percentile) of the data
Q2	The median (50th percentile) of the data
Q3	The third quartile (75th percentile) of the data
Mean	The arithmetic average of the data
Population Standard Deviation	A measure of the spread of the data
Total Absolute Difference	The sum of the absolute differences between each data point
Interquartile Range	The difference between Q3 and Q1
Upper Inner Fence	The upper boundary for identifying potential outliers
Lower Inner Fence	The lower boundary for identifying potential outliers

were all derived from the ROIs in each raster dataset to capture as much information as possible. All image data analyses were performed in RStudio (Ver. Posit, MA, USA).

2.4. Dimensionality reduction algorithms

The dataset from Section 2.3 includes 182 independent variables and one dependent variable, which may lead to multicollinearity, noise, and redundancy, and negatively impact the accuracy of the predicted results [56]. The dimensionality of the dataset in this study was reduced by the Sequential Forward Selection method, which decreases the dimensionality of the feature space by selecting the most relevant features [57]. However, traditional Sequential Forward Selection methods (e.g., Forward / Backward Stepwise Regression) have an overfitting issue, especially for a small dataset [58]. Therefore, the Sequential Cross-Validated Forward Selection (SCVFS) method, based on the theory of Sequential Forward Selection, was used to reduce the dimensionality of the dataset [59]. The steps of SCVFS are illustrated in Fig. 2, including:

(1) Initialize the machine learning model with pre-defined hyper-parameters and create empty lists (X_{opt} & *Error_Record*) to store selected variables and modeling performance on testing dataset.
(2) For each variable X_i that does not exist in X_{opt}, fit the model on the training dataset using X_i, recursively, combined with previously selected variables in X_{opt}. Use Monte Carlo Cross-Validation (MCCV), resampling the training and testing data 100 times, to evaluate the performance of machine learning model on the testing dataset. Record the average Root Mean Squared Error (RMSE) for each X_i.
(3) After evaluating all variables, select the X_i with the lowest average RMSE and add it to X_{opt}. Repeat this process again, adding one X_i per round until $n-1$ rounds are completed (where n is the total number of independent variables).

2.5. Chlorophyll content prediction models

This study evaluated four regression models used extensively in various remote sensing applications to predict leaf chlorophyll content in corn [60,61]. The selected models included Epsilon–Support Vector Regression (ε-SVR), Elastic–Net Regularization Regression (Elastic-Net), Partial Least Squares Regression (PLSR), and Random Forest Regression (RF).

The ε-SVR model uses a kernel function to map input data into a high-dimensional feature space. It finds a hyperplane that minimizes the deviation between the actual and prediction values, subject to a tolerance defined by ε. ε defines a margin of tolerance where no penalty is given to training errors. This approach helps the model generalize well, even when working with small datasets, without being overly sensitive to minor variations [62]. The SCVFS-ε-SVR model uses the SCVFS method described in Section 2.4 to select variables that are then entered into the ε-SVR for fitting. The ε-SVR was built using the function "*ksvm ()*" in the R package "*kernlab*". The hyper-parameters selected as listed as *kernel* = "*vanilladot*" (Linear Kernel), $C = 0.1$ (Lagrange Regularization), *epsilon* = *0.2* (Insensitive Loss), and all other settings set to default.

Elastic-Net is a statistical model that combines the L1 (Lasso) and L2 (Ridge) regularization methods. It can avoid multicollinearity by setting

Input: Dataset D containing independent variables $\{X_i | i \in \{1, \ldots, n\}\}$ and dependent variable Y
Output: Ranked variables list X_{opt} and their associated minimum errors
Initialize model with hyper-parameters
$X_{opt} \leftarrow$ empty list ; $Error\_Record \leftarrow$ empty list
for $k = 1$ to $n - 1$ **do**
    $Error\_List \leftarrow$ empty list
    for each X_i not in X_{opt} **do**
        $total\_error \leftarrow 0$
        for $j = 1$ to 100 **do** ▷ Monte Carlo CV
            Resample D into D_{train}, D_{test}
            Fit model using $\{X_{opt}, X_i\}$ on D_{train}
            $total\_error \leftarrow total\_error +$ Evaluate(D_{test}, RMSE)
        end for
        Append $(X_i, total\_error/100)$ to $Error\_List$
    end for
    $(best\_variable, min\_error) \leftarrow$ entry with min error in $Error\_List$
    Add $best\_variable$ to X_{opt}
    Append $(best\_variable, min\_error)$ to $Error\_Record$
end for
return $X_{opt}, Error\_Record$

Fig. 2. Example code for SCVFS.

the coefficient of each independent variable extremely low or to zero, thereby minimizing the effect of the variables in the model. This makes the modeling process more efficient as it does not require additional dimensionality reduction methods. In this study, the model was built using the function "*cv.glmnet()*" from the R package "*glmnet*" by specifying the hyper-parameter *alpha = 0.6* and parameter *family = "Gaussian"* with all other settings kept as default. The best shrinkage parameter (λ) was selected through a 5-fold cross-validation on the training dataset, choosing the λ value based on the lowest average training error. Partial Least Squares Regression is another commonly used multivariate statistical method for predictive modeling. It can also handle multicollinearity and is suitable for small datasets [63]. Unlike the PCA, the PLSR algorithm considers independent and dependent variables to derive latent variables that are most relevant to the dependent variable. The '*plsr()*' function from the '*pls*' R package was used to apply the model, with seven latent variables selected based on Leave-One-Out Cross-Validation (LOOCV) during the training process.

Random Forest is a classic ensemble learning algorithm that captures non-linear relationships between independent and dependent variables. It constructs multiple (hundreds or even thousands) decision trees on bootstrapped samples of the training data, randomly selecting subsets of independent variables during node splitting. This improves the model's robustness and helps avoid overfitting. The "*randomForest()*" function from the "*randomForest*" R package was used to build the RF model, with *ntree = 500, mtry = 12,* and *nodesize = 2* and all other settings were kept as default.

All training data was standardized following the Gaussian distribution ($\mu = 0$ & $\sigma = 1$), and the testing data were adjusted using the same scaling parameters from the training dataset. The optimal hyper-parameters for each model were determined by selecting the lowest average RMSE on the testing dataset through a grid search with 100 MCCV. Three evaluation metrics were used to assess the model performance on various datasets, including coefficient of determination (R^2), Mean Absolute Error (MAE), and RMSE, as calculated in Eqs. 2-4.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{2}$$

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \tag{3}$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \tag{4}$$

where, n denotes sample size; y_i indicates the true value; $\hat{y}_i$ means the predicted value; $\bar{y}$ is the mean of the true values.

2.6. Kriging interpolation

This study used Ordinary Kriging with the "smartmap" plugin in QGIS to estimate the shallow EC_a and deep EC_a values. A linear semi-variogram model was fitted to the experimental semi-variogram values obtained from the original dataset, resulting in an R^2 of 0.976, indicating a good explanation of the spatial data variability. The parameters "*Neighbors = 10*" and "*Radius = 39.048*" were chosen based on a preliminary trail to create a smoother interpolation map. The same polygon shapefile used to segment the plots from the orthomosaic image was also used to segment the interpolated EC_a data. The data was extracted using the same summary statistics (Table 3) to convert it from 2-dimensional to 1-dimensional data.

2.7. Statistical analysis

The study used the Pearson correlation coefficient (r) to assess the similarity between two polynomial regression models. These models were developed based on SPAD readings for each growth stage. The first model used values only from plots with ground truth data, while the second model included data from all plots, including both ground truth and predicted values from the selected machine learning model. To simulate nitrogen application rates, 286 structured sequences ranging from 0 to 285 kg N ha^{-1} in 1 kg N ha^{-1} increments. These sequences were then fitted to both models to generate predicted leaf chlorophyll content. The r was generated between the predicted values from the first and the second polynomial regression models.

3. Results and discussion

3.1. Descriptive statistics of ground truth data

The descriptive statistics of measured leaf chlorophyll content from the SPAD meter are shown in Fig. 3a, indicating that the SPAD readings were between 32.9 and 59.2 with the median value of 51.3 and a standard deviation of 6.0 SPAD units. The histogram shows a bimodal distribution peaking at around 48 and 56. The point plot in Fig. 3b shows the distribution of SPAD readings for the four growth stages. The trend of each growth stage in different nitrogen rates was shown using the polynomial regression (*degree = 2*). Growth stage V12 indicates that the SPAD readings are lower at higher nitrogen rates, however, this phenomenon is not obvious in earlier vegetation stages. In addition, V12 has relatively low SPAD readings compared to other growth stages across all treatments. This phenomenon has also been observed in other studies where corn leaf chlorophyll has lower SPAD readings in the late vegetative stages [64].

3.2. Model performance on leaf chlorophyll content estimation

The performance of the selected regression models in estimating corn leaf chlorophyll content from SPAD readings are shown in Table 4. There are a total of 182 individual variables (nine statistical measures for 20 VIs plus Contrast) as defined in Table 2. The highest accuracy for the leaf chlorophyll content was achieved by the SCVFS-ε-SVR model with the R^2 of 0.87 and MAE of 2.26 SPAD units based on the average value with a 100 MCCV. On the other hand, RF performs worst with the R^2 of 0.79 and MAE of 3.19 SPAD units. RF is the only non-linear regression model in our experiment, which is relatively more complex than linear models and may require more data to fit. [65] also concluded that RF could perform worse especially on a small dataset in remote sensing tasks.

3.3. Feature selection based on SCVFS

The RMSE was calculated using 182 independent variables based on the model SCVFS-ε-SVR by running the ε-SVR with 100 MCCV under different combinations of independent variables. The RMSE response to the number of independent variables for training and testing datasets are shown in Fig. 4.

In this study, predictive modeling was initiated with a single independent variable and observed that the RMSE exceeded 6.0 SPAD units for both training and testing datasets. As additional variables were gradually introduced, the performance of the model improved significantly. When the model included more than 20 independent variables, its performance on the testing dataset plateaued (changes of RMSE $<$ 0.01 SPAD units), indicating a state of diminishing returns, while training accuracy continued to increase. The results suggest that the model was overfitting by using more variables, which increased training time and compromised the feature space [37]. To reduce computing time and avoid overfitting, an early-stop mechanism was implanted in the forward selection process. This mechanism stops the addition of new variables when the model's performance metrics (RMSE or R^2) no longer show significant improvement.

The results visualized in Fig. 4 indicated that certain variables had a significant improvement in the performance of the ε-SVR. Specifically,

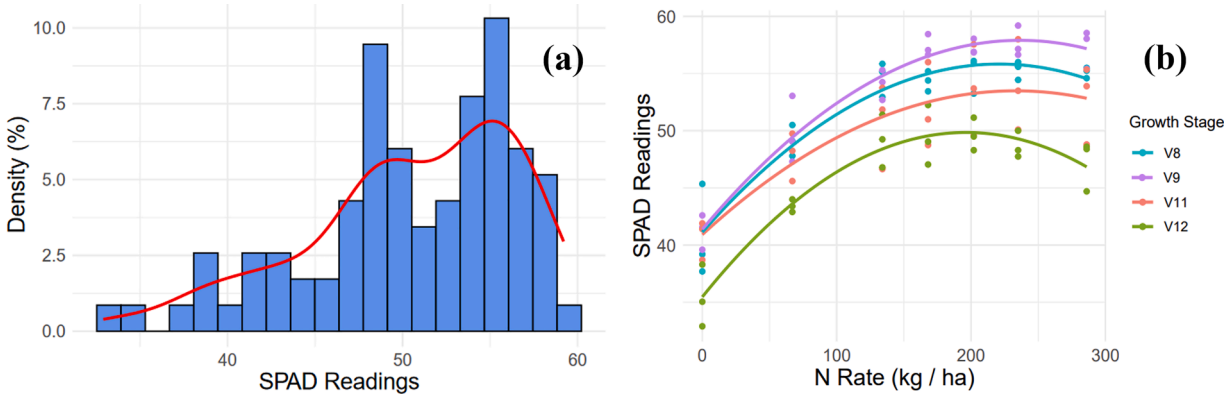

Fig. 3. Description of the ground truth from growth stages v8, v9, v11, and v12 based on the treatment 1, 2, 3, 5, 7, 9, 12 for all 3 replications. (a). Distribution of all leaf SPAD readings. (b). Trend of SPAD readings at four growth stages fitted by Polynomial regression (*degree* = *2*).

Table 4
Performance of the Elastic-Net, **RF**, ε - SVR, SCVFS - ε - SVR, and PLSR models in leaf chlorophyll content estimation using nine extracted statistic measurements from each vegetation index. Evaluation metrics R^2, MAE (SPAD value), and RMSE (SPAD value) were calculated based on the 100 MCCV testing dataset.

Model	R^2	MAE	RMSE
Elastic - Net	0.82	2.06	2.63
RF	0.79	2.51	3.19
PLSR	0.81	2.14	2.76
ε - SVR	0.83	2.02	2.59
SCVFS - ε - SVR	0.87	1.80	2.26

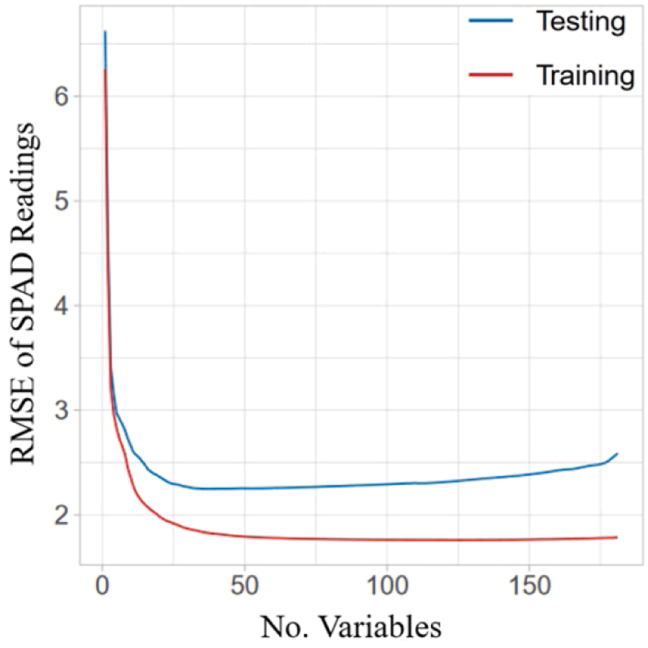

Fig. 4. Root Mean Square Error (RMSE, SPAD value) of SPAD values (indicating the leaf chlorophyll content) predicted by the SCVFS-ε-SVR algorithm. The horizontal axis shows the number of variables used for the training (red) and testing (blue) phases and the vertical axis shows the RMSE values (SPAD values).

four independent variables (i.e., Q1 of the GLI, the total absolute difference in shallow soil EC_a, the upper inner fence of both the GNDVI and RDVI) were identified as the most influential with at least a 0.2 decrease in the RMSE on the test dataset. The GLI mainly represents vegetation greenness, which corresponds to chlorophyll concentration. The total absolute difference of shallow soil EC_a also contributes significantly to the model's performance as an environmental factor with a moderate negative linear relationship with the dependent variable ($r \approx$ -0.55). Additionally, GNDVI and RDVI, which are related to crop vigor, had an impact on the model's performance. While other variables had minor improvements in explaining the variance observed in the SPAD readings on the testing dataset, their combined contribution was still significant. Interestingly, the SCVFS method rarely included extracted image features that were calculated as an average. Such an observation makes it worth comparing our method with the conventional approach of extracting and using only the mean value of the image feature.

3.4. Compare different feature extraction methods

The mean value of the collected imagery data has been extensively used to represent image or environmental features in a specified ROI for machine learning models. Using fewer variables during training requires fewer computational resources, but relying solely on the mean value may not capture all the trends or the dispersion in the original dataset. This study extracts the mean value of each VI and soil EC_a with image contrast to fit into the model ε - SVR with the SCVFS feature selection method, in order to compare the results of Section 3.3. As more variables are added, the model's performance steadily increases, reaching an RMSE of 2.81 at 11 variables on the testing dataset. However, adding too many variables leads to overfitting, where the model performs well on the training data but fails to generalize the testing data. Fig. 5 shows this pattern, which is similar to Fig. 4. It demonstrates that adding too many variables can negatively impact the model's performance.

The mean values of GLI, RDVI, contrast, ExG, RVI, CCCI, and GNDVI are the top 7 variables that contribute to at least a 0.1 decrease in RMSE of the testing dataset. Similar to Section 3.3, GLI was selected as the most contributing factor in this study. However, it is important to note that the GLI was calculated based on the mean value in this section whereas the first selected variable is Q1 of GLI in Section 3.3. Besides, these selected VIs are mainly related to the greenness of plants (e.g., GNDVI, ExG, GLI) or related to NIR reflectance (RVI, RDVI, and CCCI). While the independent variable "contrast" also plays a significant role, potentially due to the variance in canopy color, which could be caused by environmental factors. Overall, this part of the study suggests that adopting more summary statistics to describe the matrix format image and environmental features could provide an additional benefit to the performance of machine learning models.

3.5. Goodness of fit in different vegetation stages

The performance of the model on the training dataset can help us understand how well it fits the collected data and how it changes across various corn vegetation stages. This section used the same hyper-

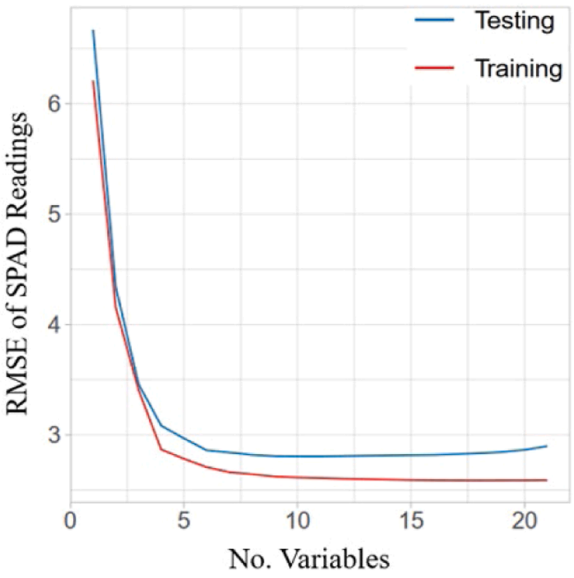

Fig. 5. R^2 of SCVFS-ε-SVR performance on training and testing dataset by considering the mean value of each vegetation index only with canopy coverage and image contrast.

parameters of the ε-SVR model and independent variables identified in Section 3.3. Data collected from different vegetative stages are individually fitted into the linear model.

The performance of each model is consistent across the four sub-sets of data collected from different dates (Table 5). The selected independent variables were able to explain 92% to 95% of the changes or fluctuations in the dependent variable for the model. This shows that the relationship between independent and dependent variables remains consistent across different corn growth stages. Therefore, by combining datasets from various periods, it might be possible to adapt the model for similar scenarios, assuming the external environmental conditions remain unchanged with those of the periods covered by the sub-datasets.

It is also worthwhile investigating whether the model can estimate the other plots (*151, 185, 219, 252, and 269 N ha*$^{-1}$) where ground truth data were not collected (Fig. 6). The model was trained based on the independent variables concluded from Section 3.3 by combining all available data with ground truth to estimate the SPAD readings for areas without the data. Fig. 6 illustrates four polynomial regression lines with a degree of 2 based on combining the collected and predicted leaf chlorophyll content from the ε-SVR. Through visual observation, the predicted values follow the pattern of the ground truth, with low SPAD readings at growth stage V12 and higher readings than others at V9. By following the procedure outlined in *Section 2.6*, the inclusion of predicted values did not alter the polynomial regression model. Consequently, the $r \geq 0.9999$ between the predicted value from models across all growth stages demonstrates the stability and accuracy of the model in predicting performance on unseen data.

Table 5
Result of how well the observed data corresponds to the ε-SVR.

Vegetation stages	R^2	DAN (days)[2]
V8	0.94	20
V9	0.95	27
V11	0.92	34
V12	0.94	40
All[1]	0.90	N/A

[1] V8 + V9 + V11+ V12. [2]: Days After Nitrogen Application*:

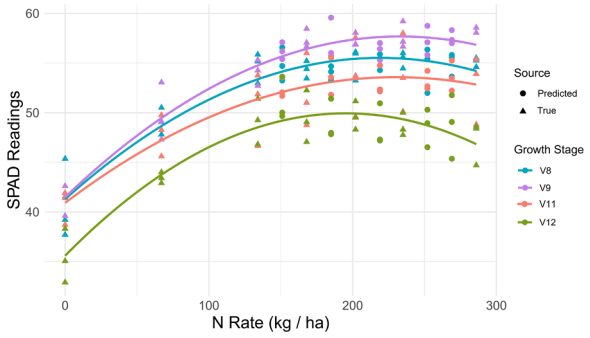

Fig. 6. The polynomial regression (*degree = 2*) fitted to the combination of predicted SPAD-readings from ε-SVR and ground truth at four growth stages. The dots show the predicted values and triangled shapes show the true values.

3.6. Future study

Leaf chlorophyll content is an important indicator of crop health [1], which has been assessed using UAV-based remote sensing and proximal sensing technologies. However, few studies have focused on the multiple vegetative stages [26,31]. Monitoring leaf chlorophyll content during the vegetative stages can help to make decisions on water and fertilizer management prior to reproductive stages, ultimately influencing yield and grain quality. In this study, a more comprehensive feature extraction and selection strategy were used to estimate leaf chlorophyll content. Instead of relying on the averages of image-derived features, a number of statistics were used to capture the distributions and variability of each image feature. A more refined and automated dimensionality reduction and improved model accuracy were achieved using the SCVFS method. In addition, this study also indicates that integrating environmental parameters (i.e., soil ECa) with UAV multispectral imagery data can improve the estimation accuracy of corn leaf chlorophyll content.

Machine learning models with limited sample sizes are at a higher risk of experiencing statistical artifacts (e.g., sampling bias, spurious correlation), leading to suspicious results [66]. To mitigate the risk of sampling bias due to limited size of data, the study employed the 100 times MCCV method to obtain more accurate results. However, this approach requires extensively higher computing time and resources. Future research should compare the result obtained from different feature selection methods such as, “recursive feature elimination” (REF), “SHapley Additive exPlanations” (SHAP), and “permutation feature importance”, to obtain more accurate understanding of the feature importance.

This study had a minimal number of ground samples. Moreover, the machine learning model was only based on one hybrid and one site year of data. This means the model needs to be improved for its universality and robustness. Therefore, future studies should include more widely cultivated hybrids and collect data from different fields with more environmental variations.

This study focused on aerial multispectral imaging, which may not provide enough information due to low spectral resolution and weak band continuity. In contrast to other sensors, hyperspectral sensors capture critical spectral information essential for assessing the physiological parameters of crops. This capability allows for a more detailed analysis of specific spectral signatures generated by the absorption and reflection of solar radiation by leaf chlorophyll. However, there is a lack of research on the benefits of including SWIR (900 nm to 2500 nm) data in real production scenarios, which could potentially enhance the model’s performance by providing more information. Future studies should explore the role of SWIR in improving the accuracy of leaf chlorophyll estimation to gain valuable insights into crop physiology and the practical application of N management.

4. Conclusion

The study compared different machine learning methods in predicting corn leaf chlorophyll content using aerial imagery and soil EC_a data. Feature extraction techniques that consider a wider range of descriptive statistics were found to improve prediction accuracy. The ε-SVR model with SCVFS feature selection performed the best, outperforming RF, PLSR, and Elastic-Net with an R^2 of 0.87, MAE of 1.80, and RMSE of 2.26. The model's performance remained stable across different vegetation growth stages. However, the study was limited by a small dataset, so further investigation with larger datasets is needed. Combining UAV imagery with environmental data can help accurately and continuously monitor the chlorophyll content of leaves during crop growth stages, helping to identify crop health issues early, and allowing for timely intervention and preventing loss of profitability.

CRediT authorship contribution statement

Fengkai Tian: Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Jianfeng Zhou:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization. **Curtis J. Ransom:** Writing – review & editing, Validation, Data curation, Conceptualization. **Noel Aloysius:** Writing – review & editing, Methodology, Investigation. **Kenneth A. Sudduth:** Writing – review & editing, Supervision, Funding acquisition, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We would like to thank Chin Nee Vong and Jing Zhou for their assistance in data collection. This research was partially funded by the USDA Agricultural Research Service and the Hatch Project of the University of Missouri. Mentioning company or trade names is for description only and does not imply endorsement by the USDA or the University of Missouri. The USDA and the University of Missouri are equal opportunity providers and employers.

Ethics statement

Not applicable: This manuscript does not include human or animal research.

Data availability

Data will be made available on request.

Reference

[1] D. Pavlović, B. Nikolić, S. Đurović, H. Waisi, A. Anđelković, D. Marisavljević, Chlorophyll as a measure of plant health: Agroecological aspects, Pesticidi i fitomedicina 29 (1) (2014) 21–34.
[2] B.M. Atta, M. Saleem, H. Ali, H.M.I. Arshad, M. Ahmed, Chlorophyll as a biomarker for early disease diagnosis, Laser. Phys. 28 (6) (2018) 065607.
[3] R. Sanchez, A. Hall, N. Trapani, R.C. De Hunau, Effects of water stress on the chlorophyll content, nitrogen level and photosynthesis of leaves of two maize genotypes, Photosynth. Res. 4 (1983) 35–47.
[4] H.M. Kalaji, P. Dąbrowski, M.D. Cetner, I.A. Samborska, I. Łukasik, M. Brestic, M. Zivcak, H. Tomasz, J. Mojski, H. Kociel, A comparison between different chlorophyll content meters under nutrient deficiency conditions, J. Plant Nutr. 40 (7) (2017) 1024–1034.
[5] Y. Zhao, Q. Han, C. Ding, Y. Huang, J. Liao, T. Chen, S. Feng, L. Zhou, Z. Zhang, Y. Chen, S. Yuan, M. Yuan, Effect of low temperature on chlorophyll *b*iosynthesis and chloroplast biogenesis of rice seedlings during greening, Int. J. Mol. Sci. 21 (4) (2020) 1390. https://www.mdpi.com/1422-0067/21/4/1390.
[6] C. Parry, J.M. Blonquist Jr, B Bugbee, In situ measurement of leaf chlorophyll concentration: analysis of the optical/absolute relationship, Plant Cell Environ. 37 (11) (2014) 2508–2520.
[7] L. Zhang, H. Zhang, Y. Niu, W. Han, Mapping maize water stress based on UAV multispectral remote sensing, Remote Sens. 11 (6) (2019) 605.
[8] N. Ahmad, M.F.A. Muttalib, M. Uda, Z. Arsat, F. Abdullah, M. Hashim, F. Azizan, M. Jusoh, S. Kamaruzaman, A. Nordin, Measurement of leaf chlorophyll content in Harumanis mango cultivated in a greenhouse using SPAD meter, in: Materials Today: Proceedings, 2023.
[9] A. Rhezali, A.E. Aissaoui, Feasibility Study of Using Absolute SPAD values for standardized evaluation of corn nitrogen status, Nitrogen 2 (3) (2021) 298–307. https://www.mdpi.com/2504-3129/2/3/20.
[10] J. Zhou, J. Zhou, H. Ye, M.L. Ali, P. Chen, H.T. Nguyen, Yield estimation of soybean breeding lines under drought stress using unmanned aerial vehicle-based imagery and convolutional neural network, Biosyst. Eng. 204 (2021) 90–103.
[11] W.H. Maes, K. Steppe, Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture, Trends. Plant Sci. 24 (2) (2019) 152–164.
[12] D.A. Sims, J.A. Gamon, Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages, Remote Sens. Environ. 81 (2-3) (2002) 337–354.
[13] A. Feng, J. Zhou, E.D. Vories, K.A. Sudduth, M. Zhang, Yield estimation in cotton using UAV-based multi-sensor imagery, Biosyst. Eng. 193 (2020) 101–114.
[14] L. Volpato, F. Pinto, L. González-Pérez, I.G. Thompson, A. Borém, M. Reynolds, B. Gérard, G. Molero, F.A. Rodrigues Jr, High throughput field phenotyping for plant height using UAV-based RGB imagery in wheat breeding lines: Feasibility and validation, Front. Plant Sci. 12 (2021) 591587.
[15] A. Pagliai, M. Ammoniaci, D. Sarri, R. Lisci, R. Perria, M. Vieri, M.E. M. D'Arcangelo, P. Storchi, S.-P. Kartsiotis, Comparison of aerial and ground 3D point clouds for canopy size assessment in precision viticulture, Remote Sens. 14 (5) (2022) 1145.
[16] F. Tian, C.J. Ransom, J. Zhou, B. Wilson, K.A. Sudduth, Assessing the impact of soil and field conditions on cotton crop emergence using UAV-based imagery, Comput. Electron. Agric. 218 (2024) 108738.
[17] C.N. Vong, L.S. Conway, A. Feng, J. Zhou, N.R. Kitchen, K.A. Sudduth, Corn emergence uniformity estimation and mapping using UAV imagery and deep learning, Comput. Electron. Agric. 198 (2022) 107008.
[18] M.L. Buchaillot, A. Gracia-Romero, O. Vergara-Diaz, M.A. Zaman-Allah, A. Tarekegne, J.E. Cairns, B.M. Prasanna, J.L. Araus, S.C. Kefauver, Evaluating maize genotype performance under low nitrogen conditions using RGB UAV phenotyping techniques, Sensors 19 (8) (2019) 1815.
[19] F. Tian, C.C. Vieira, J. Zhou, J. Zhou, P. Chen, Estimation of off-target dicamba damage on soybean using uav imagery and deep learning, Sensors 23 (6) (2023) 3241.
[20] O.S. Walsh, S. Shafian, J.M. Marshall, C. Jackson, J.R. McClintick-Chess, S. M. Blanscet, K. Swoboda, C. Thompson, K.M. Belmont, W.L. Walsh, Assessment of UAV based vegetation indices for nitrogen concentration estimation in spring wheat, Adv. Remote Sens. 7 (2) (2018) 71–90.
[21] J. Zhou, B. Wang, J. Fan, Y. Ma, Y. Wang, Z. Zhang, A systematic study of estimating potato N concentrations using UAV-based hyper-and multi-spectral imagery, Agronomy 12 (10) (2022) 2533.
[22] Betitame, K., Koparan, C., Zhang, Y., Howatt, K., Ostlie, M., Bajwa, S.G., & Sun, X. (2024). Evaluation of dicamba drift injury and yield loss on soybean using small unmanned aircraft systems (sUAS) and multispectral imaging technologies.
[23] Y. Huang, K.N. Reddy, S.J. Thomson, H. Yao, Assessment of soybean injury from glyphosate using airborne multispectral remote sensing, Pest. Manage Sci. 71 (4) (2015) 545–552.
[24] K. Dilmurat, V. Sagan, M. Maimaitijiang, S. Moose, F.B. Fritschi, Estimating crop seed composition using machine learning from Multisensory UAV Data, Remote Sens. 14 (19) (2022) 4786.
[25] B. Chen, G. Huang, X. Lu, S. Gu, W. Wen, G. Wang, W. Chang, X. Guo, C. Zhao, Prediction of vertical distribution of SPAD values within maize canopy based on unmanned aerial vehicles multispectral imagery, Front. Plant Sci. 14 (2023) 1253536.
[26] A. Narmilan, F. Gonzalez, A.S.A. Salgadoe, U.W.L.M. Kumarasiri, H.A. S. Weerasinghe, B.R. Kulasekara, Predicting canopy chlorophyll content in sugarcane crops using machine learning algorithms and spectral vegetation indices derived from uav multispectral imagery, Remote Sens. 14 (5) (2022) 1140. https://www.mdpi.com/2072-4292/14/5/1140.
[27] Q. Wu, Y. Zhang, Z. Zhao, M. Xie, D. Hou, Estimation of relative chlorophyll content in spring wheat based on multi-temporal UAV remote sensing, Agronomy 13 (1) (2023) 211.
[28] J. Ji, N. Li, H. Cui, Y. Li, X. Zhao, H. Zhang, H. Ma, Study on Monitoring SPAD values for multispatial spatial vertical scales of summer maize based on UAV multispectral remote sensing, Agriculture 13 (5) (2023) 1004. https://www.mdpi.com/2077-0472/13/5/1004.
[29] X. Yang, R. Yang, Y. Ye, Z. Yuan, D. Wang, K. Hua, Winter wheat SPAD estimation from UAV hyperspectral data using cluster-regression methods, Int J Appl Earth Observat Geoinf 105 (2021) 102618.
[30] Y. Guo, S. Chen, X. Li, M. Cunha, S. Jayavelu, D. Cammarano, Y. Fu, Machine learning-based approaches for predicting SPAD values of maize using multi-spectral images, Remote Sens. 14 (6) (2022) 1337. https://www.mdpi.com/2072-4292/14/6/1337.

[31] L. Qiao, W. Tang, D. Gao, R. Zhao, L. An, M. Li, H. Sun, D. Song, UAV-based chlorophyll content estimation by evaluating vegetation index responses under different crop coverages, Comput. Electron. Agric. 196 (2022) 106775.
[32] Abendroth, L., Elmore, R., Boyer, M., & Marlay, S. (2011). Corn growth and development. PMR 1009. Iowa State Univ. *Ext.: Ames, IA, USA*.
[33] K.A. Sudduth, N. Kitchen, G. Bollero, D. Bullock, W. Wiebold, Comparison of electromagnetic induction and direct sensing of soil electrical conductivity, Agron. J. 95 (3) (2003) 472–482.
[34] N. Otsu, A threshold selection method from gray-level histograms, Automatica 11 (285-296) (1975) 23–27.
[35] Z. Cheng, L. Qi, Y. Cheng, Y. Wu, H. Zhang, Interlacing Orchard Canopy Separation and Assessment using UAV Images, Remote Sens. 12 (5) (2020) 767. https://www.mdpi.com/2072-4292/12/5/767.
[36] B. Li, X. Xu, J. Han, L. Zhang, C. Bian, L. Jin, J. Liu, The estimation of crop emergence in potatoes by UAV RGB imagery, Plant Methods 15 (1) (2019) 1–13.
[37] S. Sarkar, V. Sagan, S. Bhadra, K. Rhodes, M. Pokharel, F.B. Fritschi, Soybean seed composition prediction from standing crops using PlanetScope satellite imagery and machine learning, ISPRS J. Photogr. Remote Sens. 204 (2023) 257–274.
[38] T.N. Carlson, D.A. Ripley, On the relation between NDVI, fractional vegetation cover, and leaf area index, Remote Sens. Environ. 62 (3) (1997) 241–252.
[39] J.F. Shanahan, J.S. Schepers, D.D. Francis, G.E. Varvel, W.W. Wilhelm, J.M. Tringe, M.R. Schlemmer, D.J. Major, Use of remote-sensing imagery to estimate corn grain yield, Agron. J. 93 (3) (2001) 583–589.
[40] F. Li, Y. Miao, G. Feng, F. Yuan, S. Yue, X. Gao, Y. Liu, B. Liu, S.L. Ustin, X. Chen, Improving estimation of summer maize nitrogen status with red edge-based spectral vegetation indices, Field. Crops. Res. 157 (2014) 111–123.
[41] D.M. El-Shikha, E.M. Barnes, T.R. Clarke, D.J. Hunsaker, J.A. Haberland, P. Pinter Jr, P.M. Waller, T.L. Thompson, Remote sensing of cotton nitrogen status using the canopy chlorophyll content index (CCCI), Trans. ASABe 51 (1) (2008) 73–82.
[42] A. Viña, A.A. Gitelson, A.L. Nguy-Robertson, Y. Peng, Comparison of different vegetation indices for the remote assessment of green leaf area index of crops, Remote Sens. Environ. 115 (12) (2011) 3468–3478.
[43] E.A.L. Salas, G.M. Henebry, A new approach for the analysis of hyperspectral data: Theory and sensitivity analysis of the Moment Distance Method, Remote Sens. 6 (1) (2013) 20–41.
[44] J.M. Chen, Evaluation of vegetation indices and a modified simple ratio for boreal applications, Canad. J. Remote Sens. 22 (3) (1996) 229–242.
[45] R. Gupta, S. Prasad, T. Nadham, G. Rao, Relative sensitivity of district mean RVI and NDVI over an agrometeorological zone, Adv. Space Res. 13 (5) (1993) 261–264.
[46] B. Matsushita, W. Yang, J. Chen, Y. Onda, G. Qiu, Sensitivity of the enhanced vegetation index (EVI) and normalized difference vegetation index (NDVI) to topographic effects: a case study in high-density cypress forest, Sensors 7 (11) (2007) 2636–2651.
[47] A. Bannari, D. Morin, F. Bonn, A. Huete, A review of vegetation indices, Remote Sens. Rev. 13 (1-2) (1995) 95–120.
[48] J. Payero, C. Neale, J. Wright, Comparison of eleven vegetation indices for estimating plant height of alfalfa and grass, Appl. Eng. Agric. 20 (3) (2004) 385–393.
[49] J. Dash, P. Curran, Evaluation of the MERIS terrestrial chlorophyll index (MTCI), Adv. Space Res. 39 (1) (2007) 100–104.
[50] F.-M. Wang, J.-f. Huang, Y.-l. Tang, X.-z. Wang, New vegetation index and its application in estimating leaf area index of rice, Rice Sci. 14 (3) (2007) 195–203.
[51] A.A. Gitelson, Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation, J. Plant Physiol. 161 (2) (2004) 165–173.
[52] G.E. Meyer, J.C. Neto, Verification of color vegetation indices for automated crop imaging applications, Comput. Electron. Agric. 63 (2) (2008) 282–293.
[53] Barbosa, B., Ferraz, G., Gonçalves, L., Marin, D., Maciel, D., Ferraz, P., & Rossi, G. (2019). RGB vegetation indices applied to grass monitoring: A qualitative analysis.
[54] H. Kukkonen, J. Rovamo, K. Tiippana, R. Näsänen, Michelson contrast, RMS contrast and energy of various spatial stimuli at threshold, Vision. Res. 33 (10) (1993) 1431–1436.
[55] J.T. Walton, D.J. Nowak, E.J. Greenfield, Assessing urban forest canopy cover using airborne or satellite imagery, Arbor. Urban Forestry 34 (6) (2008) 334–340.
[56] J.Y.-L. Chan, S.M.H. Leow, K.T. Bea, W.K. Cheng, S.W. Phoong, Z.-W. Hong, Y.-L. Chen, Mitigating the multicollinearity problem and its machine learning approach: a review, Mathematics 10 (8) (2022) 1283. https://www.mdpi.com/2227-7390/10/8/1283.
[57] P. Bermejo, J.A. Gámez, J.M. Puerta, Incremental wrapper-based subset selection with replacement: An advantageous alternative to sequential forward selection, in: 2009 IEEE Symposium on Computational Intelligence and Data Mining, 2009.
[58] H. Jabbar, R.Z. Khan, Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study), Comput. Sci. Commun. Instru. Dev. 70 (10.3850) (2015) 978–981.
[59] R.B. Bendel, A.A. Afifi, Comparison of stopping rules in forward "stepwise" regression, J. Am. Stat. Assoc. 72 (357) (1977) 46–53.
[60] L. Prado Osco, A.P. Marques Ramos, D. Roberto Pereira, É. Akemi Saito Moriya, N. Nobuhiro Imai, E. Takashi Matsubara, N. Estrabis, M. de Souza, J. Marcato Junior, W.N. Gonçalves, Predicting canopy nitrogen content in citrus-trees using random forest algorithm associated to spectral vegetation indices from UAV-imagery, Remote Sens. 11 (24) (2019) 2925.
[61] H. Yuan, G. Yang, C. Li, Y. Wang, J. Liu, H. Yu, H. Feng, B. Xu, X. Zhao, X. Yang, Retrieving soybean leaf area index from unmanned aerial vehicle hyperspectral remote sensing: Analysis of RF, ANN, and SVM regression models, Remote Sens. 9 (4) (2017) 309.
[62] G. Mountrakis, J. Im, C. Ogole, Support vector machines in remote sensing: A review, ISPRS J. Photogr. Remote Sens. 66 (3) (2011) 247–259.
[63] C.J. Ransom, N.R. Kitchen, J.J. Camberato, P.R. Carter, R.B. Ferguson, F. G. Fernández, D.W. Franzen, C.A. Laboski, D.B. Myers, E.D. Nafziger, Statistical and machine learning methods evaluated for incorporating soil and weather into corn nitrogen recommendations, Comput. Electron. Agric. 164 (2019) 104872.
[64] K. Subedi, B. Ma, Seed priming does not improve corn yield in a humid temperate environment, Agron. J. 97 (1) (2005) 211–218.
[65] P. Thanh Noi, M Kappas, Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery, Sensors 18 (1) (2017) 18.
[66] F. Fassnacht, F. Hartig, H. Latifi, C. Berger, J. Hernández, P. Corvalán, B. Koch, Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass, Remote Sens. Environ. 154 (2014) 102–114.