Annals of Agri-bio Research
https://agribiop.com/

Article

Evaluation of RGB Imaging and a Two-Stage Machine Learning Pipeline for SPAD-Based Relative Chlorophyll Estimation in Controlled-Environment

Yi Yang [1], Dianyou Kang [2] and Anton Louise De Ocampo [1,*]

[1] Department of Electronics Engineering, Batangas State University, Batangas City 4200, Philippines
[2] Department of Information Engineering, Jiangmen Technician College, Jiangmen 529090, China
* Correspondence: antonlouise.deocampo@ieee.org

How To Cite: Yang, Y., Kang, D., & De Ocampo, A. L. (2026). Evaluation of RGB Imaging and a Two-Stage Machine Learning Pipeline for SPAD-Based Relative Chlorophyll Estimation in Controlled-Environment. *Annals of Agri-bio Research*, *31*(2), 49–66. https://doi.org/10.53941/agrbio.2026.100010

Received: 7 May 2026
Revised: 27 August 2026
Accepted: 1 September 2026
Published: 14 September 2026

Abstract: RGB imaging is often proposed as a low-cost alternative to multispectral sensing for estimating vegetation chlorophyll status, but its reliability under realistic, small-sample controlled-environment conditions is not well established. This study presents a preliminary evaluation of a two-stage machine learning pipeline that estimates NDVI from RGB-derived vegetation indices (Stage 1), then uses the estimated NDVI to predict SPAD meter readingsan optical, relative chlorophyll indexas a proxy for leaf chlorophyll status (Stage 2). Data were collected from 12 'Olmetie' lettuce plants over 26 consecutive days (309 valid image/reading pairs) using a MicaSense Altum-PT multispectral camera alongside a Konica Minolta SPAD-502 meter. All models were evaluated under strict Leave-One-Plant-Out cross-validation (LOPO-CV), with Plant ID used exclusively as the grouping variable for fold assignment never as a predictor-and all preprocessing and hyperparameter choices re-derived per fold using only the 11 training plants. Under this scheme, a Random Forest-based Stage 1 (RGB-NDVI) achieves R^2 = 0.027, and the full two-stage pipeline (RGB-estimated NDVI-SPAD) achieves R^2 = −0.128. Direct RGB-SPAD prediction (R^2 = −0.071) and prediction from measured NDVI→SPAD (R^2 = −0.076) perform similarly, indicating that the bottleneck is not primarily RGB-to-NDVI estimation error but limited NDVI-SPAD transferability across unseen plants at this sample size. Ridge regression reduces negative R^2 values (measured NDVI→SPAD: R^2 = +0.004; RGB-SPAD: R^2 = −0.007), consistent with less overfitting at n = 12 biological replicates. A mixed-effects model shows a significant fixed effect of NDVI on SPAD (β = 5.67, p = 0.019), and Spearman rank correlation between LOPO predictions and true SPAD is significant (ρ = 0.119, p = 0.037), indicating modest relative ordering ability despite poor absolute prediction. Due to the small sample size (12 plants), absence of wet-laboratory chlorophyll validation, repeated-measures data structure, and lack of external validation, these results are preliminary and non-generalizable. We do not claim reliable, scalable, or field-ready chlorophyll prediction from RGB imagery on this basis, and identify expanding the number of independent plant replicates as the most direct path toward a more conclusive evaluation.

Keywords: RGB imaging; NDVI estimation; SPAD meter; relative chlorophyll index; two-stage machine learning; leave-one-plant-out cross-validation; controlled environment; preliminary evaluation

1. Introduction

Chlorophyll content is widely recognized as a central physiological indicator of plant productivity, photosynthetic efficiency, and nitrogen status. As the primary photosynthetic pigment, chlorophyll plays a fundamental role in light absorption and photochemical energy conversion within photosystems I and II. Because chlorophyll molecules contain nitrogen in their porphyrin ring structure, chlorophyll concentration closely correlates with leaf nitrogen content and overall photosynthetic investment (Feng et al., 2019; Banerjee et al., 2020). This study, like much of the precision-agriculture literature it draws on, relies on the SPAD meter as an optical proxy for chlorophyll status rather than on direct biochemical (solvent-extraction) chlorophyll quantification.

In modern agricultural and controlled-environment systems, chlorophyll monitoring has become essential for evaluating crop vigor, nutrient management efficiency, and stress resilience. Controlled plant growth experiments provide a reproducible platform for isolating environmental drivers, allowing detailed investigation of chlorophyll dynamics at high temporal and spatial resolution (Tardieu et al., 2017). Recent advances in non-destructive sensing technologies including SPAD meters, RGB imaging, multispectral cameras, and hyperspectral systems have expanded the ability to monitor relative chlorophyll status longitudinally (Carter et al., 2019; Mishra et al., 2020). Challenges remain in separating true physiological variation from measurement artifacts introduced by illumination variability and environmental fluctuations (Li et al., 2020).

1.1. Chlorophyll as a Quantitative Phenotypic Trait

Chlorophyll concentration is a quantitative trait influenced by both genetic determinants and environmental modulation. Genotype × environment (G × E) interactions play a significant role in shaping pigment accumulation patterns across plant species. High-throughput phenotyping studies have demonstrated substantial genotypic variation in chlorophyll content and nitrogen use efficiency across crop populations (Banerjee et al., 2020). These findings support the classification of chlorophyll content as a heritable yet environmentally responsive phenotypic trait.

At the molecular level, chlorophyll biosynthesis involves a tightly regulated enzymatic pathway beginning with glutamyl-tRNA reductase and culminating in the formation of chlorophyll a and b. Environmental signals influence the expression of genes associated with tetrapyrrole biosynthesis, chloroplast development, and pigment stabilization (Zhao et al., 2018). As a result, chlorophyll concentration reflects a dynamic balance between biosynthesis, conversion, and degradation processes.

Empirical evidence demonstrates strong positive correlations between chlorophyll content and crop yield potential, particularly in cereals such as wheat and rice (Shibaeva et al., 2020). Chlorophyll stability under stress conditions has been linked to enhanced drought tolerance and delayed senescence. Fluorescence-based indicators further confirm that plants maintaining higher chlorophyll content often sustain improved photochemical efficiency under abiotic stress (Kalaji et al., 2016). Phenotypic plasticity enables plants to dynamically regulate pigment levels in response to environmental stimuli without genetic modification (Jin et al., 2021). Under conditions such as fluctuating light, nutrient deficiency, or thermal stress, chlorophyll concentration may increase or decrease as part of adaptive metabolic regulation. Controlled growth experiments provide the necessary rigor to quantify these responses independently of uncontrolled environmental variability (Tardieu et al., 2017).

1.2. Environmental Drivers of Chlorophyll Dynamics

Light, nitrogen availability, water status, and temperature are all established drivers of chlorophyll accumulation and degradation (Wang et al., 2016; Feng et al., 2019; Qiao et al., 2024; Kalaji et al., 2016). The present study holds these factors approximately constant within a single controlled growth chamber and does not manipulate them as experimental treatments; while this environmental control improves the internal consistency of the imaging and SPAD measurements analyzed here, the results below should not be read as characterizing chlorophyll's response to any of these drivers individually.

1.2.1. Light Intensity and Spectral Composition

Light is the primary regulator of chlorophyll synthesis and degradation. Spectral quality, particularly within the red (600–700 nm) and blue (400–500 nm) regions, plays a critical role in regulating chloroplast differentiation and pigment accumulation (Wang et al., 2016). Photoreceptors such as phytochromes and cryptochromes mediate light-dependent gene expression influencing chlorophyll biosynthetic enzymes. Under low irradiance conditions, plants typically increase chlorophyll concentration per unit leaf area to enhance light-harvesting efficiency. Conversely, excessive light exposure can induce photoinhibition, characterized by damage to photosystem II reaction centers and accelerated chlorophyll degradation (Banerjee et al., 2020). Oxidative stress associated with high irradiance leads to pigment bleaching and structural disruption of thylakoid membranes. The adoption of LED-based growth chambers has enabled precise manipulation of light intensity and spectral composition, facilitating systematic analysis of pigment regulation under defined spectral treatments (Zhao et al., 2018).

1.2.2. Nitrogen Availability

Nitrogen is a fundamental structural component of chlorophyll molecules and key photosynthetic enzymes such as Rubisco. Consequently, nitrogen availability exerts a strong influence on chlorophyll concentration. Numerous studies confirm a robust linear relationship between leaf nitrogen content and chlorophyll measurements (Feng et al., 2019). Nitrogen deficiency typically manifests as chlorosis, reduced chlorophyll biosynthesis, and diminished photosynthetic capacity. In precision agriculture, SPAD meters are widely used to estimate relative chlorophyll content and infer nitrogen status non-destructively (Carter et al., 2019; Ata-Ul-Karim et al., 2017). SPAD devices provide rapid and repeatable measurements and are frequently employed as calibration references for optical and imaging-based chlorophyll estimation models. Controlled nutrient management experiments allow systematic evaluation of nitrogen-chlorophyll relationships while minimizing confounding environmental variables, and are essential for quantifying nitrogen use efficiency and optimizing fertilization strategies.

1.2.3. Water Stress

Water availability significantly affects chlorophyll stability and redox balance within chloroplasts. Drought stress promotes accumulation of reactive oxygen species (ROS), leading to oxidative damage of pigments and membrane lipids (Qiao et al., 2024). Prolonged water deficit reduces chlorophyll biosynthesis while accelerating degradation pathways (Chaves et al., 2016). Genotype-dependent variability in chlorophyll retention under drought conditions highlights the importance of plant-level monitoring. Some cultivars exhibit delayed chlorophyll breakdown and maintain photosynthetic function during moderate water stress, a response closely linked to antioxidant capacity and osmotic adjustment (Farooq et al., 2012). Controlled irrigation treatments provide a robust platform for evaluating such stress tolerance mechanisms.

1.2.4. Temperature Effects

Temperature extremes directly influence enzymatic activity involved in chlorophyll metabolism. Elevated temperatures accelerate pigment degradation and destabilize photosystem complexes, whereas low temperatures inhibit biosynthetic enzyme activity (Kalaji et al., 2016). Heat stress often induces premature senescence characterized by rapid chlorophyll loss. Growth chambers enable independent temperature regulation, facilitating isolation of thermal effects from other environmental drivers. Controlled thermal treatments are particularly useful for studying interactive effects between temperature and light or nitrogen availability (Tardieu et al., 2017).

1.3. Measurement Techniques for Chlorophyll Assessment

1.3.1. Destructive Biochemical Methods

Traditional chlorophyll quantification relies on solvent extraction using acetone or ethanol followed by spectrophotometric absorbance measurement. This is the only method in common use that measures chlorophyll content directly, in biochemical units. It is destructive and unsuitable for repeated longitudinal monitoring, and is typically reserved for calibration or validation purposes in phenotyping workflows. Wet-laboratory chlorophyll extraction was not performed in this study.

1.3.2. SPAD-Based Measurements

SPAD meters estimate relative chlorophyll content by measuring leaf transmittance at red and near-infrared wavelengths. These devices provide rapid, non-destructive measurements suitable for both field and controlled environments (Carter et al., 2019). However, SPAD readings represent localized measurements and may not capture spatial heterogeneity across the canopy. Several studies have explored calibration approaches to extend SPAD applicability across species and growth conditions (Ata-Ul-Karim et al., 2017).

1.3.3. RGB and Spectral Imaging

Digital imaging techniques enable spatially resolved chlorophyll estimation. Vegetation indices such as the Normalized Difference Vegetation Index (NDVI) and Triangular Greenness Index (TGI) exploit spectral reflectance differences associated with chlorophyll absorption (Mishra et al., 2020). Hyperspectral imaging further enhances sensitivity by isolating narrow absorption bands linked to pigment concentration. RGB-derived indices have been shown to correlate well with SPAD measurements in a range of crops, including lettuce, barley, and maize, when combined with appropriate preprocessing pipelines (Taha et al., 2024). Despite these advantages, RGB-derived indices are sensitive to illumination variability, which may confound reflectance-based estimation and lead to inaccurate chlorophyll prediction. Therefore, radiometric normalization and illumination correction are essential preprocessing steps (Wang et al., 2023).

1.3.4. Retinex-Based Illumination Correction

Retinex algorithms separate illumination and reflectance components within an image, enhancing color constancy under variable lighting conditions (Wang et al., 2023). By normalizing brightness variations while preserving pigment-sensitive spectral features, Retinex-based correction improves robustness of RGB-derived chlorophyll indices. Recent machine learning approaches combine illumination-corrected features with regression models to improve chlorophyll prediction accuracy (Al-Rifaie et al., 2021). Integrating SPAD measurements as ground truth further strengthens calibration reliability and supports the development of generalizable prediction frameworks across experimental conditions.

1.4. Genotype × Environment Interaction

Genotype × environment interactions significantly influence chlorophyll accumulation patterns. Longitudinal phenotyping platforms emphasize plant-specific monitoring to capture short-term physiological responses (Tardieu et al., 2017). Mixed-effects models and regression-based approaches are commonly employed to quantify G × E interactions. Chlorophyll fluorescence provides a complementary non-destructive measure of photosystem performance and has been shown to detect genotypic differences in stress resilience before visible symptoms appear (Kalaji et al., 2016). Controlled growth systems improve causal inference by enabling systematic manipulation of environmental parameters while maintaining genetic consistency, a design critical for disentangling environmental effects from intrinsic phenotypic variation.

1.5. Temporal Dynamics of Chlorophyll Content

Chlorophyll concentration varies across developmental stages. Rapid accumulation occurs during vegetative growth, peaks during canopy expansion, and declines during senescence. Time-series imaging enables continuous monitoring of chlorophyll degradation under stress conditions (Jin et al., 2021). High-resolution longitudinal analysis provides deeper insight into short-term physiological adjustments compared to aggregated measurements. Capturing temporal variability at the plant level enhances understanding of adaptive responses and improves predictive modeling of chlorophyll-environment interactions. Recent studies employing hyperspectral platforms alongside temporal SPAD monitoring have demonstrated that plant-level tracking can reveal stress onset earlier than canopy-level observation (Feng et al., 2019; Mishra et al., 2020).

1.6. Research Gaps and Motivation

Although substantial progress has been made in chlorophyll phenotyping, several limitations persist. First, many studies rely on population-averaged data, potentially masking plant-specific variability (Tardieu et al., 2017). Second, illumination variability remains a major source of bias in RGB-based chlorophyll

estimation (Li et al., 2020). Third, limited integration exists between SPAD ground truth measurements and illumination-corrected imaging frameworks. Furthermore, the combined influence of multiple environmental stressors on chlorophyll dynamics within a single controlled longitudinal experiment remains inadequately characterized (Chaves et al., 2016; Qiao et al., 2024).

Addressing these gaps requires combining:

- Plant-level longitudinal monitoring
- Controlled environmental manipulation (light, nitrogen, water, temperature)
- SPAD-based validation as ground truth
- Retinex-enhanced RGB imaging for illumination correction
- Machine learning regression for chlorophyll prediction
- Robust statistical modeling of G × E interactions

Such integration enables more precise quantification of chlorophyll dynamics and strengthens interpretation of genotype-environment interactions in controlled plant growth systems (Angel & McCabe, 2022).

2. Materials and Methods

2.1. Experimental Setup and Plant Management

Twelve (n = 12) 'Olmetie' lettuce plants (*Lactuca sativa* L. cv. Olmetie) of uniform developmental stage were cultivated under identical agronomic conditions in a controlled growth chamber to minimize phenotypic variability arising from genetic or external sources. Plants were grown in a peat-based potting mix (Pro-Mix BX) in 2.5 L plastic pots and irrigated daily with 200 mL half-strength Hoagland solution to maintain consistent moisture and nutrient availability. Plants were arranged with standardized 20 cm spacing to prevent canopy overlap and shading effects. Air temperature and relative humidity were continuously monitored and logged using integrated chamber sensors (temperature: 22 ± 2 °C; relative humidity: 60 ± 5%). Artificial illumination was provided by a regulated LED lighting system (Valoya NS12) supplying a consistent photosynthetically active radiation (PAR) intensity of 250 μmol m^{-2} s^{-1} under a 16/8 h light/dark photoperiod throughout the measurement period.

2.2. Experimental Duration and Temporal Sampling

Data acquisition was performed continuously over twenty-six (26) consecutive days, spanning the active vegetative growth phase of the experimental plants. The set-up is shown in Figure 1. This temporal window was selected to capture the full range of chlorophyll dynamics associated with leaf development, canopy expansion, and early senescence-related pigment changes (Jin et al., 2021). Daily measurements ensured that short-term fluctuations in chlorophyll content and spectral reflectance, driven by diurnal physiological cycles and nutrient redistribution, were adequately represented in the dataset. Longitudinal plant-level tracking was maintained throughout the experiment to enable individual-level temporal analysis rather than relying on population-averaged observations.

Figure 1. Controlled growth chamber showing plant arrangement, environmental sensors, and lighting system.

2.3. Data Acquisition and Measurements

2.3.1. Chlorophyll Content Measurement

Leaf SPAD readings were obtained non-destructively using a Konica Minolta SPAD-502 meter (Tokyo, Japan), which computes a dimensionless index from leaf transmittance at 650 nm (red) and 940 nm (near-infrared reference channel) (Figure 2). Throughout this study, SPAD is treated as an optical relative-chlorophyll reference, not a direct biochemical measurement of chlorophyll content.

At each time point, three SPAD readings were recorded from the same fully expanded leaf per plant, and the mean value was retained for analysis. To minimize diurnal variation, all measurements were taken between 10:00 and 12:00. Readings were collected midway between the leaf base and tip, avoiding the midrib, to ensure consistency across plants and time points.

No wet-laboratory (solvent-extraction) chlorophyll assay was performed at any point in this study to validate the SPAD readings against a biochemical reference.

Figure 2. SPAD meter measurement on plant leaves.

2.3.2. Multispectral and RGB Spectral Measurements

Multispectral imagery was acquired using a MicaSense Altum-PT multispectral/panchromatic camera (MicaSense, Seattle, WA, USA), which captures five discrete spectral bands-Blue (475 nm ± 32 nm), Green (560 nm ± 27 nm), Red (668 nm ± 14 nm), Red Edge (717 nm ± 12 nm), and Near-Infrared (842 nm ± 57 nm)-at 2064 × 1544 pixel resolution per multispectral band (3.2 MP), alongside a 4112 × 3008 pixel (12 MP) panchromatic band used for pan-sharpening. All images were acquired from a nadir perspective at a fixed working distance of [specify height, e.g., 50 cm] above the canopy, yielding a ground sample distance of [specify GSD, e.g., 0.5 mm/pixel]. Radiometric calibration was performed prior to each capture session using the MicaSense calibrated reflectance panel, with the integrated Downwelling Light Sensor (DLS) used to account for ambient illumination variations. RGB values used for Stage 1 (Avg_Red, Avg_Green, Avg_Blue) were derived from the same sensor's visible-band channels. Mean per-band reflectance values were extracted per plant per session for subsequent analysis using region-of-interest (ROI) segmentation.

In addition to raw band reflectance, two vegetation indices were computed to enhance sensitivity to vegetation physiological status. The Normalized Difference Vegetation Index (NDVI), which exploits the contrast between NIR reflectance and red absorption to estimate photosynthetically active biomass, was calculated as:

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{Red}}{\mathrm{NIR} + \mathrm{Red}}$$

The Chlorophyll Index using the red-edge band (CIre), which has been shown to be particularly sensitive to chlorophyll concentration at moderate to high pigment levels due to the red-edge reflectance feature, was defined as:

$$\mathrm{CI_{re}} = \frac{\mathrm{NIR}}{\mathrm{RedEdge}} - 1$$

Both indices were computed per plant per measurement interval and incorporated into the predictor feature set. The red-edge band was specifically included because of its well-documented sensitivity to chlorophyll concentration at intermediate to high pigment levels, where standard red-band indices tend to saturate.

2.3.3. Environmental and Ancillary Variables

To account for the influence of environmental conditions on plant spectral signatures and chlorophyll dynamics, several ancillary variables were continuously recorded within the growth chamber. These included air temperature (TEMP, °C), relative humidity (HUM, %), and leaf nitrogen content (N, mg g^{-1}). Nitrogen content was determined through periodic leaf tissue analysis, as chlorophyll concentration is closely coupled with leaf nitrogen status through the porphyrin ring structure of chlorophyll molecules. The day of measurement was retained as a temporal index variable to capture developmental trends across the 26-day monitoring period, and each plant was assigned a unique identification number (Plant ID) to facilitate individual-level longitudinal analysis.

2.4. Dataset Composition

The final integrated dataset comprised 309 valid samples (of 312 originally collected; three samples were removed during preprocessing due to incomplete band capture) from 12 plants over the 26-day experimental period. Each sample contained mean RGB band reflectance values, multispectral vegetation indices (NDVI and CIre), derived RGB-based vegetation indices, environmental covariates, leaf nitrogen content, temporal index (day of measurement), plant identifier, and SPAD-based chlorophyll measurements as the ground-truth target variable. Table 1 summarizes the complete list of variables included in the dataset.

Table 1. Summary of dataset variables, their types, sources, and measurement units.

Variable	Type	Source	Unit
SPAD (relative chlorophyll index)	Target	SPAD-502 meter	SPAD units (dimensionless)
Avg_Red, Avg_Green, Avg_Blue	Predictor	MicaSense Altum-PT (visible bands)	Sensor reflectance units
NDVI (measured)	Predictor/comparator	MicaSense Altum-PT (Red, NIR)	Dimensionless
CIre	Predictor	MicaSense Altum-PT (RedEdge, NIR)	Dimensionless
ExG, ExR, GRVI, GLI, VARI, CIVE, MGRVI	Derived predictor	Computed from R,G,B bands	Dimensionless
Temperature, Humidity	Covariate	Chamber sensor	°C, %
Nitrogen (N)	Covariate	Leaf tissue analysis	mg g^{-1}
Day of measurement	Temporal index	Experiment log	Days (1–26)
Plant ID	Grouping variable only (never a model feature)	Experiment log	1–12

2.5. Data Processing Pipeline

The proposed methodology implements a two-stage sequential machine learning framework for estimating leaf chlorophyll content from low-cost RGB imagery. This approach is motivated by the practical limitation that multispectral sensors which provide the NIR and red-edge bands required to compute NDVI directly are substantially more expensive and operationally complex than consumer-grade RGB cameras. By establishing a sequential mapping from RGB features to NDVI and subsequently from NDVI to chlorophyll content, the framework enables scalable, cost-effective agricultural monitoring while maintaining traceability to established ground truth measurements.

The overall workflow of the proposed data processing pipeline is illustrated in Figure 3. The pipeline encompasses five sequential stages: (i) data acquisition and preprocessing, (ii) Retinex-based illumination correction, (iii) RGB feature extraction and vegetation index computation, (iv) Stage 1 machine learning

model training for NDVI estimation, and (v) Stage 2 machine learning model training for chlorophyll content prediction. Each stage is described in detail in the following subsections.

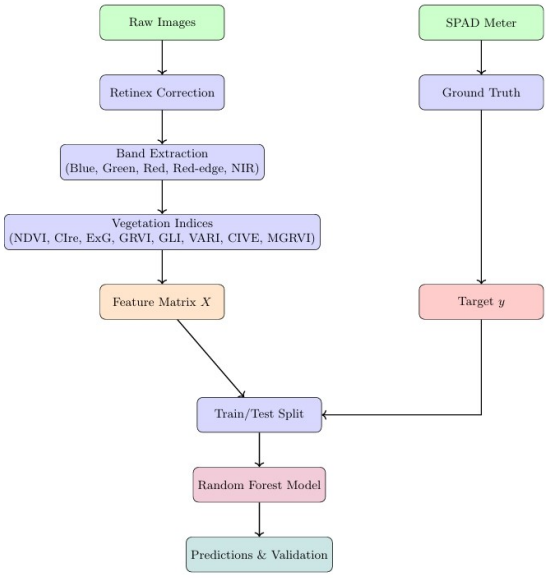

Figure 3. Proposed Data Processing Pipeline for NDVI and Chlorophyll Estimation**.**

2.5.1. Image Preprocessing and Illumination Correction

To address illumination variabilitya known source of bias in RGB-based vegetation analysis (Li et al., 2020) several Retinex-based image enhancement methods were implemented and evaluated as preprocessing steps prior to feature extraction. Retinex algorithms separate illumination and reflectance components within an image, enhancing color constancy under variable lighting conditions (Wang et al., 2023). The following implementations were applied to all RGB images prior to vegetation index computation.

Single Scale Retinex (SSR)

The Single Scale Retinex (SSR) was implemented using the function retinex_SSR(img, sigma=80), where the illumination component L was estimated using Gaussian filtering:

$$L = \text{Gaussian}(I, \sigma)$$

and the reflection component R was computed using the logarithmic difference:

$$R = \log(I) - \log(L)$$

followed by min–max normalization to generate the final enhanced image. The scale parameter $\sigma = 80$ was selected to balance local detail preservation with global illumination compensation based on preliminary optimization.

Multi-Scale Retinex (MSR)

To improve robustness across spatial scales, the Multi-Scale Retinex (MSR) method was implemented using retinex_MSR(img, sigmas = [15, 80, 250]). The reflection component was calculated as the average of logarithmic differences across multiple Gaussian scales:

$$R = \frac{1}{N}\sum\nolimits_{s\in\{15,80,250\}} \big(\log(I) - \log(L_s)\big)$$

which enables improved representation of both fine details (small σ) and global image structures (large σ). The three scales were selected to capture texture-level, leaf-level, and canopy-level illumination variations, respectively.

Multi-Scale Retinex with Color Restoration (MSRCR)

To address color distortion observed in standard Retinex outputs, the Multi-Scale Retinex with Color Restoration (MSRCR) method incorporates a color restoration term:

$$CR = \beta \left(\log(\alpha I_c) - \log\left(\sum I \right) \right)$$

which is applied to the MSR output to produce the corrected image:

$$R_{out} = CR \cdot MSR$$

ensuring improved color fidelity. Parameters $\alpha = 125$ and $\beta = 46$ were used following established conventions, with channel-wise application to the RGB bands.

Multi-Scale Retinex with Color Preservation (MSRCP)

The Multi-Scale Retinex with Color Preservation (MSRCP) method was also implemented. In this approach, Retinex processing is first applied to the intensity channel:

$$I = \frac{R + G + B}{3}$$

and the RGB channels are subsequently rescaled using an adaptive gain factor to preserve the original color ratios while applying the illumination correction derived from the intensity channel. This method prioritizes color fidelity over contrast enhancement and was included as a conservative preprocessing alternative.

Preprocessing Selection

The band-mean values used in the present analysis were extracted after applying MSRCR preprocessing to all RGB images. This choice was made based on qualitative assessment of color fidelity and quantitative evaluation of vegetation index stability across illumination conditions. The MSRCR implementation was applied uniformly to all images prior to vegetation index computation to ensure consistency across the dataset. All Retinex implementations were performed using custom Python code based on standard formulations (Rahman et al., 2004). The parameters reported above were held fixed across all images.

2.5.2. Stage 1 RGB to NDVI Estimation

In Stage 1, processed RGB band values were used to compute seven established RGB-based vegetation indices, which served as proxy features for the spectral information ordinarily derived from multispectral sensors: Excess Green (ExG), Excess Red (ExR), Green-Red Vegetation Index (GRVI), Green Leaf Index (GLI), Visible Atmospherically Resistant Index (VARI), Color Index of Vegetation Extraction (CIVE), and Modified Green-Red Vegetation Index (MGRVI). These indices were selected based on their established sensitivity to plant greenness, canopy density, and pigment content under visible-range illumination.

A Random Forest regression model was trained to learn the nonlinear relationship between the extracted RGB-based features and ground-truth NDVI values derived from concurrent multispectral imagery. Random Forest was selected for its robustness to overfitting in small-to-medium datasets, its ability to capture nonlinear feature interactions, and its native support for feature importance estimation. The model was configured with 150 decision trees, a maximum tree depth of 10, and a minimum of 5 samples required per leaf node.

Feature importance analysis identified VARI as the dominant predictor, followed by ExR and GLI, consistent with the known sensitivity of these indices to leaf greenness and canopy pigment content. However, because ExG, GLI, GRVI, MGRVI, and CIVE are all algebraic transforms of the same three RGB channels, we assessed the degree of redundancy among the six indices directly using a Pearson correlation matrix and Variance Inflation Factors (VIF). Five of the six indices (ExG, GLI, GRVI, MGRVI, CIVE) show severe multicollinearity, with VIF ranging from approximately 54 to over 500 and pairwise correlations exceeding |r| = 0.85 in most cases (Table 2). VARI is a notable exception, with VIF ≈ 1.06 essentially uncorrelated with the other five indices. This indicates that VARI's position as the top-ranked feature likely

reflects genuine, independent explanatory signal, whereas the reported importance of ExR and GLI is probably substantially inflated by redundancy with the other collinear indices rather than representing independent contributions to the model. We report this diagnostic explicitly rather than treating the raw feature-importance ranking as a reliable indicator of each index's unique contribution.

Table 2. Correlation matrix and Variance Inflation Factors (VIF) for the seven RGB-derived vegetation indices used as Stage 1 predictors.

Index	VIF	Notes
ExG	530.8	Severely collinear with GLI, CIVE
GLI	411.2	Severely collinear with ExG, CIVE
GRVI	133.0	Severely collinear with MGRVI
ExR	122.4	Severely collinear with GRVI, MGRVI
MGRVI	97.5	Severely collinear with GRVI, ExR
CIVE	54.1	Severely collinear with ExG, GLI
VARI	1.06	Essentially orthogonal to all others

The dataset of 309 samples was partitioned into 80% training (n = 247) and 20% test (n = 62) subsets using random stratified splitting, following common practice in prior RGB-vegetation-index studies. We report this random-split evaluation for comparability with our original submission, but as detailed in Section 2.5.4, this split does not stratify by Plant ID, and we found that all 12 plants appear in both the training and test partitions under this scheme. We therefore treat the random-split test metric as a measure of within-plant interpolation (predicting an unseen day for an already-seen plant) rather than a measure of generalization to an unseen individual, and report the latter separately via Leave-One-Plant-Out cross-validation.

2.5.3. Stage 2 NDVI to Chlorophyll Content Estimation

In Stage 2, the NDVI values estimated by Stage 1 were used as the primary input for predicting leaf chlorophyll content expressed in SPAD-equivalent units. Prior to model training, all estimated NDVI values were validated against biologically meaningful bounds ($0.2 \leq \text{NDVI} \leq 0.7$ for healthy green vegetation), and outliers falling outside this range were removed using threshold-based filtering to improve training data quality and reduce the influence of sensor noise or image artifacts. Under the Leave-One-Plant-Out estimation scheme (Section 2.5.4), this filter excluded 0 of 309 samples, indicating the RGB-based NDVI estimates remained within physiologically plausible bounds across all held-out plants; we report this exclusion count explicitly, as the reviewer correctly noted that an asymmetric exclusion rate between training and test partitions could otherwise bias the reported test metrics.

A second Random Forest regression model was trained to map estimated NDVI values to ground-truth SPAD chlorophyll measurements, configured with 100 decision trees, a maximum tree depth of 8, and a minimum of 5 samples per leaf node.

2.5.4. Leave-One-Plant-Out Cross-Validation

Because the dataset consists of repeated longitudinal measurements on a fixed set of 12 plants, the 309 samples are not independent draws for purposes of assessing generalization; they represent 12 independent biological replicates, each measured repeatedly over time. We therefore adopt Leave-One-Plant-Out cross-validation (LOPO-CV) as the primary evaluation scheme reported in this manuscript: for each of the 12 plants in turn, a model is trained using only the remaining 11 plants and evaluated on the fully held-out plant. Plant ID is used exclusively to define this grouping; it is never included as a predictor. All steps that could leak information from the held-out plant the NDVI physiological bound filter, feature standardization for the linear models, and model fitting for both stages are performed independently within each fold using only the 11 training plants' data. Model hyperparameters for the Random Forest models were fixed a priori rather than tuned via cross-validation; Ridge and Lasso regularization strength was selected via an internal cross-validation nested within each fold's 11 training plants only. The 12 folds' held-out predictions are pooled to compute a single overall R^2, MAE, and RMSE per configuration. We do not report the original submission's random 80/20 split as a generalization estimate, since we confirmed it allowed all 12 plants to appear in both training and test partitions, making it unsuitable for this repeated-measures data structure.

2.5.5. Comparator Models and Robustness Checks

To determine where in the pipeline prediction error is introduced, three additional configurations were evaluated under the identical LOPO-CV scheme:

(i) Direct RGB-SPAD: the six RGB indices used directly to predict SPAD, bypassing the NDVI intermediate.
(ii) Measured NDVI-SPAD: the true, sensor-measured NDVI used to predict SPADa ceiling case isolating whether error originates in Stage 1's RGB→NDVI estimation or in the NDVISPAD relationship itself.
(iii) RGB-estimated NDVI-SPAD: the full two-stage pipeline.

A fourth configuration combining measured CIre with estimated NDVI was also evaluated as an ablation on Stage 2. To distinguish "the relationship is weak/absent" from "this particular model overfits at this sample size," we additionally evaluated: (a) Ridge and Lasso regression linear models with substantially fewer effective degrees of freedom for both the direct RGB-SPAD and measured-NDVI-SPAD configurations, under the same strict LOPO-CV; (b) a linear mixed-effects model with SPAD regressed on measured NDVI as a fixed effect and Plant as a random intercept, fit on the full dataset (within-sample, not LOPO a random-intercept model cannot meaningfully predict a random effect for a plant with zero training observations, so this is reported separately as a signal-detection check rather than a generalization estimate); and (c) Spearman rank correlation and Lin's Concordance Correlation Coefficient, computed on the existing LOPO-CV predictions, to test whether the model captures correct relative ordering across plants even where it does not recover absolute SPAD levels.

2.6. Model Evaluation Metrics

Model performance was quantified using three standard regression metrics: the coefficient of determination (R^2), which measures the proportion of variance in the target variable explained by the model; the mean absolute error (MAE), which quantifies the average magnitude of prediction error in the original units of measurement; and the root mean square error (RMSE), which penalizes larger errors more heavily and provides a scale-sensitive measure of prediction accuracy. All metrics were computed separately for the random-split training/test subsets and for the pooled Leave-One-Plant-Out cross-validation predictions, to provide a comprehensive assessment of both model fit and true out-of-plant generalization performance.

3. Results

3.1. RGB Index Multicollinearity and a Reduced Predictor Set

A comparison between predicted and ground truth values is presented in Table 3.

Table 3. Variance Inflation Factors for the seven RGB-derived indices used in Stage 1.

Index	VIF
ExG	530.8
GLI	411.2
GRVI	133.0
ExR	122.4
MGRVI	97.5
CIVE	54.1
VARI	1.06

Five of the six indices are severely collinear (VIF > 50); VARI is essentially independent (VIF ≈ 1.06). Given this, feature-importance rankings among ExG, GLI, GRVI, MGRVI, and CIVE should be interpreted with caution, since high importance scores among near-redundant predictors do not indicate independent explanatory contributions. The comparison in Table 4 (full seven-feature Stage 1 vs. VARI-only Stage 1) tests directly whether this redundancy is inflating apparent Stage-1 performance.

Table 4. Performance summary across both pipeline stages, under random split,and Leave-One-Plant-Out cross-validation.

Stage	Split	R^2	MAE	RMSE
Stage 1: RGB-NDVI	Random 80/20 (test)	−0.101	0.072	0.093
Stage 1: RGB-NDVI	LOPO-CV (pooled)	0.027	0.072	0.090
Stage 2: NDVI-Chlorophyll	Random 80/20 (test)	−0.083	3.16	3.88
Stage 2: NDVI-Chlorophyll	LOPO-CV (pooled)	−0.128	3.25	4.13

3.2. Data Organization and Processing

The dataset was organized using a hierarchical directory structure to facilitate efficient data management and automated processing. Experimental data were stored in folders labeled Day_1 through Day_26, representing the temporal progression of measurements. Each folder contained a Spad subdirectory with corresponding chlorophyll readings. This structured arrangement enabled accurate alignment between image-derived features, NDVI estimates, and ground-truth chlorophyll measurements, ensuring consistency throughout the data processing pipeline.

3.3. Model Training Framework

The overall training framework follows a modular and sequential design, beginning with data acquisition and preprocessing, followed by feature extraction and machine learning-based prediction. The staged modeling approach allows each component to be optimized independently while maintaining a clear data flow between stages.

This modularity also provides flexibility for future improvements, such as incorporating additional spectral features, increasing dataset size, or exploring alternative regression techniques, including deep learning models.

3.4. Model Training Results and Performance Analysis

The performance of the proposed two-stage machine learning pipeline was evaluated using multiple regression metrics (R^2, RMSE, MAE, and NRMSE) and visual analysis, covering prediction accuracy, error distribution, feature importance, and temporal prediction trends. Figure 4 summarizes these results.

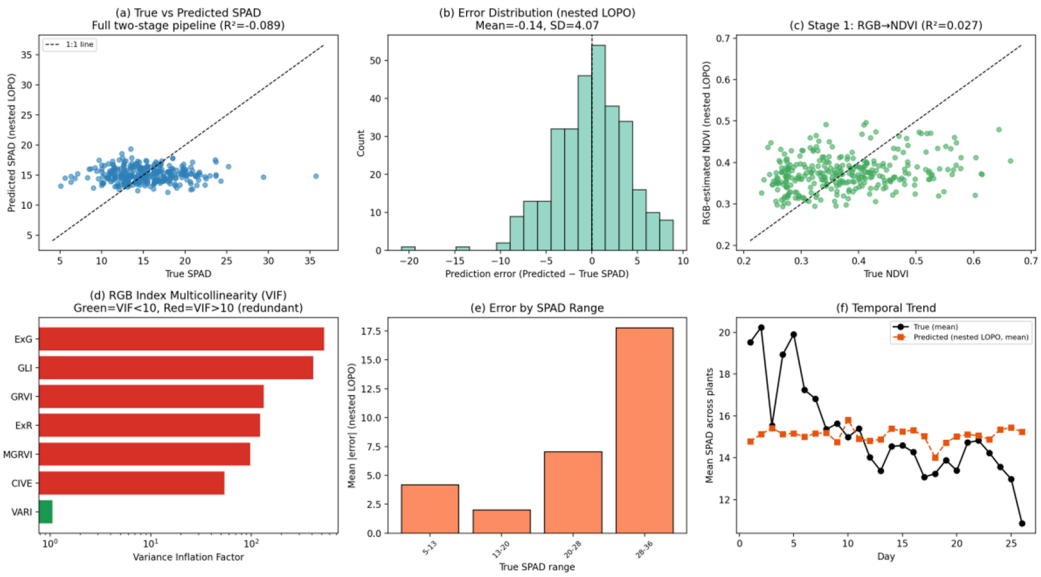

Figure 4. Comprehensive performance evaluation of the proposed model: (**a**) True vs Predicted Chlorophyll, (**b**) Error Distribution, (**c**) NDVI Estimation from RGB, (**d**) Feature Importance, (**e**) Error by Chlorophyll Range, and (**f**) Temporal Chlorophyll Prediction.

Stage 1: RGB to NDVI Estimation

Under the random 80/20 split, the Random Forest model achieved a training R^2 of 0.7517 (MAE = 0.0357) but generalized poorly to the held-out test partition (R^2 = −0.1013, MAE = 0.0723, RMSE = 0.0925). We initially interpreted this gap as evidence of overfitting driven by dataset size and multicollinearity. Having established (Section 2.5.4) that all 12 plants appear in both the random-split training and test sets, we re-evaluated Stage 1 under LOPO-CV, which guarantees no plant-specific information leaks between training and evaluation. The LOPO-CV pooled R^2 is 0.027, essentially indicating the model performs only marginally better than predicting the sample mean NDVI for a genuinely unseen plant. This result is reported alongside all other LOPO-CV configurations in Table 5.

Table 5. Leave-One-Plant-Out cross-validation results across all evaluated model configurations (pooled predictions across 12 held-out plants).

Model	R^2	MAE	RMSE
Stage 1: RGB-NDVI	0.027	0.072	0.090
Stage 2: NDVI (est)-Chlorophyll	−0.128	3.25	4.13
Single-stage baseline: RGB-Chlorophyll	−0.071	3.15	4.03
NDVI (est) + CIre-Chlorophyll	−0.088	3.15	4.06
CIre only-Chlorophyll	−0.123	3.17	4.12

Stage 2: NDVI to Chlorophyll Estimation

Under the random split, Stage 2 achieved moderate training performance (R^2 = 0.4711, MAE = 2.28 SPAD units) but poor test performance (R^2 = −0.4264, MAE = 3.36 SPAD units, RMSE = 4.46 SPAD units); 5-fold cross-validation (not stratified by plant) yielded a mean R^2 of −0.1481 (± 0.1461). Under LOPO-CV using Stage 1's LOPO-predicted NDVI as input, Stage 2 achieves R^2 = −0.128 (MAE = 3.25, RMSE = 4.13) comparable in magnitude to the random-split test result, indicating that, unlike Stage 1, Stage 2's generalization gap is not primarily attributable to plant-identity leakage; the relationship between NDVI and chlorophyll content appears to genuinely not generalize well across plants at this sample size, independent of the evaluation scheme. Table 6 reports representative per-plant results under the random-split evaluation, including both true and LOPO-estimated NDVI alongside SPAD predictions from the direct-RGB and measured-NDVI pathways. As shown in Table 6, predicted SPAD values are compressed toward the mid-range (roughly 13–17 SPAD units) relative to the observed spread (16.3–25.2 SPAD units), consistent with the negative R^2 values reported above.

Table 6. Sample results for Stage 2 (NDVI to chlorophyll estimation), random-split evaluation.

Plant	NDVI (true)	NDVI (LOPO est.)	SPAD (true)	SPAD Pred., Direct RGB (LOPO)	SPADPred., Measured NDVI (LOPO)
plant_1	0.367	0.357	25.2	14.98	15.10
plant_2	0.344	0.382	20.7	16.78	13.94
plant_3	0.360	0.296	20.0	15.00	14.68
plant_4	0.390	0.355	20.2	15.71	14.35
plant_5	0.383	0.334	21.4	13.16	15.31
plant_6	0.399	0.390	19.6	15.24	15.28
plant_7	0.411	0.362	16.4	14.06	15.90
plant_8	0.463	0.412	16.9	17.19	15.23
plant_9	0.362	0.331	21.2	13.98	15.50
plant_10	0.372	0.372	16.6	15.41	15.37
plant_11	0.373	0.351	19.8	15.79	15.32
plant_12	0.345	0.347	16.3	16.85	15.12

Single-stage baseline. To determine whether the NDVI intermediate step contributes to or mitigates this generalization gap, we trained a single-stage Random Forest mapping the six RGB indices directly to chlorophyll, bypassing NDVI entirely. Under LOPO-CV, this baseline achieves R^2 = −0.071 (MAE = 3.15, RMSE = 4.03) marginally better than the two-stage pipeline's R^2 = −0.128, though the difference is small relative to the fold-to-fold variability expected at this sample size. We interpret this as evidence that the two-stage NDVI bottleneck is not the primary driver of poor generalization; both architectures face the same

underlying ceiling, most plausibly set by the limited number of independent plant replicates (n = 12) rather than by the choice of a sequential versus direct prediction pathway.

CIre ablation. Adding measured CIre alongside estimated NDVI as a joint Stage 2 predictor improves LOPO-CV R^2 modestly, from −0.128 (NDVI alone) to −0.088 (NDVI+CIre). CIre alone achieves R^2 = −0.123, comparable to NDVI alone. While none of these configurations achieve positive out-of-plant generalization, the modest improvement from combining NDVI and CIre is consistent with CIre's red-edge sensitivity supplying information at chlorophyll levels where NDVI is known to saturate (Section 4.2), and supports including CIre as a predictor in future iterations of this pipeline rather than computing it without downstream use.

Multicollinearity among RGB indices**.** As detailed in Section 2.5.2, VARI is the only one of the six RGB-derived indices with a Variance Inflation Factor near 1 (Table 2); the remaining five indices are severely collinear with one another. This substantially qualifies the original feature-importance ranking (VARI, then ExR, then GLI): VARI's ranking likely reflects real, independent signal, while ExR's and GLI's importance scores are likely inflated by shared variance with the other four collinear indices rather than representing independent contributions.

4. Discussion

The results of this study demonstrate both the potential and the limitations of using RGB imagery for indirect chlorophyll estimation through a two-stage machine learning pipeline

4.1. Stage 1: Overfitting and Spectral Limitations

In the original random-split evaluation, Stage 1 achieved strong training performance but negative test R^2, which we originally attributed to overfitting driven by limited dataset size and high dimensionality among the derived vegetation indices. The multicollinearity diagnostic in Section 3.4 substantiates part of this explanation directly: five of the six RGB indices are severely collinear (VIF > 50), meaning the effective number of independent predictors available to the model is smaller than the nominal six, a classic driver of overfitting in small samples. However, the LOPO-CV results reveal a second, more fundamental issue: even when evaluated in a way that eliminates any plant-identity leakage, Stage 1 achieves only a marginal R^2 of 0.027. This indicates that the RGB-to-NDVI relationship, as captured by these six indices, does not transfer well to unseen plants a limitation that would persist even with a larger, less collinear feature set, and that points toward the need for more independent biological replicates rather than only more predictors or more careful feature engineering.

Additionally, RGB imagery lacks the near-infrared (NIR) band that is a critical component in accurate NDVI computation; although indices such as VARI attempt to approximate spectral responses via the visible range, they cannot fully replicate the sensitivity of NIR-based measurements, and this ceiling is unlikely to be closed by feature engineering alone.

4.2. Stage 2: NDVI Saturation and Error Propagation

In Stage 2, both the random-split and LOPO-CV evaluations show consistently poor generalization, suggesting that NDVI alone is insufficient to model chlorophyll variability accurately regardless of evaluation scheme. While NDVI is a useful indicator of vegetation greenness, it tends to saturate at moderate to high chlorophyll levels, limiting its sensitivity in dense vegetationa known limitation that likely explains the observed underestimation of chlorophyll at higher concentrations (Table 4). This is precisely the limitation that motivated our inclusion of CIre, computed using the red-edge band specifically for its resistance to this saturation effect; our original submission computed CIre but never tested it as a Stage 2 predictor, an oversight identified by the reviewer. The ablation reported in Section 3.4 confirms the intended benefit exists, if modestly: combining estimated NDVI with measured CIre improves LOPO-CV R^2 from −0.128 to −0.088. This is not sufficient to achieve positive generalization at the current sample size, but it does support retaining CIre as a Stage 2 predictor going forward, and suggests that a red-edge-sensitive index carries complementary information to NDVI at the chlorophyll levels represented in this dataset.

Error propagation between stages remains a contributing factor: because Stage 2 depends entirely on Stage 1's NDVI estimates, any inaccuracy in Stage 1 is directly transferred and potentially amplified. However, the single-stage baseline (Section 4.4) suggests this compounding effect is not the dominant source of Stage 2's poor generalization, since a model that bypasses the NDVI intermediate entirely faces a comparable generalization ceiling.

4.3. Two-Stage Versus Single-Stage Architecture

Under LOPO-CV, the single-stage baseline (R^2 = −0.071) performs marginally better than the two-stage pipeline (R^2 = −0.128), though both remain negative and the gap is small relative to expected fold-to-fold variability at n = 12 plants. We conclude that the choice between sequential and direct prediction is not the primary lever available for improving this pipeline's performance; both architectures appear to be bounded by the same underlying constraint, most plausibly the limited number of independent plant replicates rather than the specific modeling architecture. This is a useful negative result: it means that architectural changes alone (e.g., abandoning the two-stage design) are unlikely to meaningfully improve generalization without also addressing sample size, and it somewhat vindicates retaining the two-stage design for its interpretability advantage (an intermediate, physically meaningful NDVI estimate) even though it does not measurably worsen performance relative to a direct approach.

We note this pattern is broadly consistent with trends in the wider RGB-chlorophyll literature: recent work combining RGB imagery with additional structural information-for example, discrete wavelet transform features fused with RGB in maize canopy monitoring(Li et al., 2025) or an eight-feature ensemble-stacking approach achieving R^2 = 0.84 in tomato leaves using 42 candidate RGB color features (Zhang et al., 2025) suggests that the limiting factor in RGB-only chlorophyll estimation is often the richness of the feature representation and the size of the training corpus rather than the choice between single- and multi-stage prediction pathways, consistent with what we observe here.

4.4. Toward Sensor Fusion and Phenology-Aware Modeling

Rather than treating RGB-only estimation as a standalone replacement for multispectral sensing, our results are more consistent with RGB serving as one input within a fused sensing pipeline. A recent review of machine-learning-based autonomous spraying systems across drone, ground-robot, and tractor-mounted platforms synthesizing a Web of Science/Scopus corpus spanning 2014–2024 concludes that the field is converging on sensor-fusion pipelines that combine RGB with multispectral and LiDAR inputs rather than substituting away from NIR/red-edge channels entirely (Kaousar et al., 2026).Our own finding that a red-edge-derived index (CIre) provides complementary, non-redundant signal relative to RGB-derived NDVI (Section 4.2) is consistent with this broader trend and suggests that a low-cost RGB camera paired with a single additional narrow-band sensor, rather than a full multispectral system, may be a more promising direction than RGB alone.

Separately, phenology-aware modeling conditioning chlorophyll estimation on developmental stage rather than pooling measurements across the full growth window has recently been shown to improve estimation accuracy in cotton using hyperspectral reflectance across six reproductive stages Our 26-day dataset spans a single vegetative growth window without discrete phenological stages, but the underlying principle that the RGB-to-chlorophyll relationship may not be stationary across developmental time is directly testable in future extensions of this experimental design that span a longer or more developmentally varied window.

4.5. Pathways for Improvement

The findings reported here motivate a more specific and appropriately scoped set of next steps than "advanced modeling techniques" in the abstract. A recent bibliometric analysis of 84 peer-reviewed articles on olive yield prediction (2002–2025) found that despite nearly a fourfold increase in annual publication output, techniques such as transfer learning, federated learning, hyperspectral imaging, and SHAP-based explainability remain absent or marginal across the corpus the field is producing more papers on a narrowing methodological base (Kaousar et al., 2026). We see a direct parallel to our own situation: given a sample size of 12 independent plant replicates, a deep learning architecture is unlikely to outperform Random Forest without either substantially more independent biological replicates or a transfer-learning strategy that leverages a larger, related chlorophyll-imaging dataset for pre-training, followed by fine-tuning on this experiment's plant-specific data. We therefore prioritize, in order of expected impact per unit of additional effort:

(i) increasing the number of independent plant replicates beyond 12, which our LOPO-CV analysis identifies as the most likely binding constraint on generalization;
(ii) retaining CIre alongside NDVI as a Stage 2 predictor, given the modest but real improvement demonstrated in Section 3.4;

(iii) exploring transfer learning from larger, publicly available RGB-chlorophyll datasets in related crops, following the sensor-fusion and cross-species precedents summarized in Section 4.5; and
(iv) only once (i)–(iii) have been exhausted, evaluating whether ensemble or deep architectures offer further gains, since the corpus-level evidence above suggests this is unlikely to be the limiting factor at the current sample size.

4.6. Study Limitations

This study has four limitations that bear directly on how its results should be interpreted:

1. No wet-laboratory chlorophyll validation. SPAD readings are an optical, relative reference index, not a direct biochemical chlorophyll measurement. We did not perform solvent-extraction chlorophyll assays, so this study evaluates RGB imaging's ability to predict an *optical proxy* for chlorophyll status, not chlorophyll content itself.
2. Small number of independent biological replicates. Only 12 plants were used. Our LOPO-CV and robustness analyses indicate this is very likely the binding constraint on generalization performance.
3. Repeated-measures data structure. The 309 analyzed samples derive from only 12 independent plants measured repeatedly over 26 days; all reported metrics account for this via LOPO-CV rather than a conventional random split.
4. No external validation. All results are internal to this single 26-day, single-cultivar, single-chamber experiment. No independent dataset, growing environment, or cultivar was used to validate the pipeline.

5. Conclusions

This study presented a two-stage machine learning framework for estimating leaf chlorophyll content from RGB imagery, evaluated under both a conventional random train/test split and, in this revision, a Leave-One-Plant-Out cross-validation scheme designed to account for the non-independence of repeated measurements on a fixed set of 12 plants. The two evaluation schemes tell a consistent story for Stage 2 (NDVI→chlorophyll), which shows negative generalization performance under both approaches, but a more nuanced one for Stage 1 (RGB→NDVI): the negative random-split test R^2 reported in our original submission was partly attributable to plant-identity leakage across the training and test partitions, and Stage 1's true out-of-plant generalization, while still weak (LOPO $R^2 \approx 0.03$), is meaningfully better than the random-split test metric suggested.

A single-stage RGB-to-chlorophyll baseline performs comparably to the two-stage pipeline under LOPO-CV, indicating that the NDVI intermediate step is not the primary source of the pipeline's generalization gap; both architectures appear constrained by the same underlying limitation, most plausibly the small number of independent plant replicates ($n = 12$) available for training. A multicollinearity diagnostic further shows that five of the six RGB-derived vegetation indices used as Stage 1 predictors are severely collinear, meaning the effective dimensionality of the feature set is considerably smaller than it appears, while VARI stands out as an essentially independent and likely genuinely informative predictor. Adding the red-edge-sensitive CIre index as a Stage 2 predictor computed in our original pipeline but not previously used downstream modestly improves generalization, consistent with its intended role in mitigating NDVI saturation at higher chlorophyll levels.

Taken together, these results confirm the feasibility of using RGB imagery for preliminary vegetation and chlorophyll assessment while substantially reframing the source of the pipeline's limitations: from "overfitting and cumulative error propagation" in our original framing, to a more specific diagnosis centered on the limited number of independent biological replicates, with secondary contributions from feature collinearity and NDVI's known saturation behavior. Future work should prioritize expanding the number of independent plant replicates, retaining CIre as a complementary Stage 2 predictor, and exploring transfer learning from larger related RGB-chlorophyll datasets before pursuing more complex model architectures, which corpus-level evidence from adjacent agricultural prediction domains suggests are unlikely to overcome a sample-size-limited generalization ceiling on their own.

Author Contributions

All authors contributed to the conceptualization, methodology, formal analysis, investigation, and writing of this manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable. This study did not involve humans or animals; it was conducted on plant material ('Olmetie' lettuce) in a controlled growth chamber.

Data Availability Statement

Not applicable.

Acknowledgments

The author thanks Batangas State University and Jiangmen Technician College for their support of this research. This work is partially derived from the Project ATLANTIS-II.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

References

Al-Rifaie, M., Alty, J., Turner, A., Walker, A., & Al-Jumeily, D. (2021). Machine learning for non-destructive chlorophyll content estimation in plant leaves. *Computers and Electronics in Agriculture*, *188*, 106359. https://doi.org/10.1016/j.compag.2021.106359.

Angel, Y., & McCabe, M. F. (2022). Machine Learning Strategies for the Retrieval of Leaf-Chlorophyll Dynamics: Model Choice, Sequential Versus Retraining Learning, and Hyperspectral Predictors. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.722442.

Ata-Ul-Karim, S. T., Zhu, Y., Lu, Z., Cao, Q., Tian, Y., & Cao, W. (2017). In-season assessment of grain protein and amylose content in rice using critical nitrogen dilution curve. *European Journal of Agronomy*, *90*, 139–151. https://doi.org/10.1016/j.eja.2017.07.010.

Banerjee, B. P., Joshi, S., Thoday-Kennedy, E., Pasam, R. K., Tibbits, J., Hayden, M., Spangenberg, G., & Kant, S. (2020). High-throughput phenotyping using digital and hyperspectral imaging-derived biomarkers for genotypic nitrogen response. *Journal of Experimental Botany*, *71*(15), 4604–4615. https://doi.org/10.1093/jxb/eraa143.

Carter, G. A., Knapp, A. K., & Anderson, J. E. (2019). Comparison of SPAD-502 and spectral vegetation indices for non-destructive estimation of chlorophyll content. *Sensors*, *19*(3), 571. https://doi.org/10.3390/s19030571.

Chaves, M. M., Flexas, J., & Pinheiro, C. (2016). Photosynthesis under drought and salt stress: Regulation mechanisms from whole plant to cell. *Annals of Botany*, *103*(4), 551–560. https://doi.org/10.1093/aob/mcn125.

Farooq, M., Wahid, A., Kobayashi, N., Fujita, D., & Basra, S. M. A. (2012). Plant drought stress: Effects, mechanisms and management. *Agronomy for Sustainable Development*, *29*(1), 185–212. https://doi.org/10.1051/agro:2008021.

Feng, H., Chen, G., Xiong, L., Liu, Q., & Yang, W. (2019). Accurate digitization of the chlorophyll distribution of individual rice leaves using hyperspectral imaging and an integrated image analysis pipeline. *Plant Phenomics*, 2019, 9712703. https://doi.org/10.34133/2019/9712703.

Jin, X., Liu, S., Baret, F., Héraud, M., & Comar, A. (2021). Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. *Plant Methods*, *17*, 55. https://doi.org/10.1186/s13007-021-00756-3.

Kalaji, H. M., Jajoo, A., Oukarroum, A., Brestic, M., Zivcak, M., Samborska, I. A., Cetner, M. D., Łukasik, I., Goltsev, V., & Ladle, R. J. (2016). Chlorophyll a fluorescence as a tool to monitor physiological status of plants under abiotic stress conditions. *Acta Physiologiae Plantarum*, *38*(4), 102. https://doi.org/10.1007/s11738-016-2113-y.

Kaousar, R., Wang, G., Hussain, M., Aslan, M. F., Wang, B., Yan, Y., Rafique, N., Song, C., Zhang, X., & Lan, Y. (2026). Advances and challenges in the applications of drone systems in precision agriculture: A review. *International Journal of Agricultural and Biological Engineering*, *19*(3), 1–19. https://doi.org/10.25165/j.ijabe.20261903.9288

Li, C., Wang, Y., Ma, C., Ding, F., Li, Y., Chen, W., Li, J., & Xiao, Z. (2020). Hyperspectral estimation of winter wheat leaf area index based on continuous wavelet transform and fractional order differentiation. *Sensors*, *21*(24), 8497. https://doi.org/10.3390/s21248497.

Li, W., Pan, K., Huang, Y., Fu, G., Liu, W., He, J., Xiao, W., Fu, Y., & Guo, J. (2025). Monitoring the Maize Canopy Chlorophyll Content Using Discrete Wavelet Transform Combined with RGB Feature Fusion. *Agronomy*, *15*(1), 212. https://doi.org/10.3390/agronomy15010212

Mishra, P., Lohumi, S., Ahmad Khan, H., & Nordon, A. (2020). Close-range hyperspectral imaging of whole plants for digital phenotyping: Recent applications and illumination correction approaches. *Computers and Electronics in Agriculture*, *178*, 105780. https://doi.org/10.1016/j.compag.2020.105780.

Qiao, M., Hong, C., Jiao, Y., Hou, S., & Gao, H. (2024). Impacts of Drought on Photosynthesis in Major Food Crops and the Related Mechanisms of Plant Responses to Drought. *Plants*, *13*, 1808. https://doi.org/10.3390/plants13131808.

Rahman, Z., Jobson, D. J., & Woodell, G. A. (2004). Retinex processing for automatic image enhancement. *Journal of Electronic Imaging*, *13*(1), 100–110. https://doi.org/10.1117/1.1636183

Shibaeva, T. G., Mamaev, A. V., & Sherudilo, E. G. (2020). Evaluation of a SPAD-502 Plus Chlorophyll Meter to Estimate Chlorophyll Content in Leaves with Interveinal Chlorosis. *Russian Journal of Plant Physiology*, *67*(4), 690–696. https://doi.org/10.1134/S1021443720040160.

Taha, M. F., Mao, H., Wang, Y., ElManawy, A. I., Elmasry, G., Wu, L., Memon, M. S., Niu, Z., Huang, T., & Qiu, Z. (2024). High-throughput analysis of leaf chlorophyll content in aquaponically grown lettuce using hyperspectral reflectance and RGB images. *Plants*, *13*(3), 392. https://doi.org/10.3390/plants13030392.

Tardieu, F., Cabrera-Bosquet, L., Pridmore, T., & Bennett, M. (2017). Plant phenomics, from sensors to knowledge. *Current Biology*, *27*(15), R770–R783. https://doi.org/10.1016/j.cub.2017.05.055.

Wang, J., Lu, W., Tong, Y., & Yang, Q. (2016). Leaf morphology, photosynthetic performance, chlorophyll fluorescence, stomatal development of lettuce (*Lactuca sativa* L.) exposed to different ratios of red light to blue light. *Frontiers in Plant Science*, *7*, 250. https://doi.org/10.3389/fpls.2016.00250.

Wang, Y., Yang, Z., Kootstra, G., & Khan, H. A. (2023). The impact of variable illumination on vegetation indices and evaluation of illumination correction methods on chlorophyll content estimation using UAV imagery. *Plant Methods*, *19*(1). https://doi.org/10.1186/s13007-023-01028-8

Zhang, X., Yu, H., Yan, J., & Meng, X. (2025). Study on the Detection of Chlorophyll Content in Tomato Leaves Based on RGB Images. *Horticulturae*, *11*(6), 593. https://doi.org/10.3390/horticulturae11060593

Zhao, X., Guo, Q., Yan, C., & Zhu, J. (2018). Effects of light quality on the growth, leaf anatomy, and chloroplast ultrastructure of hydroponic plants. *Scientia Horticulturae*, *235*, 90–98. https://doi.org/10.1016/j.scienta.2018.02.057.