# Classification of maize hybrids using UAV-based multispectral remote sensing and machine learning algorithms

## Clasificación de híbridos de maíz utilizando detección remota multiespectral basada en UAV y algoritmos de aprendizaje automático

**João Lucas Gouveia de Oliveira[1], Dthenifer Cordeiro Santana[1], Izabela Cristina de Oliveira[1], Ricardo Gava[1], Fábio Henrique Rojo Baio[1], Carlos Antônio da Silva Junior[2], Larissa Pereira Ribeiro Teodoro[1], Paulo Eduardo Teodoro[1], and Job Teixeira de Oliveira[1]***

**ABSTRACT**

Novel methodologies for phenotypic evaluation in maize have been developed through the integration of advanced sensing technologies and machine learning algorithms. The aim of this study was to identify the most accurate machine learning algorithm for the classification of maize hybrids and to determine the optimal input data to enhance model performance. Seven maize hybrids were used in the experiment. After 60 d of crop emergence, the remotely piloted aircraft SenseFly® eBee RTK was used to obtain reflectance values at the following spectral bands (SB): blue (475 nm, B_475), green (550 nm, G_550), red (660 nm, R_660), red edge (735 nm, RE_735) and near-infrared (790 nm, NIR_790). Following the acquisition of spectral band (SB) data, vegetation indices (VIs) were calculated. The resulting dataset was subsequently analyzed using machine learning techniques, evaluating six algorithms: artificial neural networks (ANN), J48 decision trees (J48), REPTree (DT), random forest (RF), support vector machine (SVM) and logistic regression (LR) as the baseline model. Three accuracy metrics were used to evaluate the performance of the algorithms in classifying maize hybrids: correct classifications (CC), Kappa coefficient, and F-score. Among the algorithms tested, ANN showed the highest performance in all three metrics, proving its superiority and potential for real-world applications. Although all three input configurations enhanced classification accuracy for ANN algorithm, the optimal approach is to use only SB as input due to reduced data processing time and increased simplicity.

**Key words:** artificial neural networks, spectral bands, spectral curve, vegetation indices, machine learning classification, UAV-based image analysis.

**RESUMEN**

Se han desarrollado nuevas metodologías para la evaluación fenotípica en maíz mediante la integración de tecnologías de detección avanzadas y algoritmos de aprendizaje automatizado. El objetivo de este trabajo fue identificar el algoritmo de aprendizaje automático más preciso para la clasificación de híbridos de maíz y determinar los datos de entrada que mejoran el rendimiento del modelo. Se utilizaron siete híbridos de maíz. A los 60 d de emergencia del cultivo se utilizó la aeronave pilotada remotamente SenseFly® eBee RTK para obtener la reflectancia en las siguientes bandas espectrales (BE): azul (475 nm, B_475), verde (550 nm, G_550), rojo (660 nm, R_660), borde rojo (735 nm, RE_735) y NIR (790 nm, NIR_790). Luego de la adquisición de datos de la banda espectral (BE), se calcularon los índices de vegetación (Vis). El conjunto de datos resultante se analizó posteriormente utilizando técnicas de aprendizaje automático, evaluando seis algoritmos: redes neuronales artificiales (RNA), árboles de decisión J48 (J48), REPTree (DT), bosque aleatorio (BA), máquina de vectores de soporte (MVS) y regresión logística (RL) como enfoque de referencia. Se utilizaron tres métricas de precisión para evaluar el desempeño de los algoritmos en la clasificación de híbridos de maíz: clasificaciones correctas (CC), coeficiente Kappa y F-score. Entre los algoritmos probados, el algoritmo de ANN se destacó con el mayor desempeño en las tres métricas, demostrando su superioridad y potencial para aplicaciones reales en clasificación. Aunque las tres configuraciones de entrada mejoraron la precisión de clasificación para el algoritmo RNA, el enfoque óptimo es utilizar solo las BE como entrada debido al menor tiempo de procesamiento de datos y la mayor simplicidad.

**Palabras clave:** redes neuronales artificiales, bandas espectrales, curva espectral, índices de vegetación, clasificación mediante aprendizaje automático, análisis de imágenes basado en UAV.

Received for publication: February 10, 2025. Accepted for publication: April 25, 2025. Doi: 10.15446/agron.colomb.v43n1.118781

[1] Universidade Federal de Mato Grosso do Sul, Chapadão do Sul, MS (Brazil).
[2] Universidade Estadual do Mato Grosso (Brazil).
* Corresponding author: job.oliveira@hotmail.com

## Introduction

Maize (*Zea mays* L.) is a highly versatile crop with numerous applications, representing a substantial portion of cultivation in major grain-producing countries (Liu *et al*., 2023). This versatility results from the extensive selection and improvement processes that the plant has undergone. These processes have enabled maize to overcome global challenges, improve productivity and nutritional quality, and meet diverse human needs (Swarup *et al*., 2021).

Over the years, plant breeding techniques have advanced, especially at the molecular level, such as genetic sequencing of several species and gene interventions, enabling the development of populations resistant to biotic and abiotic stresses (Iqbal *et al*., 2021; Rivero *et al*., 2022; Zafar *et al*., 2022). However, phenotypic analysis has not progressed at the same rate, facing challenges such as a shortage of qualified personnel capable of efficiently and accurately selecting desirable characteristics. Therefore, it is essential to develop and implement advanced techniques to enhance the precision and performance of phenotypic evaluations (Santos *et al*., 2014).

The advancement in technologies such as remote sensing, combined with efforts to accelerate maize genetic improvement, has led to the development of new high-precision phenotyping (FAP) techniques. These techniques address the extended duration required by traditional methods for the evaluation and selection of materials (Herzig *et al*., 2021). FAP provides significant advantages in reducing fieldwork time, labor and costs by using sensors to characterize phenotypes with high accuracy and efficiency in a non-destructive manner (Andrade *et al*., 2021; Dobbels *et al*., 2019; Santana *et al*., 2023). FAP also enables the correlation of several plant traits, such as the selection of soybean genotypes based on precocity and grain productivity (Santana *et al*., 2022).

The emergence of remotely piloted aircraft (unmanned aerial vehicle - UAV) was significantly benefited agriculture by integrating various sensors to capture data across different wavelengths. This technology has become an important tool for non-destructive field data collection throughout the crop cycle (Das Choudhury *et al*., 2019). It is capable of collecting data with high speed and spatial resolution (Kar *et al*., 2021).

The amount of data generated through remote sensing is substantial, posing challenges for traditional statistical methods in correlating sensor data with plant characteristics, due to the non-linearity between spectral variables and physiological or morphological characteristics of plants (Van Eeuwijk *et al*., 2019). Thus, there is a need for enhanced data processing capacity, which can be achieved through the application of machine learning (ML) algorithms.

Machine learning (ML) algorithms enable the prediction of leaf nitrogen concentration and height in maize (Osco *et al*., 2020). Additionally, ML algorithms can be used for classifying different soybean cultivars (Gava *et al*., 2022), evaluating leaf nutritional traits (Santana, Teodoro *et al*., 2023), and accurately classifying genotypes based on industrial soybean characteristics (Santana, Teixeira *et al*., 2023). Corn plants subjected to different irrigation management practices exhibit distinct spectral behaviors, allowing their classification through machine learning modeling (Oliveira *et al*., 2025).

This study hypothesizes that maize hybrids can be effectively distinguished based on spectral data using machine learning models. The aim was to identify the most accurate machine learning algorithm for classifying maize hybrids and to determine the optimal input data to enhance the performance of these models.

## Materials and methods

### Experiment conditions

The experiment was conducted in the experimental area of the Federal University of Mato Grosso do Sul (18°41'33" S, 52°40'45" W, 810 m a.s.l.), located in Chapadão do Sul, Brazil. The conventional soil preparation method, consisting of plowing and leveling harrowing, was employed. The climate of the region is classified as Tropical Savanna (Aw) according to the Köppen classification. The soil, classified as clayey Dystrophic Red Oxisol (Santos *et al*., 2018), exhibits the following characteristics in the 0-0.20 m layer: $pH_{(H_2O)}$ = 6.2; exchangeable Al (cmol $kg^{-1}$) = 0.0; Ca+Mg (cmol $kg^{-1}$) = 3.59; P (mg $kg^{-1}$) = 41.3; K (cmol $kg^{-1}$) = 0.16; organic matter (%) = 1.97; base saturation (BS, %) = 45; aluminum saturation (%) = 0.0; sum of bases (cmol $kg^{-1}$) = 1.92; cation exchange capacity (CEC) (cmol $kg^{-1}$) = 4.25.

Seed sowing was carried out using a seeder, with seeds placed 0.45 m apart between rows. Additionally, 200 kg $ha^{-1}$ of monoammonium phosphate (MAP) with a formulation of 11-52-00 was applied. Top dressing was performed using 150 kg $ha^{-1}$ of urea (45% N) when the plants reached

the V4 growth stage. The experiment was conducted using a strip-plot design, with plot dimensions of 3.8 m x 4.9 m (18.62 $m^2$). The maize hybrids used in the experiment were: H1 (AS 1868), H2 (DKB 360), H3 (FS 615 PWU), H4 (K 7510 VIP3), H5 (NK 520 VIP3), H6 (P 3858 PWU), and H7 (SS 182E VIP3). Each hybrid's seed amount per meter was 2.67, 2.68, 2.74, 2.65, 2.64, 2.63, and 2.55, respectively.

### Acquisition and processing of multispectral images

A flight was conducted 60 d after crop emergence, at the stage of full bloom for the maize varieties, using the SenseFly® eBee RTK, a fixed-wing remotely piloted aircraft with autonomous takeoff and landing capabilities. The eBee was equipped with a Parrot® Sequoia multispectral sensor (Parrot Drones SAS). The images were obtained at 09:00 a.m. under clear skies, at an altitude of 100 m and a spatial resolution of 0.10 m. The aerial survey employed Real-Time Kinematic (RTK) technology, which enables image acquisition with a positional accuracy of 2.5 cm. Image mosaicking and orthorectification were subsequently performed using Pix4Dmapper software.

Radiometric calibration of the entire scene was conducted using a calibrated reflective surface provided by the manufacturer. The Parrot Sequoia multispectral sensor is equipped with a sunshine sensor for the calibration of acquired reflectance values. The sensor captured data across the following spectral bands (SB): blue (475 nm, B_475), green (550 nm, G_550), red (660 nm, R_660), red edge (735 nm, RE_735) and near-infrared (790 nm, NIR_790). Following the acquisition of spectral band (SB) data, 41 vegetation indices (VIs) (Suppl. Tab. 1 (S1)) were calculated, ranging from simple ratios to indices incorporating atmosphere and soil correction (Oliveira *et al*., 2025; Ramos *et al*., 2020), using ESRI ArcGIS 10.5 geographic information systems software.

### Machine learning models and statistical analysis

The data were subjected to machine learning analyses, evaluating six algorithms: artificial neural networks (ANN) (Egmont-Petersen *et al*., 2002), J48 decision tree (J48) (Quinlan, 1993) and REPTree (Al Snousy *et al*., 2011), random forest (RF) (Belgiu *et al*., 2016), support vector machine (SVM) (Nalepa *et al*., 2019), and logistic regression (LR) (Štepanovský *et al*., 2017). Logistic regression (LR) was chosen as the baseline model due to its widespread adoption in binary and multiclass classification problems. It is frequently used as a reference in the literature because of its interpretability, computational efficiency, and consistent performance on linearly separable data. Its comparison with traditional methods (such as discriminant analysis or Naive Bayes) was implicitly addressed through stratified k-fold cross-validation (k=10 with 10 replicates), ensuring a robust and generalizable evaluation. Although LR does not always outperform more complex approaches, its use as a baseline allows for quantifying the actual improvement of the proposed hybrid models, distinguishing significant gains from marginal ones. Additionally, its simplicity facilitates the identification of biases or underfitting, making it a valid substitute for traditional methods in scenarios where variable relationships are predominantly linear. All algorithm parameters were established according to the default configuration of Weka 3.8.5, which proved efficient for the problem under study, with its impact on model performance validated through comparative tests.

Three accuracy metrics were used to evaluate the performance of the algorithms in classifying maize hybrids: correct classification (CC), Kappa coefficient and F-score. An analysis of variance (ANOVA) was conducted to assess the significance of ML algorithms, input variables and their interactions. To illustrate model performance and statistical significance, boxplots were generated using the mean values of CC, Kappa and F-score, grouped according to the Scott-Knott test (Scott & Knott, 1974) at a 5% significance level. All analyses and graphical representations were performed using the ggplot2 and ExpDes.pt packages in R (R Core Team, 2013).

## Results and discussion

The reflectance of each hybrid at specific wavelengths is shown in Figure 1. Hybrids H4 and H7 exhibited the highest reflectance values across all wavelengths, with statistically significant similarities observed only with hybrid H6 at the B_475 wavelength and with hybrid H1 at the RE_735 and NIR_790 wavelengths. Hybrids H3 and H5 exhibited the lowest reflectance values at the B_475, R_660 and G_550 wavelengths. At the RE_735 wavelength, hybrids H5 and H6 showed reduced reflectance, while at the NIR_790 wavelength, hybrid H6 demonstrated the lowest reflectance.

Performance evaluation of the machine learning algorithms was conducted using three parameters: correct classification (CC), F-score and Kappa coefficient. A significant interaction was observed between the input data and the machine learning algorithms for all accuracy parameters assessed.

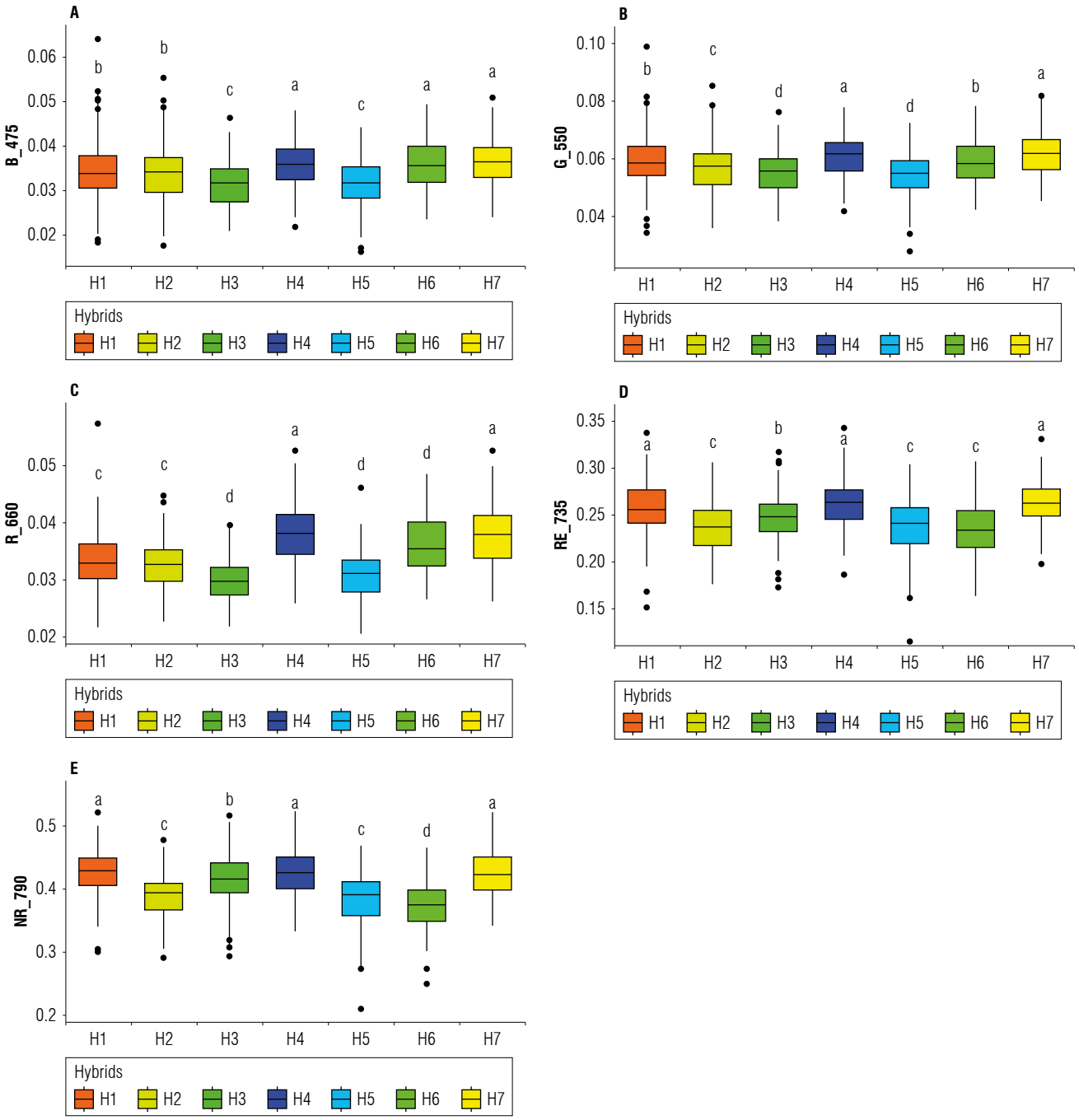


**FIGURE 1.** Boxplots illustrating the multispectral reflectance of seven maize hybrids across five spectral bands: blue (B_475 nm, A), green (G_550 nm, B), red (R_660 nm, C), red edge (RE_735 nm, D) and near-infrared (NIR_790 nm, E). The data were analyzed using the Scott-Knott test at a 5% significance level. Hybrids followed by the same letters for each wavelength do not significantly differ from each other.

Based on the correct classification metric, the artificial neural network (ANN) was identified as the best-performing algorithm among all input datasets (Fig. 2). When analyzing the ANN algorithm, no statistically significant differences were observed among the input datasets. In contrast, the RF, LR, DT, and SVM algorithms demonstrated superior performance with the IVs and SB+IVs inputs. For the J48 algorithm, the best performance was achieved with the SB and IVs inputs.

Evaluation of the algorithms using F-score metric revealed that ANN was the most effective algorithm for classification among all input datasets (Fig. 3). For the ANN, LR, J48, and DT algorithms, no statistically significant differences

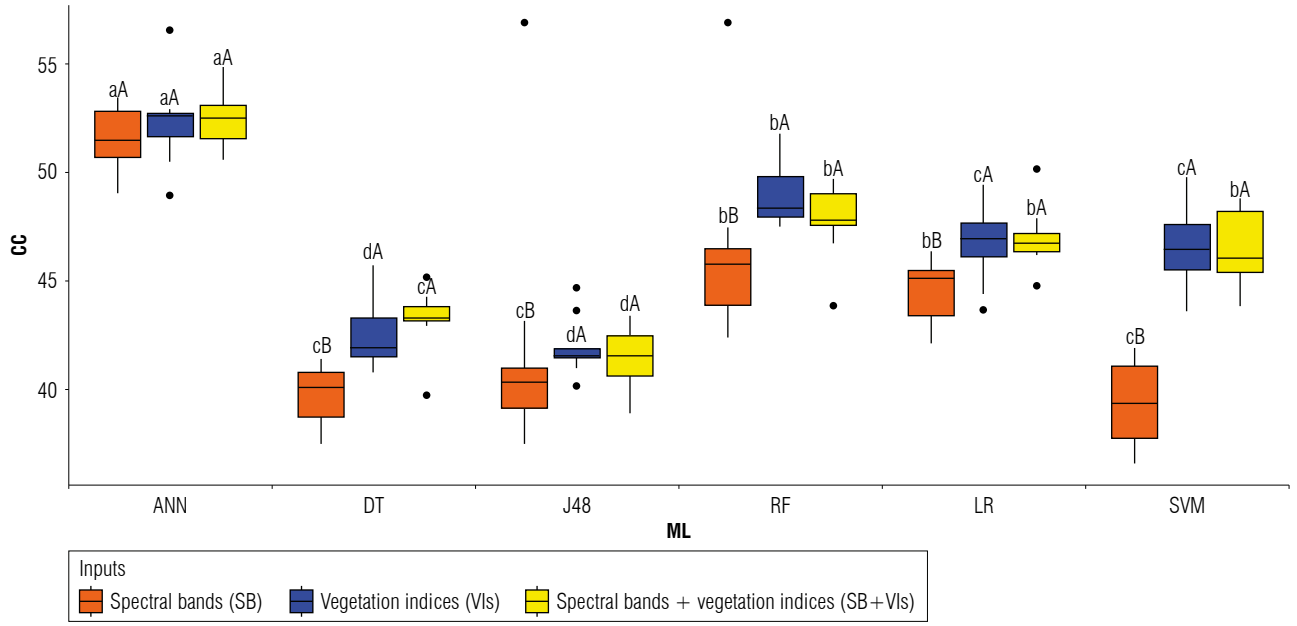


**FIGURE 2.** Boxplot illustrating the percentage of correct classification (CC) for the significant interaction between the machine learning models and the tested input datasets. Means followed by the same uppercase letters (for the different inputs) and the same lowercase letters (for the different ML algorithms) do not differ significantly according to the Scott-Knott test at a 5% significance level. Algorithms: Artificial neural networks (ANN), J48 decision trees (J48), REPTree (DT), random forest (RF), support vector machine (SVM), and logistic regression (LR).

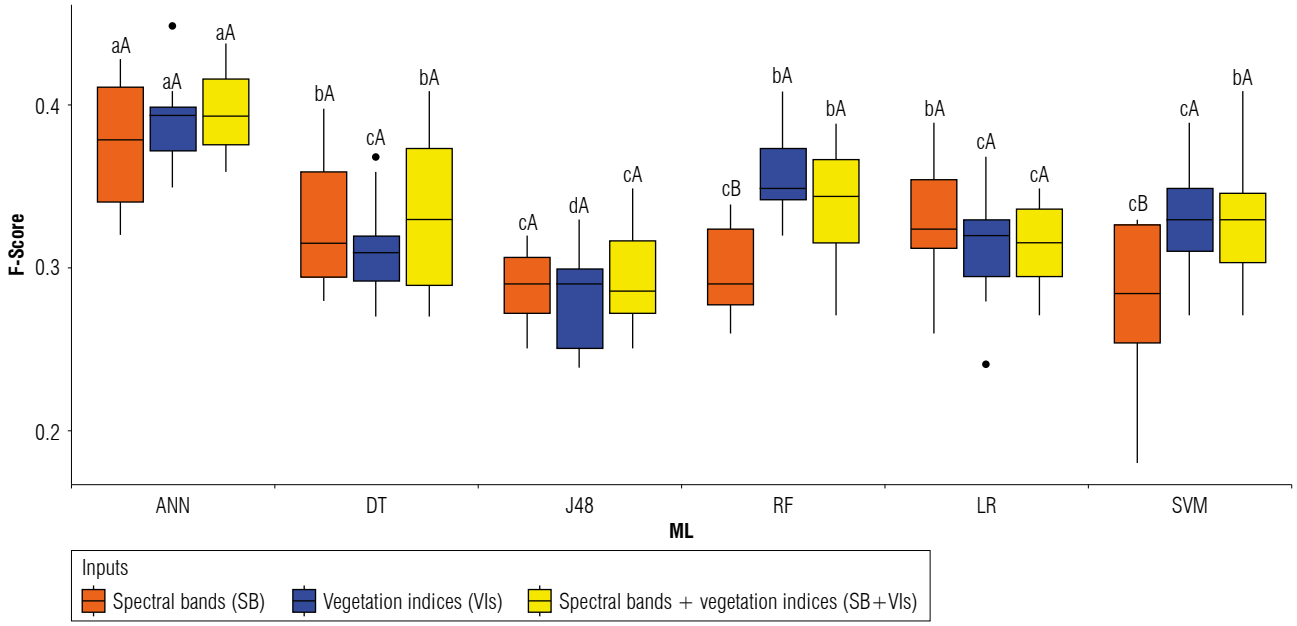


**FIGURE 3.** Boxplot illustrating the percentage of F-score metric for the significant interaction between the machine learning models and the tested input datasets. Means followed by the same uppercase letters (for the different inputs) and the same lowercase letters (for the different ML algorithms) do not differ significantly according to the Scott-Knott test at a 5% significance level.

were observed among the different input datasets. Conversely, the RF and SVM algorithms exhibited optimal performance with IVs and SB+IVs inputs.

Analysis using the Kappa metric indicated that the artificial neural network (ANN) was the most effective algorithm for classifying all input datasets (Fig. 4). For both the ANN and J48 algorithms, no statistically significant differences

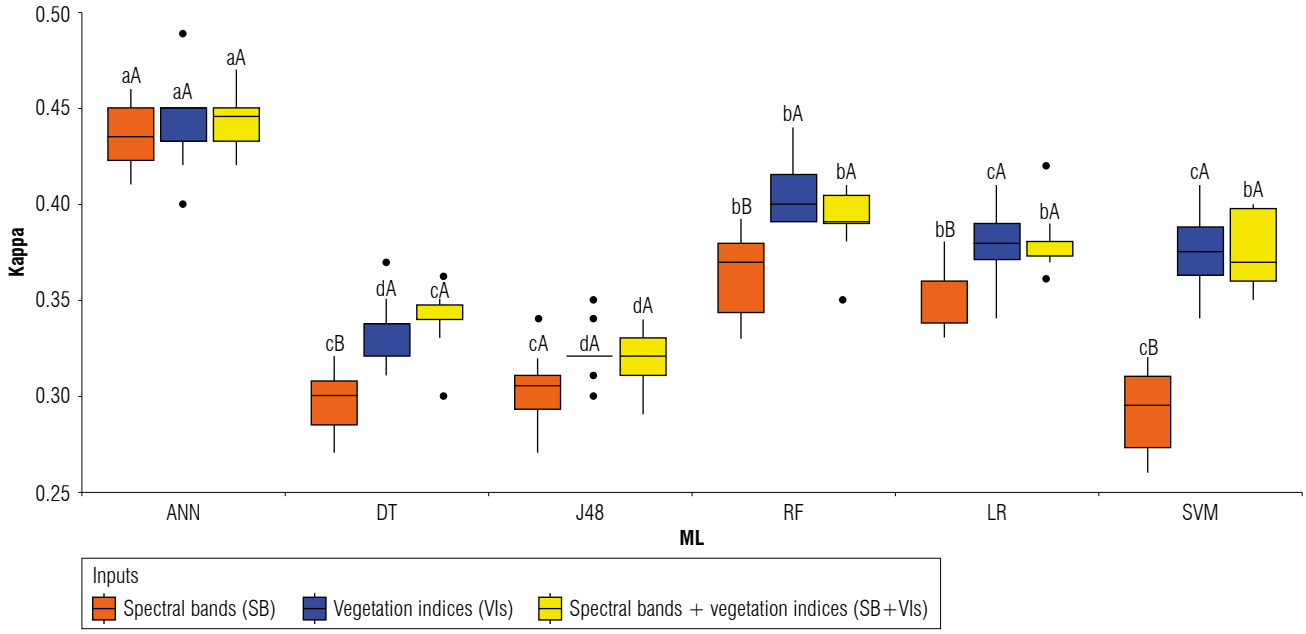


**FIGURE 4.** Boxplot illustrating the percentage of Kappa metric for the significant interaction between the machine learning models and the tested input datasets. Means followed by the same uppercase letters (for the different inputs) and the same lowercase letters (for the different ML algorithms) do not differ significantly according to the Scott-Knott test at a 5% significance level.

were observed among the input datasets. In contrast, the RF, LR, DT, and SVM algorithms demonstrated optimal performance with VI and SB+VI inputs.

The spectral signature of plants generally exhibits low reflectance in the visible region of the electromagnetic spectrum due to the absorption of these wavelengths by pigments for photosynthesis. In contrast, reflectance increases in the near-infrared and red-edge regions, above 700 nm (Taiz *et al*., 2017). A study mapping corn crops using Google Earth Engine (GEE), with calculations of vegetation indices, classification by the random forest (RF) algorithm, followed by analysis of the confusion matrix, kappa, and validation statistics, found satisfactory results; kappa values of 81.26% and 84.61% were reported among the image systems considered (Maciel Junior *et al*., 2024).

The visible spectrum is divided into spectral bands with distinct roles in plant physiology. For instance, blue (475 nm) and red (660 nm) wavelengths are critical for photosynthesis, significantly enhancing plant growth and development (Lu *et al*., 2021; Shi *et al*., 2019). The absorption of blue light positively regulates the expression of certain photosynthesis-related genes and increases the rate of electron transport, whereas the absorption of red light influences the expression of other genes involved in photosynthesis (Li *et al*., 2017; Wu *et al*., 2014).

The spectral behavior of hybrids varies according to their genetic background. In this context, hybrids H3 and H5 exhibited lower reflectance values across the observed range, suggesting greater use of wavelengths for photosynthesis, due to reduced reflectance and increased absorption. Notably, within the blue region of the spectral signature at 475 nm, minimal differentiation was observed among the hybrids. In contrast, the 550 nm wavelength, which corresponds to the green region of the spectrum, shows a pronounced peak in reflectance due to the strong reflection of green light by chloroplasts (Nishio, 2000). The H4 hybrid showed the highest reflectance in the 550 nm range compared to the other hybrids. From 735-790 nm onwards, a prominent increase in the reflectance within the VIS/NIR region is observed, attributed to the abundance of chlorophyll and organization of the internal leaf structure (Hennessy *et al*., 2020). This increase can serve as an indicator of stress, senescence, or disease incidence in vegetation (Atta *et al*., 2023; Dawson *et al*., 1998; Gholizadeh *et al*., 2016; Zahir *et al*., 2022). Thus, hybrids H1 and H3 exhibited high reflectance at 735 nm and lower reflectance at 475 nm and 660 nm. This result suggests that these hybrids have a high concentration of chlorophyll, as indicated by their increased reflectance at 735 nm. Additionally, the low reflectance in blue and red wavelengths implies intense photosynthetic activity in these hybrids. Based on the detailed physiological information provided by the spectral bands,

it is possible to investigate relationships between different spectral bands (Oliveira *et al*., 2023) through mathematical models known as vegetation indices (VIs) (Pantaleão *et al*., 2022; Silva *et al*., 2020). The use of these VIs can provide even more accurate information about plant physiological characteristics, including agronomic responses such as productivity (Santana, Santos *et al*., 2022), leaf nitrogen concentration, and maize plant height (Osco *et al*., 2020).

Distinguishing maize hybrids based on their spectral characteristics can reduce the costs of the genotype selection process. However, the large volume of data generated can hinder information processing. Consequently, the use of machine learning algorithms to classify plants or leaves (Osco *et al*., 2020) is a fast and accurate alternative for genotype differentiation. ML involves training algorithms on specific datasets to recognize patterns, which can then be applied to plant genetic improvement studies (Niazian & Niedbała, 2020).

Among the machine learning (ML) algorithms used, artificial neural networks (ANNs) are among the most well-known and widely employed techniques, particularly for classification tasks (Vidyarthi *et al*., 2020). A key advantage of neural networks over other ML algorithms is their ability to efficiently model complex nonlinear relationships, without requiring prior assumptions (Qi & Zhang, 2001).

Among the three accuracy metrics evaluated, the ANN algorithm achieved the highest performance in classifying maize hybrids, regardless of the input data used. Consequently, the use of SB can simplify and accelerate procedural steps, as it does not require the complex mathematical model calculations needed for IVs (Gava *et al*., 2022). The use of only specific wavelengths enhances algorithm accuracy in various agricultural analyses, such as predicting days to maturity, plant height and yield in soybeans (Teodoro *et al*., 2021). Identifying the most effective machine learning algorithms and input variables improves the performance of models for accurate classification (Oliveira *et al*., 2025). Our findings indicate that maize hybrids can be effectively distinguished using multispectral data, revealing distinct spectral behaviors and unique spectral signatures among different hybrids. Furthermore, the use of machine learning algorithms can be an important tool for classifying these hybrids. It is important to highlight the novelty of this research in maize cultivation. While similar techniques have been applied to other crops, such as soybeans, there is limited literature addressing genotype classification, both commercial and in the improvement process, in maize cultivation. The use of hyperspectral sensors represents a promising approach for obtaining more comprehensive insights into maize hybrids and is suggested as a potential direction for future research.

## Conclusions

The Artificial Neural Networks (ANN) algorithm demonstrated the highest performance among the three accuracy metrics evaluated.

Although all three input configurations improved classification accuracy of the ANN algorithm, the most effective approach was using only the spectral bands, as this strategy reduces data processing time and simplifies the overall procedure.

Despite the promising results, this study has some limitations. First, the analysis was restricted to a specific geographic region and a limited number of maize hybrids, which may affect the generalizability of the model to other environments or crop varieties. Second, the use of only multispectral data (without additional features such as thermal or LiDAR data) might not capture all relevant phenotypic variations among hybrids. Third, the performance of the ANN algorithm, while superior in this study, may vary with larger datasets or more complex crop traits.

### Acknowledgments

The authors would like to thank the Federal University of Mato Grosso do Sul (UFMS) and Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul – FUNDECT.

### Conflict of interest statement

The authors declare that there is no conflict of interests regarding the publication of this article.

### Author's contributions

JLGO: conceptualization, methodology, validation, formal analysis, research, writing - original draft; DCS, JTO: experiment design, data analysis, manuscript writing and editing; ICO: validation, writing - review and editing; RG: experiment design, data analysis, manuscript writing and editing; FHRB: validation, resources, writing - review and editing; CASJ, LPRT, PET: validation, writing - review and editing. All authors approved the final version of the manuscript.

## Literature cited

Al Snousy, M. B., El-Deeb, H. M., Badran, K., & Al Khlil, I. A. (2011). Suite of decision tree-based classification algorithms on cancer gene expression data. *Egyptian Informatics Journal*,

*12*(1), 73–82. https://www.sciencedirect.com/science/article/pii/S1110866511000223

Andrade, S. M., Teodoro, L. P. R., Baio, F. H. R., Campos, C. N. S., Roque, C. G., Silva Junior, C. A., Coradi, P. C., & Teodoro, P. E. (2021). High-throughput phenotyping of soybean genotypes under base saturation stress conditions. *Journal of Agronomy and Crop Science*, *207*(5), 814–822. https://doi.org/10.1111/jac.12513

Atta, B. M., Saleem, M., Bilal, M., ul Rehman, A., & Fayyaz, M. (2023). Early detection of stripe rust infection in wheat using light-induced fluorescence spectroscopy. *Photochemical & Photobiological Sciences*, *22*, 115–134. https://doi.org/10.1007/s43630-022-00303-2

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing, 114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Das Choudhury, S., Samal, A., & Awada, T. (2019). Leveraging image analysis for high-throughput plant phenotyping. *Frontiers in Plant Science, 10*, Article 508. https://doi.org/10.3389/fpls.2019.00508

Dawson, T. P., & Curran, P. J. (1998). A new technique for interpolating the reflectance red edge position. *International Journal of Remote Sensing, 19*(11), 2133–2139. https://doi.org/10.1080/014311698214910

Dobbels, A. A., & Lorenz, A. J. (2019). Soybean iron deficiency chlorosis high-throughput phenotyping using an unmanned aircraft system. *Plant Methods, 15*, Article 97. https://doi.org/10.1186/s13007-019-0478-9

Egmont-Petersen, M., de Ridder, D., & Handels, H. (2002). Image processing with neural networks – A review. *Pattern Recognition, 35*(10), 2279–2301. https://doi.org/10.1016/S0031-3203(01)00178-9

Gava, R., Santana, D. C., Cotrim, M. F., Rossi, F. S., Teodoro, L. P. R., Silva Junior, C. A., & Teodoro, P. E. (2022). Soybean cultivars identification using remotely sensed image and machine learning models. *Sustainability, 14*(12), Article 7125. https://doi.org/10.3390/su14127125

Gholizadeh, A., Mišurec, J., Kopačková, V., Mielke, C., & Rogass, C. (2016). Assessment of red-edge position extraction techniques: A case study for Norway spruce forests using Hymap and simulated Sentinel-2 data. *Forests, 7*(10), Article 226. https://doi.org/10.3390/f7100226

Hennessy, A., Clarke, K., & Lewis, M. (2020). Hyperspectral classification of plants: A review of waveband selection generalisability. *Remote Sensing, 12*(1), Article 113. https://doi.org/10.3390/rs12010113

Herzig, P., Borrmann, P., Knauer, U., Klück, H. C., Kilias, D., Seiffert, U., Pillen, K., & Maurer, A. (2021). Evaluation of RGB and multispectral unmanned aerial vehicle (UAV) imagery for high-throughput phenotyping and yield prediction in barley breeding. *Remote Sensing, 13*(14), Article 2670. https://doi.org/10.3390/rs13142670

Iqbal, A., Khan, R. S., Khan, M. A., Gul, K., Jalil, F., Shah, D. A., Rahman, H., & Ahmed, T. (2021). Genetic engineering approaches for enhanced insect pest resistance in sugarcane. *Molecular Biotechnology, 63*, 557–568. https://doi.org/10.1007/s12033-021-00328-5

Kar, S., Purbey, V. K., Suradhaniwar, S., Korbu, L. B., Kholová, J., Durbha, S. S., Adinarayana, J., & Vadez, V. (2021). An ensemble machine learning approach for determination of the optimum sampling time for evapotranspiration assessment from high-throughput phenotyping data. *Computers and Electronics in Agriculture, 182*, Article 105992. https://doi.org/10.1016/j.compag.2021.105992

Li, Y., Xin, G., Wei, M., Shi, Q., Yang, F., & Wang, X. (2017). Carbohydrate accumulation and sucrose metabolism responses in tomato seedling leaves when subjected to different light qualities. *Scientia Horticulturae, 225*, 490–497. https://doi.org/10.1016/j.scienta.2017.07.053

Liu, G., Yang, Y., Guo, X., Liu, W., Xie, R., Ming, B., Xue, J., Wang, K., Li, S., & Hou, P. (2023). A global analysis of dry matter accumulation and allocation for maize yield breakthrough from 1.0 to 25.0 Mg $ha^{-1}$. *Resources, Conservation and Recycling, 188*, Article 106656. https://doi.org/10.1016/j.resconrec.2022.106656

Lu, Z., Meng, Y., Fan, H., Lu, J., Zhong, X., Ou, Y., Mo, H., & Zhou, L. (2021). Luminescent properties of $Mn^{4+}$-doped $LaTiSbO_6$ deep-red-emitting phosphor for plant growth LEDs. *Journal of Luminescence, 236*, Article 118100. https://doi.org/10.1016/j.jlumin.2021.118100

Maciel Junior, I. C., Dallacort, R., Boechat, C. L., Teodoro, P. E., Teodoro, L. P. R., Rossi, F. S., Oliveira-Júnior, J. F., Della-Silva, J. L., Baio, F. H. R., Lima, M., & Silva Junior, C. A. (2024). Maize crop detection through geo-object-oriented analysis using orbital multi-sensors on the Google Earth engine platform. *AgriEngineering, 6*(1), 491–508. https://doi.org/10.3390/agriengineering6010030

Nalepa, J., & Kawulok, M. (2019). Selecting training sets for support vector machines: A review. *Artificial Intelligence Review, 52*, 857–900. https://doi.org/10.1007/s10462-017-9611-1

Niazian, M., & Niedbała, G. (2020). Machine learning for plant breeding and biotechnology. *Agriculture, 10*(10), Article 436. https://doi.org/10.3390/agriculture10100436

Nishio, J. N. (2000). Why are higher plants green? Evolution of the higher plant photosynthetic pigment complement. *Plant, Cell & Environment, 23*(6), 539–548. https://doi.org/10.1046/j.1365-3040.2000.00563.x

Oliveira, J. F., Alcântara, J. F., Santana, D. C., Teodoro, L. P. R., Baio, F. H. R., Coradi, P. C., Silva Junior, C. A., & Teodoro, P. E. (2023). Spectral variables as criteria for selection of soybean genotypes at different vegetative stages. *Remote Sensing Applications: Society and Environment, 32*, Article 101026. https://doi.org/10.1016/j.rsase.2023.101026

Oliveira, J. L. G., Santana, D. C., Oliveira, I. C., Gava, R., Baio, F. H. R., Silva Junior, C. A.,Teodoro, L. P. R., Teodoro, P. E., & Oliveira, J. T. (2025). Classification of irrigation management practices in maize hybrids using multispectral sensors and machine learning techniques. *Engenharia Agrícola, 45*, Article e20240164. https://doi.org/10.1590/1809-4430-Eng.Agric.v45e20240164/2025

Osco, L. P., Marcato Junior, J., Ramos, A. P. M., Furuya, D. E. G., Santana, D. C., Teodoro, L. P. R., Gonçalves, W. N., Baio, F. H. R., Pistori, H., & Silva Junior, C. A. (2020). Leaf nitrogen concentration and plant height prediction for maize using UAV-based multispectral imagery and machine learning techniques. *Remote Sensing, 12*(19), Article 3237. https://doi.org/10.3390/rs12193237

Pantaleão, A. A., Teodoro, L. P. R., Argentel Martínez, L., González Aguilera, J., Campos, C. N. S., Baio, F. H. R., Silva Júnior, C. A., & Teodoro, P. E. (2022). Soybean base saturation stress:

Selecting populations for multiple traits using multivariate statistics. *Journal of Agronomy and Crop Science, 208*(2), 168–177. https://doi.org/10.1111/jac.12564

Qi, M., & Zhang, G. P. (2001). An investigation of model selection criteria for neural network time series forecasting. *European Journal of Operational Research, 132*(3), 666–680. https://doi.org/10.1016/S0377-2217(00)00171-5

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann. https://dl.acm.org/doi/10.5555/583200

Ramos, A. P. M., Osco, L. P., Furuya, D. E. G., Gonçalves, W. N., Santana, D. C., Teodoro, L. P. R., Silva Junior, C. A., Capristo-Silva, G. F., Li, J., & Baio, F. H. R. (2020). A random forest ranking approach to predict yield in maize with UAV-based vegetation spectral indices. *Computers and Electronics in Agriculture, 178*, Article 105791. https://doi.org/10.1016/j.compag.2020.105791

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf

Rivero, R. M., Mittler, R., Blumwald, E., & Zandalinas, S. I. (2022). Developing climate-resilient crops: Improving plant tolerance to stress combination. *The Plant Journal, 109*(2), 373–389. https://doi.org/10.1111/tpj.15483

Santana, D. C., Cunha, M. P. O., Santos, R. G., Cotrim, M. F., Teodoro, L. P. R., Silva Junior, C. A., Baio, F. H. R., & Teodoro, P. E. (2022). High-throughput phenotyping allows the selection of soybean genotypes for earliness and high grain yield. *Plant Methods, 18*, Article 13. https://doi.org/10.1186/s13007-022-00848-4

Santana, D. C., Santos, R. G., Teodoro, L. P. R., Silva Junior, C. A., Baio, F. H. R., Coradi, P. C., & Teodoro, P. E. (2022). Structural equation modelling and factor analysis of the relationship between agronomic traits and vegetation indices in corn. *Euphytica, 218*, Article 44. https://doi.org/10.1007/s10681-022-02997-y

Santana, D. C., Teixeira Filho, M. C. M., Silva, M. R., Chagas, P. H. M., Oliveira, J. L. G., Baio, F. H. R., Campos, C. N. S., Teodoro, L. P. R., Silva Junior, C. A., Teodoro, P. E., & Shiratsuchi, L. S. (2023). Machine learning in the classification of soybean genotypes for primary macronutrients' content using UAV–multispectral sensor. *Remote Sensing, 15*(5), Article 1457. https://doi.org/10.3390/rs15051457

Santana, D. C., Teodoro, L. P. R., Baio, F. H. R., Santos, R. G., Coradi, P. C., Biduski, B., Silva Junior, C. A., Teodoro, P. E., & Shiratsuchi, L. S. (2023). Classification of soybean genotypes for industrial traits using UAV multispectral imagery and machine learning. *Remote Sensing Applications: Society and Environment, 29*, Article 100919. https://doi.org/10.1016/j.rsase.2023.100919

Santos, H. G., Jacomine, P. K. T., Anjos, L. H. C., Oliveira, V. A., Lumbreras, J. F., Coelho, M. R., Almeida, J. A., Araujo Filho, J. C., Oliveira, J. B., & Cunha, T. J. F. (2018). *Sistema brasileiro de classificação de solos* (5th ed.). Embrapa. https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1094003

Santos, T. T., & Yassitepe, J. E. C. T. (2014). Fenotipagem de plantas em larga escala: um novo campo de aplicação para a visão computacional na agricultura. In S. M. F. S. Massruhá, M. A. A. Leite, A. Luchiari Junior, & L. A. S. Romani (Eds.), *Tecnologias da informação e comunicação e suas relações com a agricultura* (pp. 85–100). Embrapa. https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1010708/1/capitulo0508814.pdf

Scott, A. J., & Knott, M. (1974). Cluster analysis method for grouping means in the analysis of variance. *Biometrics, 30*(3), 507–512. https://doi.org/10.2307/2529204

Shi, L., Han, Y. J., Wang, H., Shi, D., Geng, X., & Zhang, Z. (2019). High-efficiency and thermally stable far-red emission of $Mn^{4+}$ in double cubic perovskite $Sr_9Y_2W_4O_{24}$ for plant cultivation. *Journal of Luminescence, 208*, 307–312. https://doi.org/10.1016/j.jlumin.2018.12.065

Silva, V. S., Silva, C. A., Mohan, M., Cardil, A., Rex, F. E., Loureiro, G. H., Almeida, D. R. A., Broadbent, E. N., Gorgens, E. B., Dalla Corte, A. P., Silva, E. A., Valbuena, R., & Klauberg, C. (2020). Combined impact of sample size and modeling approaches for predicting stem volume in *Eucalyptus* spp. forest plantations using field and LiDAR data. *Remote Sensing, 12*(9), Article 1438. https://doi.org/10.3390/rs12091438

Štepanovský, M., Ibrová, A., Buk, Z., & Velemínská, J. (2017). Novel age estimation model based on development of permanent teeth compared with classical approach and other modern data mining methods. *Forensic Science International, 279*, 72–82. https://doi.org/10.1016/j.forsciint.2017.08.005

Swarup, S., Cargill, E. J., Crosby, K., Flagel, L., Kniskern, J., & Glenn, K. C. (2021). Genetic diversity is indispensable for plant breeding to improve crops. *Crop Science, 61*(2), 839–852. https://doi.org/10.1002/csc2.20377

Taiz, L., Zeiger, E., Møller, I. M., & Murphy, A. (2017). *Fisiologia e desenvolvimento vegetal* (6th ed.). Artmed Editora.

Teodoro, P. E., Teodoro, L. P. R., Baio, F. H. R., Silva Junior, C. A., Santos, R. G., Ramos, A. P. M., Pinheiro, M. M. F., Osco, L. P., Gonçalves, W. N., Carneiro, A. M., Pistori, H., & Shiratsuchi, L. S. (2021). Predicting days to maturity, plant height, and grain yield in soybean: A machine and deep learning approach using multispectral data. *Remote Sensing, 13*(22), Article 4632. https://doi.org/10.3390/rs13224632

Van Eeuwijk, F. A., Bustos-Korts, D., Millet, E. J., Boer, M. P., Kruijer, W., Thompson, A., Malosetti, M., Iwata, H., Quiroz, R., Kuppe, C., Muller, O., Blazakis, K. N., Yu, K., Tardieu, F., & Chapman, S. C. (2019). Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Science, 282*, 23–39. https://doi.org/10.1016/j.plantsci.2018.06.018

Vidyarthi, V. K., Jain, A., & Chourasiya, S. (2020). Modeling rainfall-runoff process using artificial neural network with emphasis on parameter sensitivity. *Modeling Earth Systems and Environment, 6*, 2177–2188. https://doi.org/10.1007/s40808-020-00833-7

Wu, Q., Su, N., Shen, W., & Cui, J. (2014). Analyzing photosynthetic activity and growth of *Solanum lycopersicum* seedlings exposed to different light qualities. *Acta Physiologiae Plantarum, 36*, 1411–1420. https://doi.org/10.1007/s11738-014-1519-7

Zafar, M. M., Mustafa, G., Shoukat, F., Idrees, A., Ali, A., Sharif, F., Shakeel, A., Mo, H., Youlu, Y., Ali, Q., Razzaq, A., Ren, M., & Li, F. (2022). Heterologous expression of *cry3Bb1* and *cry3* genes for enhanced resistance against insect pests in cotton. *Scientific Reports, 12*, Article 10878. https://doi.org/10.1038/s41598-022-13295-x

Zahir, S. A. D. M., Omar, A. F., Jamlos, M. F., Azmi, M. A. M., & Muncan, J. (2022). A review of visible and near-infrared (Vis-NIR) application in plant stress detection. *Sensors and Actuators A: Physical, 338*, Article 113468. https://doi.org/10.1016/j.sna.2022.113468

**SUPPLEMENTARY TABLE S1**. List of 41 vegetation indices.

| Acronym | Vegetation index (VI) | Equation |
|---|---|---|
| ARVI2 | Atmospheric resilient vegetation index 2 | $-0.18 + 1.17 * [\frac{(R_{nir} - R_{red})}{(R_{nir} + R_{red})}]$ |
| ATSAVI | Transformed soil-adjusted vegetation index | $1.22 * [\frac{(R_{nir} - 1.22 * R_{red} - 0.03)}{(1.22 * R_{nir} + R_{red} - 1.22 * 0.03 + 0.0891 + 1.22^2)}]$ |
| CCCI | Canopy chlorophyll content index | $\frac{(R_{nir} - R_{rededge})/(R_{nir} + R_{rededge})}{(R_{nir} - R_{red})/(R_{nir} + R_{red})}$ |
| CIgreen | Green chlorophyll index | $\frac{NIR}{green} - 1$ |
| CTVI | Correct vegetation index transformed | $\frac{NDVI + 0.5}{NDVI + 0.5} * \sqrt{NDVI + 0.5}$ |
| CVI | Chlorophyll vegetation index | $NIR\frac{red}{green^2}$ |
| DVI | Differentiated vegetation index | $\frac{R_{nir}}{R_{red}}$ |
| EVEI2 | Enhanced vegetation index 2 | $\frac{2.5 * (R_{nir} - R_{red})}{(R_{nir} + 2.4 * R_{red} + 1)}$ |
| EVI | Improved vegetation index | $2.5 * \frac{(R_{NIR} - R_{RED})}{((R_{NIR}) + (C1 * R_{NIR}) - (C2 * R_{BLUE}) + L)}$ |
| GDVI | NIR difference/vegetation index by green difference | $R_{nir} - R_{green}$ |
| GEMI | Global environmental monitoring index | $\frac{2 * (1 - 0.25 * 2) - ((R_{red} - 0.125)}{(1 - R_{red})}$ |
| GLI | Green leaf index | $\frac{(2 * R_{green} - R_{red} - R_{blue})}{(2 * R_{green} + R_{red} + R_{blue})}$ |
| GNDVI | Green normalized vegetation index | $\frac{(R_{nir} - R_{green})}{(R_{nir} + R_{green})}$ |
| GSAVI | Green soil adjusted vegetation index | $\frac{(1 + L) * (R_{nir} - R_{green})}{(R_{nir} + R_{green} + L)}$ |
| GTVI | Green triangle vegetation index | $\frac{NDVI + 0.5}{NDVI + 0.5} * [(\sqrt{NDVI} + 0.5)]$ |
| IAF | Leaf area index | $\frac{\ln(\frac{0.69 - SAVI}{0.59})}{0.91}$ |
| IPVI | Infrared vegetation percentage index | $R_{nir}/\left(R_{nir} + \frac{R_{red)}}{2}\right) * (NDVI + 1)$ |
| LnRE | Red rim vegetation index | $(\ln R_{NIR} - \ln_{REDGE}) * 100$ |
| LogR | Log r | $\log\frac{R_{nir}}{R_{red}}$ |
| MCARI | Modified chlorophyll absorption reflectance index | $R_{700} - R_{670} - 0.2(R_{700} - R_{(550)})\frac{R_{700}}{R_{670}}$ |

To be continue

| Acronym | Vegetation index (VI) | Equation |
|---|---|---|
| MSAVI | Modified soil-adjusted vegetation index | $\frac{2\text{nir}+1-\sqrt{(2\text{nir}+1)^2-(8\text{nir}-\text{red})}}{2}$ |
| MSR | Modified soil proportion | $\frac{2*\text{R}_{\text{nir}}+1-\sqrt{(2*\text{R}_{\text{nir}}+1)^2-8*(\text{R}_{\text{NIR}}-\text{R}_{\text{red}})}}{2}$ |
| MSRNir _ Red | Modified simple rate nir/red | $(\frac{\text{R}_{\text{nir}}}{\text{R}_{\text{red}}}-1)/\sqrt{\frac{\text{R}_{\text{nir}}}{\text{R}_{\text{red}}}+1}$ |
| MTVI | Modified triangular vegetation index | $\frac{1.5\,[1.2(\text{nir}-\text{green})-2.5(\text{red}-\text{green})]}{\sqrt{(2\text{nir}+1)^2-(6\text{nir}-5\sqrt{\text{red}})}-0.5}$ |
| NDRE | Rededge normalized difference vegetation index | $\frac{(\text{R}_{\text{NIR}}-\text{R}_{\text{EDGE}})}{(\text{R}_{\text{NIR}}+\text{R}_{\text{EDGE}})}$ |
| NDVI | Normalized difference vegetation index | $\frac{(\text{R}_{\text{NIR}}-\text{R}_{\text{RED}})}{(\text{R}_{\text{NIR}}+\text{R}_{\text{RED}})}$ |
| NGRDI | Normalized green-red difference index | $\frac{\text{R}_{\text{green}}-\text{R}_{\text{red}}}{\text{R}_{\text{green}}}/(\text{R}_{\text{nir}}+\text{R}_{\text{red}}+\text{R}_{\text{green}})$ |
| NIR/G | Simple ratio nir and green | $\frac{\text{R}_{\text{nir}}}{\text{R}_{\text{green}}}$ |
| NIR/R | Simple ratio nir and red | $\frac{\text{R}_{\text{nir}}}{\text{R}_{\text{red}}}$ |
| NIR/RE | Simple ratio nir and rededge | $\frac{\text{R}_{\text{nir}}}{\text{R}_{\text{re}}}$ |
| Norm _ R _ 1 | R normal | $\frac{\text{R}_{\text{red}}}{(\text{R}_{\text{nir}}+\text{R}_{\text{red}}+\text{R}_{\text{green}})}$ |
| OSAVI | Optimized soil-adjusted vegetation index | $\frac{(1+0.16)*(\text{R}_{\text{nir}}-\text{R}_{\text{red}})}{(\text{R}_{\text{nir}}+\text{R}_{\text{red}}+0.16)}$ |
| PNDVI | | $\frac{\text{R}_{\text{nir}}-(\text{R}_{\text{green}}+\text{R}_{\text{red}}+\text{R}_{\text{blue}})}{\text{R}_{\text{nir}}+(\text{R}_{\text{green}}+\text{R}_{\text{red}}+\text{R}_{\text{blue}})}$ |
| RDVI | Renormalized difference vegetation index | $\frac{(\text{R}_{\text{nir}}-\text{R}_{\text{red}})}{\sqrt{\text{R}_{\text{nir}}+\text{R}_{\text{red}}}}$ |
| SAVI | Soil-adjusted vegetation index | $(1+0.5)\frac{\text{nir}-\text{red}}{\text{nir}+\text{red}+0.5}$ |
| SQRT _ IR _ R | | $\sqrt{\frac{\text{R}_{\text{nir}}}{\text{R}_{\text{red}}}}$ |
| TVI | Transformed vegetation index | $\sqrt{\frac{\text{R}_{\text{nir}}-\text{R}_{\text{red}}}{\text{R}_{\text{nir}}+\text{R}_{\text{red}}}}$ |
| VARI | | $\frac{\text{R}_{\text{green}}-\text{R}_{\text{red}}}{\text{R}_{\text{green}}+\text{R}_{\text{red}}-\text{R}_{\text{bllue}}}$ |
| WDRVI | Wide dynamic range vegetation index | $\frac{(0.1\text{R}_{\text{nir}}-\text{R}_{\text{red}})}{(0.1*\text{R}_{\text{nir}}+\text{R}_{\text{red}})}$ |

$R_{NIR}$: reflectance in the near infrared range; $R_{GREEN}$: reflectance in the green range; $R_{RED}$: reflectance in the red range; Rre: reflectance in the red transition range (Red-edge); L: ground effect correction factor.