# Estimation of leaf chlorophyll index in corn using smartphone images and machine learning

**Charles Cardoso Santana[1], Smegnew Moges Mintesinot[2], Daniel Marçal de Queiroz[2] , Flávio Souza Santos[2], André Luiz de Freitas Coelho[2]**

[1] Minas Gerais Agricultural Research Agency – Pitangui Institute of Agricultural Technology, Pitangui, MG, 35650-000, Brazil

[2] Department of Agricultural Engineering, Federal University of Viçosa (UFV), Viçosa, MG, 36570-900, Brazil

**Corresponding author**: Charles Cardoso Santana (charles.santana@epamig.br)

**A paper from the Proceedings of the**
**17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture**
**13-15 July 2026**
**Porto Alegre, Brazil**

***Abstract.***

Accurate estimation of the Leaf Chlorophyll Index (LCI) in corn is fundamental for nitrogen management in precision agriculture, as nitrogen availability directly affects chlorophyll production and photosynthetic capacity. Conventional field assessment techniques are time-consuming and labor-intensive, while smartphone use provides a practical and low-cost alternative for obtaining high-resolution data in near-real-time. The objective of this study was to develop and validate a non-destructive image processing pipeline to estimate the Leaf Chlorophyll Index in corn using RGB images captured with a smartphone camera under field conditions. The preprocessing pipeline integrated illumination correction and color calibration using reference patches with known RGB values (Dark Green, Green, Light Green, Orange, White, and Black), combined with geometric alignment using white square detection and projective transformation. Corn leaves were segmented from background elements using a two-stage approach: preliminary mask generation using HSV thresholding with empirically determined hue ranges (0.03–0.36 and 0.39–0.683) and saturation values (0.10–0.45), followed by refinement with GrabCut using 500 SLIC superpixels for precise boundary delineation. From the segmented leaf images, eleven RGB-based vegetation indices were computed, including Excess Green (ExG), Normalized Green-Blue Difference Index (NGBDI), Dark Green Color Index (DGCI), and Color Index of Vegetation Extraction (CIVE), along with statistical descriptors (mean, skewness, and kurtosis) of the R, G, and B channels. Leaf chlorophyll index was also measured using a SPAD-502 chlorophyll meter. Machine learning models were developed in Python 3.8.10 to estimate the chlorophyll index based on smartphone-derived data. The following algorithms were tested: Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Support Vector Machine (SVM). On the validation dataset, XGBoost demonstrated superior performance (RMSE = 2.87, $R^2$ = 0.80), followed by Random Forest (RMSE = 3.08, $R^2$ = 0.77), while SVM exhibited notably lower accuracy (RMSE = 4.48, $R^2$

= 0.52). Performance remained robust on an independent test set, with XGBoost and Random Forest achieving RMSE values of 3.39 and 3.56, respectively. Feature importance analysis consistently identified NGBDI as the most important predictor across all models, with a gain-based importance of 236.99 in XGBoost, followed by ExGR, DGCI, CIVE, and ExG. Paired t-tests and Wilcoxon signed-rank tests with Bonferroni correction confirmed statistically significant differences among models (corrected $p < 0.001$), with ensemble-based algorithms significantly outperforming SVM. These results demonstrate that the developed pipeline provides a reliable and non-destructive approach for supporting nitrogen management in corn production systems, with potential for integration into real-time decision support tools and remote sensing workflows.

***Keywords.***
*Image processing, Leaf segmentation, RGB vegetation indices, Color calibration, Proximal remote sensing, Precision nitrogen management.*

## Introduction

Global food security remains one of the United Nations' key sustainable development objectives, as the world's population is expected to increase by more than 2 billion people, from 8.2 billion in 2024 to around 10.3 billion in 2084 (Lam, 2024). This rapid population growth drives the urgent need to increase global cereal production by 900 million tons a year to meet future food demand. Within these growing demands, corn (*Zea mays*) holds a critical position as the third-largest cereal crop worldwide, with annual production exceeding 1.1 billion tons, serving both human and animal nutritional needs (Sarvestani *et al.*, 2025).

However, traditional corn-growing methods often involve uniform applications of irrigation, fertilizers, and herbicides, which fail to take into account the geographical variability of agricultural fields(Li *et al.*, 2022). Low nitrogen use efficiency (NUE) is a recurring problem in this technique. Uniform fertilization is unable to cope with localized variations in nitrogen demand caused by soil variability, topographical variations, and uneven water distribution. In addition, the lack of temporal precision in these applications leads to both over- and under-fertilization. While nitrogen deficiency impairs photosynthesis, reduces productivity, and compromises overall plant health, excess nitrogen can lead to nutrient leaching, eutrophication, increased disease susceptibility, and higher production costs(Tang *et al.*, 2025).

Plant nitrogen status can be effectively determined by monitoring chlorophyll concentration, as nitrogen availability directly affects chlorophyll production. Absorption of solar radiation by leaves is mainly controlled by photosynthetic pigments, in particular chlorophyll a and b, which are important markers of productivity and photosynthetic capacity (Li *et al.*, 2010). Therefore, accurate estimation of chlorophyll content provides important information on crop health, nitrogen use efficiency, and stress responses.

Non-destructive techniques for estimating chlorophyll concentration have been developed and widely used. Imagery-based methods for real-time monitoring of chlorophyll and photosynthetic activity have been applied due to advances in remote sensing technology (Rigon *et al.*, 2016). Digital cameras, in particular, have proven to be useful tools for agricultural monitoring and analysis when combined with image segmentation and color pattern analysis.

However, traditional field assessment techniques are time-consuming, labor-intensive, and not feasible for large-scale farms. In contrast, remote sensing using RGB cameras provides near-real-time, high-resolution data that reveals crop health, stress factors, and geographic variability. The potential of this method to improve nitrogen management techniques has been highlighted by recent research showing that it is possible to determine nitrogen status in corn using a digital RGB camera. However, existing studies typically rely on fixed vegetation index formulations rather than integrating illumination correction, reference color chart calibration, leaf segmentation, and exhaustive feature selection within a unified image processing framework. To address this gap, the present study proposes an RGB image-based pipeline for estimating maize leaf chlorophyll content from field-acquired photographs, combining color calibration using reference color charts, leaf segmentation, and vegetation index extraction, with outputs correlated against SPAD meter readings and evaluated across multiple regression frameworks to support nitrogen fertilization recommendations.

## 2. Materials and methods

This study aims to develop a non-destructive image processing pipeline for estimating chlorophyll content in corn leaves using high-resolution RGB images captured under field conditions. The estimated chlorophyll values are validated using SPAD-502-meter readings, and the results will serve as a basis for nitrogen fertilization recommendations.

## 2.1 Development of preprocessing pipeline and color calibration

### 2.1.1 Image acquisition and SPAD measurement

High-resolution RGB images of corn leaves were captured under natural daylight using a digital camera. A custom-designed reference board was used for each image, which included four white square markers positioned in the corners for geometric correction and five circular color patches, as shown in Figure 2.1 with known RGB values for color calibration. All images were acquired at a fixed camera angle and distance, with efforts taken to minimize shadows and maintain consistent natural illumination across samples.

To obtain ground-truth chlorophyll content, SPAD values were measured using a SPAD-502 Plus chlorophyll meter. For each imaged leaf, three readings were taken at different points along the lamina and subsequently averaged to derive a representative SPAD value per leaf. The statistical summary of the measured SPAD values is presented in Table 1.

**Table 1.** Statistical summary of the measured leaf SPAD values of corn leaves using a SPAD-502Plusents obtained, yielding a single, representative SPAD value.

| Statistic | Min | Mean | Max | Std | CV (%) |
|---|---|---|---|---|---|
| Values | 26.600 | 49.565 | 64.600 | 6.4312 | 12.975 |

### 2.1.2 White square detection and geometric alignment

To ensure geometric consistency and correct perspective distortion in field-acquired corn leaf images, four white square reference markers were placed at the corners of a calibration board within the scene. The acquired RGB image was first converted to grayscale and normalized to enhance contrast. A binary mask was then generated by thresholding high-intensity regions (i.e., pixels with grayscale intensity > 0.9), isolating the white markers from the background.

Subsequently, morphological operations such as hole filling and small object removal were applied to refine the mask. Connected components were extracted and analyzed based on geometric properties, including area, extent, and solidity, to identify square-like regions. The four largest valid components were selected as candidate markers, and their centroids were computed to serve as corner points for geometric correction.

To maintain a consistent spatial mapping, the centroids were sorted into a fixed order: top-left, top-right, bottom-right, and bottom-left. A projective transformation matrix was computed using these sorted points via MATLAB's fitgeotrans function, mapping the distorted reference board to a canonical rectangular plane. This transformation was then applied to the entire RGB image using imwarp, yielding a geometrically corrected output. The resulting image was visually verified and stored for subsequent analysis steps, including color calibration and automated leaf segmentation.

### 2.1.3 Color calibration using reference circles and gamma correction

To account for lighting inconsistencies and camera sensor response, color calibration was performed using five circular reference patches placed on a standardized calibration board within each image. These patches included Dark Green, Green, Light Green, Orange, White, and Black, each with known true RGB values (Table 2).

**Table 2.** True values of the RGB band of calibration panels

| Values | Dark Green | Green | Light Green | Orange | White | Black |
|---|---|---|---|---|---|---|
| R | 57 | 68 | 164 | 195 | 219 | 58 |
| G | 111 | 134 | 170 | 90 | 227 | 60 |
| B | 72 | 73 | 68 | 70 | 226 | 61 |
| gamma | 0.8 | 0.6 | 0.8 | 1.5 | 2.1 | 0.8 |

(a) (b)

**Figure 1.** White square detection and geometric alignment (a) original image with reference color circles and (b) detected white square for geometric alignment and rectification.

The reference color patches were manually selected using user-guided circular region identification. For each patch, the center was interactively chosen, and a circular mask with a fixed radius was applied to isolate the patch area. The mean RGB values within each masked region were computed and compared to the known true values provided by a spectrally calibrated source.

To correct for systematic deviations in color response, a linear scaling factor $S_c$ was calculated for each color channel (R, G, B) as equations 1 and 2 (Finlayson *et al.*, 2015):

$$S_c = \frac{U_{ct}}{U_{cm}} \tag{1}$$

Where, $U_{ct}$ represents the true (target) RGB value for the color patch,
$U_{cm}$denotes the measured RGB value obtained from the image.

Each pixel channel value of the original image I(x,y)) was scaled using its respective factor to obtain the linearly corrected image ($I_c(x,y)$ ):

$$I_c(x,y) = S_c \times I(x,y) \tag{2}$$

The scaling was applied independently to each color channel. Resulting pixel values were clipped to the valid range between 0 and 255 to preserve display and processing integrity. This linear transformation assumes a direct proportional relationship between the observed and true color values, which is appropriate under relatively uniform natural lighting. It compensates for systematic deviations arising from lighting conditions, sensor sensitivity, or other environmental factors.

After linear calibration, a gamma correction step was applied to adjust for the nonlinear luminance response of digital imaging systems. Gamma values for each color patch were determined empirically based on the expected brightness curve and visual consistency across the dataset (see Table 2).
The gamma correction was implemented using the following transformation as equation 3:

$$I_{out}(x,y) = 255 \times \sqrt[\gamma]{\frac{I_{in}(x,y)}{255}} \tag{3}$$

Where: $I_{in(x,y)}$ is the input pixel intensity (typically in the range 0 to 255. $\gamma$: is the gamma value (e.g., 0.8, 1.5, etc.) and $I_{out}(x,y)$ is the gamma-corrected output intensity.

Each color patch was corrected independently using its assigned gamma value. This step enhanced visual uniformity and color fidelity across varying brightness levels, particularly in the green bands relevant to chlorophyll estimation.

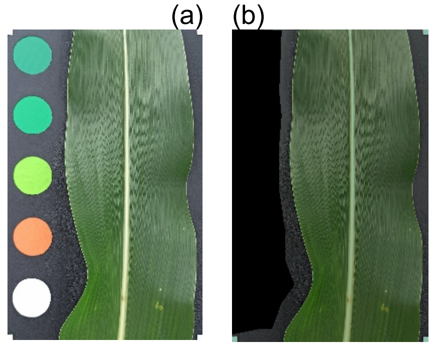


**Figure 2**. Color calibration and gamma correction (a) Original image showing reference color patches. (b) Image after color correction and gamma adjustment.

To prepare images for analysis, a polygon-based manual erase tool was provided to mask unwanted calibration color patches after image calibration. This interactive tool enabled precise exclusion of non-leaf areas and reference color patches. The fully color-corrected, gamma-adjusted, and cleaned images were then saved for subsequent processing, including vegetation index computation and SPAD-based chlorophyll estimation.

## 2.2 Corn leaf segmentation

To isolate the corn leaf region from the background and reference board, a two-stage segmentation approach was implemented, combining color-based thresholding and GrabCut refinement. This hybrid method leverages both spectral and spatial information to accurately delineate leaf boundaries in complex field images.

Stage 1: Preliminary mask generation using HSV thresholding

The input RGB image was first converted to the Hue, Saturation, and Value (HSV) color space to decouple chromaticity from intensity. The theoretical foundation of the HSI color model supports effective color separation, as described by Equations 4 to 7.

Normalized RGB values:

$$r = \frac{R}{R+G+B}, \qquad g = \frac{G}{R+G+B}, \qquad b = \frac{B}{R+G+B} \tag{4}$$

Hue (H), Saturation (S), and Intensity (I) are computed as:

$$H = \begin{cases} \theta & if\ B \leq G \\ 360° - \theta, & if\ B > G \end{cases} \tag{5}$$

$$\theta = \frac{\frac{1}{2}[(R-G)+(R-B)]}{(R-G)^2+(R-B)\,(G-B)^{1/2}} \tag{6}$$

$$S = 1 - \frac{3}{R+G+B}[min(R,G,B)] \tag{7}$$

A rough binary mask was then generated by thresholding the hue and saturation channels, which are particularly effective for detecting green vegetation. Pixels with hue values within the ranges (0.03 to 0.36) and (0.39 to 0.683), and saturation values between (0.10 to 0.45), were selected. These thresholds were empirically determined for each image based on individual leaf characteristics and illumination conditions. To improve mask quality, morphological operations were applied. Hole-filling (imfill) and small object removal (imopen) reduced noise and ensured connectivity within leaf regions. The result was a binary mask $M_o(x, y)$ where:

$$M_o(x,y) = \begin{cases} 1, & if \;\; Hmin \le H(x,y) \le H\,max \; and \; S(x,y) \ge Smin \\ 0, & otherwise \end{cases} \tag{8}$$

Where, H(x,y) and S(x,y) are the hue and saturation at pixel location (x,y), and Hmin, Hmax, Smin are the chosen thresholds for the minimum and maximum hue and minimum saturation channels, respectively.

Stage 2: GrabCut Refinement

To enhance segmentation precision, the initial mask was expanded via morphological dilation to define a region of interest (ROI) surrounding the leaf. This ROI served as the input for the GrabCut algorithm, which was used to refine the segmentation boundary through iterative modeling of foreground (leaf) and background color distributions.

Before GrabCut, the image was over-segmented into 500 superpixels using the SLIC (Simple Linear Iterative Clustering) algorithm to preserve local structure and improve boundary adherence. GrabCut then performed graph-cut optimization on these regions, producing a refined binary mask $M_{\text{GrabCut}}$(x,y), that labeled each pixel as either foreground (leaf) or background.

The final segmented image $I_{seg}$(x,y,c) was obtained by applying the binary mask to each color channel of the original RGB image as equation 9:

$$I_{seg}((x,y) = \begin{cases} I(x,y,c), & if \; M_{\text{GrabCut}}(x,y) = 1 \\ 0, & otherwise \end{cases} \tag{9}$$

This effectively removed all non-leaf pixels, producing an image that isolates the corn leaf for downstream processing, including vegetation index computation and chlorophyll content estimation.

(a) Original RGB image (b) Initial HSV Mask (c) Final segmented leaf

**Figure 3.** Corn leaf segmentation pipeline: (a) original RGB image with reference board, (b) rough binary mask using HSV thresholding, and (c) final segmented image after GrabCut refinement.

## 2.3. Vegetation index computation

To quantify leaf greenness and estimate chlorophyll content, several RGB-based vegetation indices were computed from the segmented corn leaf image. A binary mask was first created to isolate the plant region by excluding pixels where all RGB values were zero (i.e., black background). The red (R), green (G), and blue (B) channels were then extracted, and only the pixel values within the plant region were analyzed.

The segmentation mask classified each pixel as plant or non-plant based on its RGB values: pixels where all three channels were zero (i.e., black) were classified as background, and then leaf pixels retained their original RGB values.

Segmented image Plant region mask

**Figure 4.** Segmentation mask for RGB computation

To isolate the leaf region, the following mask is computed:

$$M(x,y) = \begin{cases} 1 \ \ if\ R(x,y) \neq 0\ and\ G(x,y) \neq 0\ and\ B(x,y) \neq 0 \\ 0\ \ if\ R(x,y) = G(x,y) = B(x,y) = 0 \end{cases}$$

Only pixels within the plant region (i.e., M(x,y)=1) were retained for further analysis. For each of the red (R), green (G), and blue (B) channels, statistical descriptors such as mean, skewness, and kurtosis were computed to characterize the color distribution and symmetry of the plant tissue.

Using the mean R, G, and B values of the plant region, several well-established vegetation indices were calculated. These indices were selected for their relevance to chlorophyll-related reflectance properties, particularly within the green spectral band. The mathematical formulations and corresponding references are summarized in Table 3.

**Table 3.** Formulas and references for RGB-based vegetation indices used to estimate leaf greenness and chlorophyll content

| Index | Full name | Formula | Citation | Equation |
|---|---|---|---|---|
| ExG | Excess Green | $2G - R - B$ | (Meyer & Neto, 2008) | 1 |
| ExR | Excess Red | $1.4R - G$ | (Meyer & Neto, 2008) | 2 |
| ExGR | Exces Green and Red Index | ExG -ExR | (Meyer & Neto, 2008) | 3 |
| GRVI | Green-Red Vegetation Index | $\frac{G - R}{G + R}$ | (Meyer & Neto, 2008) | 4 |
| NGBDI | Normalized Green-Blue Difference Index | $\frac{G - B}{G + B}$ | (Meyer & Neto, 2008) | 5 |
| NDI | Normalized Difference Index (NDI): | $\frac{G - R}{G + R + B}$ | (Tucker, 1979) | 6 |
| VEG | Vegetative Index | $\frac{G^2}{R \times B}$ | ( Li & Chen, 2011) | 7 |
| VARI | Visible Atmospherically Resistant Index | $\frac{G - R}{G + R - B}$ | (Haboudane et al., 2002) | 8 |
| RGRI | Red-Green Ratio Index | $\frac{R}{G}$ | (Yao et al., 2024) | 9 |
| CIVE | Color Index of Vegetation Extraction (CIVE): | $0.441 \times R - 0.881 \times G + 0.385 \times B + 18.7745$ | (Xie et al., 2018) | 10 |
| DGCI | Dark green color index | $\frac{\frac{(H - 60)}{60} + (1 - S) + (1 - V)}{3}$ | (Rigon et al., 2016) | 11 |

Where H, S, and V are the hue, saturation, and value components of the image converted to HSV color space.

All computed indices and statistical measures were exported to Excel for further analysis and correlation with SPAD chlorophyll meter readings.

### 2.4 Machine learning-based chlorophyll prediction using vegetation indices

Machine learning models were developed to predict corn chlorophyll concentration from vegetation indices (VIs) derived from RGB imagery. Analyses were performed in Python (v3.8.10). A comparative evaluation was conducted using Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Support Vector Machine (SVM) algorithms, chosen for their capability to model complex nonlinear and multivariate relationships. Models were trained and validated using the selected VIs as predictors and corresponding chlorophyll values as targets. Performance was assessed with root mean square error (RMSE) and $R^2$ on both training and validation sets. Predicted versus observed chlorophyll values were plotted to visualize accuracy and bias.

Feature importance was extracted from RF and XGBoost models to identify the most influential VIs, while permutation importance and SHAP values were used to interpret Support vector machine predictions when applicable. Statistical validation of model performance differences was conducted using paired hypothesis tests (paired t-test or Wilcoxon signed-rank test) on validation residuals. Multiple comparisons were corrected with the Bonferroni method to ensure robustness. Test statistics and p-values were reported to determine significant differences among models.

## 3.Result and discussion

### 3.1. Model performance comparison

Three machine learning algorithms, such as Random Forest (RF), XGBoost, and Support Vector Machine (SVM), were evaluated for their ability to estimate chlorophyll concentration using vegetation indices derived from corn leaf images. Model performance was assessed using the Root Mean Square Error (RMSE) and the coefficient of determination ($R^2$) to quantify prediction accuracy.

Among the models, XGBoost demonstrated the best performance on the validation dataset, achieving an RMSE of 2.87 and an $R^2$ of 0.80, indicating high predictive accuracy and a strong fit to the observed data. Random Forest followed closely with an RMSE of 3.08 and $R^2$ of 0.77, while SVM performed notably worse (RMSE = 4.48, $R^2$ = 0.52), suggesting limited ability to capture nonlinear relationships in the data. When applied to an independent test set, XGBoost and Random Forest maintained robust performance (RMSE = 3.39 and 3.56, respectively), whereas SVM again exhibited comparatively lower accuracy (RMSE = 4.48).

These results suggest that XGBoost is the most effective and reliable model for chlorophyll estimation using RGB-derived vegetation indices in this context. Its combination of low error and high explanatory power ($R^2$ = 0.80) demonstrates strong potential for integration into remote sensing workflows and real-time decision support tools in precision agriculture.

**Table 4.** Model performance metrics

| Model | RMSE(validation) | $R^2$(Validation) |
|---|---|---|
| XGBoost | 2.87 | 0.80 |
| Random Forest | 3.08 | 0.77 |
| SVM | 4.48 | 0.52 |

### 3.2 Feature importance analysis

Feature importance was analyzed to identify the most influential vegetation indices for each model. In the XGBoost model, NGBDI had the highest gain-based importance (236.99), followed by ExGR, DGCI, CIVE, and ExG. This suggests that narrow-band greenness indices provided more informative signals for estimating chlorophyll concentration.

The Random Forest model selected nine key features using impurity-based measures, with NGBDI, DGCI, and kurtB ranking highest. Similarly, the SVM model highlighted NGBDI, skewG, and kurtB as top predictors. Notably, NGBDI consistently ranked as the most important feature across all models, underscoring its relevance in chlorophyll estimation from UAV imagery.

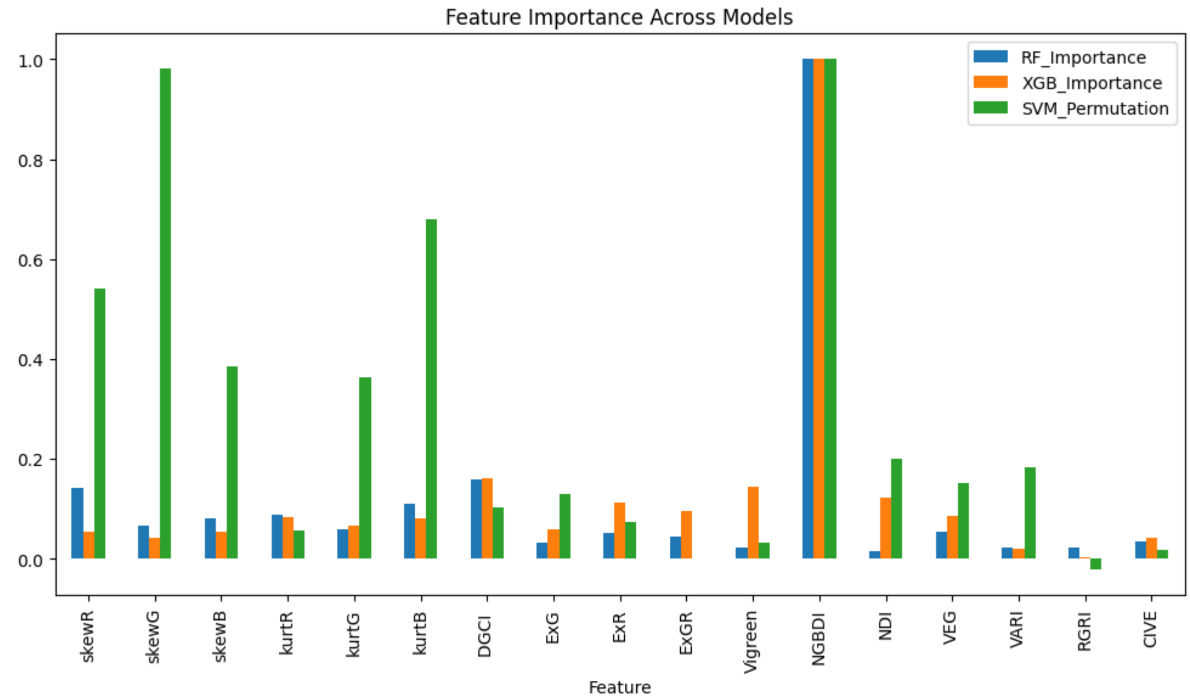


**Figure 5.** Feature importance across machine learning models. Normalized values highlight (from left to right) statistical descriptors (e.g., skewG, kurtB), greenness indices (e.g., ExGR, DGCI), and NGBDI dominating model predictions.

The feature importance analysis across all three machine learning models, such as XGBoost, Random Forest, and Support Vector Machine, consistently identified the Normalized Green-Blue Difference Index (NGBDI) as the most influential predictor for estimating chlorophyll concentration from digital RGB imagery. In the XGBoost model, NGBDI achieved the highest gain-based importance score (236.99), indicating it contributed most significantly to the model's predictive power. Other important indices in XGBoost included ExGR, DGCI, and CIVE, suggesting that indices emphasizing green reflectance carry strong predictive relevance.

Similarly, the Random Forest model ranked NGBDI as the top feature based on impurity reduction, followed by DGCI and kurtB, highlighting the importance of indices involving green and blue channels as well as statistical descriptors. The SVM model also prioritized NGBDI, alongside skewG and kurtB, underscoring the combined utility of spectral and texture-based features.

The consistently high ranking of NGBDI across all models suggests that greenness indices emphasizing the contrast between green and blue channels are particularly sensitive to chlorophyll concentration. This is likely due to the physiological behavior of chlorophyll pigments, which strongly absorb blue light while reflecting green. DGCI and kurtB also demonstrated substantial influence in the Random Forest model, in agreement with previous studies reporting strong correlations between DGCI and SPAD-based chlorophyll measurements (Rigon *et al.*, 2016). The inclusion of kurtosis and skewness measures (e.g., kurtB, skewG) indicates that the shape of pixel value distributions—beyond mean intensities—can offer meaningful insights into leaf color and condition.

Overall, the cross-model consensus on the importance of NGBDI reinforces its utility as a reliable and robust indicator for non-destructive chlorophyll estimation using RGB imagery. These findings highlight the potential of integrating both spectral indices and statistical features to improve prediction accuracy in precision agriculture applications.

### 3.3. Statistical comparison of model predictions

Figure 6 illustrates the predictive performance of three machine learning models of Random Forest, XGBoost, and Support Vector Machine (SVM) in estimating SPAD values. The scatter plot compares actual SPAD readings with the corresponding predicted values from each model. Most points align closely with the 1:1 dashed line, indicating a high level of agreement between true and predicted values. However, a noticeable spread exists, especially for SPAD values below 40, suggesting reduced accuracy in this range. Among the models, the XGBoost and SVM predictions exhibit slightly tighter clustering around the line, implying comparatively better consistency. Overall, the models show reliable predictive capacity, supporting their use for non-destructive chlorophyll estimation.

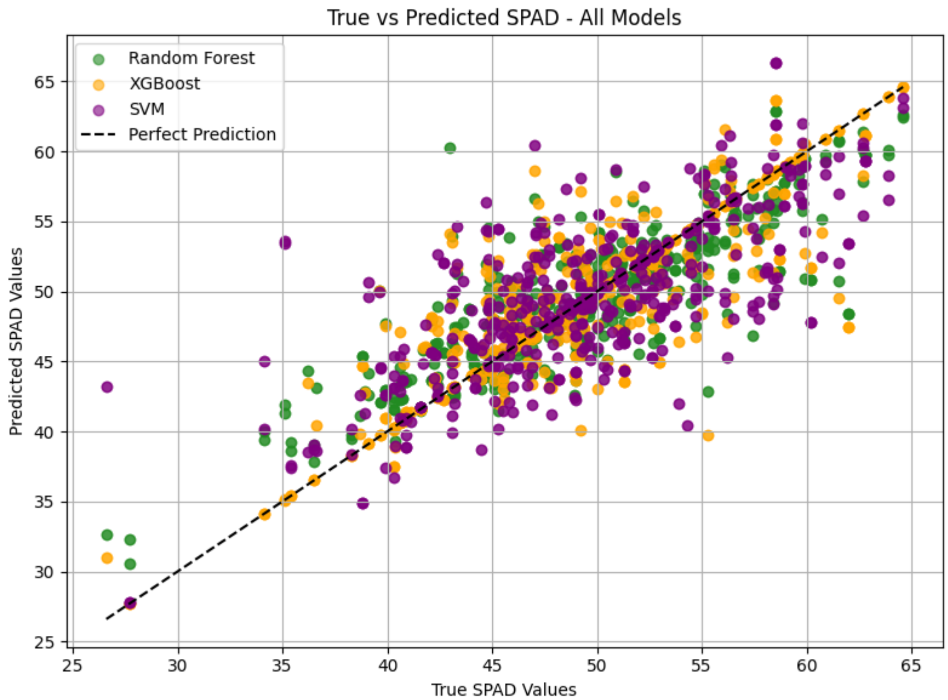


**Figure 4.3.** Scatter plot comparing true versus predicted SPAD values for all models: Random Forest (green), XGBoost (orange), and SVM (purple). The dashed black line represents the 1:1 line, indicating perfect prediction.

To determine whether the differences in model performance were statistically significant, both parametric (paired t-tests) and non-parametric (Wilcoxon signed-rank) tests were conducted on the absolute residuals. All pairwise comparisons yielded statistically significant differences after Bonferroni correction (corrected $p < 0.001$). Specifically, Random Forest significantly outperformed SVM (corrected $p = 0.0000$) and XGBoost (corrected $p = 0.0003$). XGBoost also significantly outperformed SVM (corrected $p = 0.0000$).

These results confirm that ensemble-based models (RF and XGBoost) provide more accurate and reliable predictions than kernel-based models such as SVM. This is likely due to their ability to capture complex, non-linear relationships and handle high-dimensional feature spaces more effectively.

### 3.4. Implications and model robustness

The consistent high performance of Random Forest and XGBoost across all metrics and statistical tests demonstrates the robustness of ensemble learning algorithms for UAV-based chlorophyll estimation. Their ability to model complex nonlinearities and feature interactions makes them well-suited for high-resolution vegetation index data analysis.

The dominance of NGBDI and other greenness indices across all models underscores their predictive strength, guiding future selection of spectral indices for sensor development and monitoring strategies. For example, NGBDI has shown strong correlations with chlorophyll and nitrogen content in various crops (e.g., −0.629 with chlorophyll and −0.611 with nitrogen in sugarcane). Studies in wheat also support NGBDI's robustness for non-destructive nitrogen monitoring.

## 4. Conclusion

Under this project, an image-processing pipeline was developed and validated for estimating chlorophyll content in corn leaves. The system integrates illumination correction, color calibration, and leaf segmentation, enabling precise extraction of vegetation indices. Machine learning models, particularly XGBoost and Random Forest, demonstrated strong predictive performance, with RMSE around 3 SPAD units and $R^2$ near 0.80 on both validation and test datasets. Feature importance analysis consistently identified the normalized greenness-based NGBDI index as the most informative predictor. Additionally, reducing the feature set in the Random Forest enhanced model generalization highlights the benefit of feature selection.

Statistical tests confirmed that ensemble models significantly outperform SVM in chlorophyll estimation accuracy. These findings suggest that the developed pipeline provides a reliable foundation for precision nitrogen management in corn cropping systems. Future work should focus on integrating these predictions into decision-support tools, calibrating chlorophyll estimates to agronomic thresholds, and conducting field trials to inform nitrogen application strategies directly.

## 5. Acknowledgements

The authors gratefully acknowledge the technical and institutional support of the Instituto Tecnológico de Agropecuária de Pitangui (EPAMIG ITAP) and the financial support of the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) in the development of this work (Project APQ-06075-24).

## 6. Reference

Finlayson, G. D., MacKiewicz, M., & Hurlbert, A. (2015). Color Correction Using Root-Polynomial Regression. *IEEE Transactions on Image Processing*, *24*(5), 1460–1470. https://doi.org/10.1109/TIP.2015.2405336

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, *81*(2–3), 416–426. https://doi.org/10.1016/S0034-4257(02)00018-4

Lam, D. (2024). The Next 2 Billion: Can the World Support 10 Billion People? *Population and Development Review*, *1*, 63–102. https://doi.org/10.1111/padr.12685

Li, A. J., & Chen, S. R. (2011). Stereoselective one-pot synthesis of (E)-3-(4-arylpiperazin-1-yl)-1- arylprop-2-en-1-ones. *Journal of Chemical Research*, *35*(1), 8–10. https://doi.org/10.3184/174751911x12964930076625

Li, Y., Chen, D., Walker, C. N., & Angus, J. F. (2010). Estimating the nitrogen status of crops using a digital camera. *Field Crops Research*, *118*(3), 221–227. https://doi.org/10.1016/j.fcr.2010.05.011

Li, Z., Guan, K., Zhou, W., Peng, B., Jin, Z., Tang, J., Grant, R. F., Nafziger, E. D., Margenot, A. J., Gentry, L. E., DeLucia, E. H., Yang, W. H., Cai, Y., Qin, Z., Archontoulis, S. V., Fernández, F. G., Yu, Z., Lee, D. K., & Yang, Y. (2022). Assessing the impacts of pre-growing-season weather conditions on soil nitrogen dynamics and corn productivity in the U.S. Midwest. *Field Crops Research*, *284*(October 2021). https://doi.org/10.1016/j.fcr.2022.108563

Meyer, G. E., & Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture*, *63*(2), 282–293. https://doi.org/10.1016/j.compag.2008.03.009

Rigon, J. P. G., Capuani, S., Fernandes, D. M., & Guimarães, T. M. (2016). A novel method for the estimation of soybean chlorophyll content using a smartphone and image analysis. *Photosynthetica*, *54*(4), 559–566. https://doi.org/10.1007/s11099-016-0214-x

Sarvestani, G. S., Edalat, M., Shirzadifar, A., & Pourghasemi, H. R. (2025). Early Season Dominant Weed Mapping in Maize Field Using Unmanned Aerial Vehicle (UAV) Imagery: Towards Developing Prescription Map. *Smart Agricultural Technology*, 100956. https://doi.org/10.1016/j.atech.2025.100956

Tang, Y., Li, F., Hu, Y., & Yu, K. (2025). NTRI: A novel spectral index for developing a precise nitrogen diagnosis model across pre- and post-anthesis stages of maize plants. *Field Crops Research*, *325*(January), 109829. https://doi.org/10.1016/j.fcr.2025.109829

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Xie, Q., Dash, J., Huang, W., Peng, D., Qin, Q., Mortimer, H., Casa, R., Pignatti, S., Laneve, G., Pascucci, S., Dong, Y., & Ye, H. (2018). Vegetation Indices Combining the Red and Red-Edge Spectral Information for Leaf Area Index Retrieval. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(5), 1482–1492. https://doi.org/10.1109/JSTARS.2018.2813281

Yao, Y., Yue, J., Liu, Y., Yang, H., Feng, H., Shen, J., Hu, J., & Liu, Q. (2024). Classification of Maize

Growth Stages Based on Phenotypic Traits and UAV Remote Sensing. *Agriculture (Switzerland)*, *14*(7). https://doi.org/10.3390/agriculture14071175