

## HOW MULTILAYER PERCEPTRON ARCHITECTURE AND TRAINING HYPERPARAMETERS AFFECT PREDICTIVE PERFORMANCE IN UAV-BASED SOYBEAN PHENOTYPING

João Amaro Ferreira Vieira Netto<sup>1</sup>, Paulo Cesar Lima Sousa<sup>2</sup>, Paulo Eduardo Teodoro<sup>3</sup>, Leonardo Lopes Bhering<sup>4</sup>

### ABSTRACT:

Multilayer perceptron (MLP) models have been increasingly applied to high-throughput phenotyping using unmanned aerial vehicle (UAV)-derived multispectral imagery. However, the effects of network architecture and training hyperparameters on predictive performance remain poorly understood. This study evaluated the number of neurons, hidden layers, training epochs, and cross-validation folds, as well as the performance of different combinations of these factors in MLP models to predict soybean phenotypic traits using vegetation indices. A total of 32 soybean genotypes were evaluated over three growing seasons, generating 384 experimental plots. Seven vegetation indices derived from UAV-based multispectral imagery were used to predict plant height (PH), first pod insertion height (FPIH), days to maturity (DM), and grain yield (GY). Under the conditions evaluated, increasing network width improved predictive performance up to a plateau, whereas additional hidden layers provided no consistent advantage. Likewise, under the conditions studied, extending training from 10,000 to 30,000 epochs improved prediction accuracy, but further increases produced only marginal gains. Models evaluated using 15-fold cross-validation consistently achieved the highest predictive performance. The effects of architectural and training parameters also depended on the biological nature of the target trait, with PH and FPIH showing higher predictive performance than DM and GY. Overall, the results demonstrate that optimizing architecture and training parameters can improve model performance, but the extent of this improvement depends on the ability of vegetation indices to represent the target phenotypic trait. These findings provide practical guidance for configuring MLP models in UAV-based high-throughput soybean phenotyping.

**Keywords:** artificial neural networks; high-throughput phenotyping; multispectral imagery; vegetation indices; remote sensing.

## COMO A ARQUITETURA DE PERCEPTRON MULTICAMADAS E OS HIPERPARÂMETROS DE TREINAMENTO AFETAM O DESEMPENHO PREDITIVO NA FENOTIPAGEM DE SOJA BASEADA EM VANTS

### RESUMO:

Modelos de Perceptron Multicamadas (MLP) têm sido cada vez mais aplicados à fenotipagem de alto rendimento utilizando imagens multiespectrais derivadas de veículos aéreos não tripulados (UAVs). No entanto, os efeitos da arquitetura da rede e dos hiperparâmetros de treinamento sobre o desempenho preditivo permanecem pouco compreendidos. Este estudo avaliou o número de neurônios, das camadas ocultas, das épocas de treinamento e das partições da validação cruzada, bem como o desempenho das diferentes

<sup>1</sup>Estudante de doutorado em Genética e Melhoramento na Universidade Federal de Viçosa, Departamento de Biologia Geral, Viçosa, MG, Brasil. [joao.netto@ufv.br](mailto:joao.netto@ufv.br). <https://orcid.org/0009-0002-9292-4894>. <sup>2</sup> Estudante de Agronomia pela Universidade Federal de Viçosa, Viçosa, MG, Brasil. [paulo.c.sousa@ufv.br](mailto: paulo.c.sousa@ufv.br). <https://orcid.org/0009-0007-0533-448X>. <sup>3</sup> Professor Adjunto, Universidade Federal de Mato Grosso do Sul, Chapadão do Sul, MS, Brazil. [paulo.teodoro@ufms.br](mailto:paulo.teodoro@ufms.br). <https://orcid.org/0000-0002-8236-542X>. <sup>4</sup>Professor Associado. Doutor em Genética e Melhoramento. Universidade Federal de Viçosa, Departamento de Biologia Geral, Viçosa, MG, Brasil. [leonardo.bhering@ufv.br](mailto:leonardo.bhering@ufv.br). <https://orcid.org/0000-0002-6072-0996>.

✉ Autor Correspondente.

combinações desses fatores, nos modelos MLP para prever caracteres fenotípicos da soja a partir de índices de vegetação. Um total de 32 genótipos de soja foi avaliado ao longo de três safras, gerando 384 parcelas experimentais. Sete índices de vegetação derivados de imagens multiespectrais obtidas por UAV foram utilizados para prever altura de plantas (PH), altura de inserção da primeira vagem (FPIH), dias para maturidade (DM) e produtividade de grãos (GY). Nas condições avaliadas, o aumento da largura da rede melhorou o desempenho preditivo até um platô, enquanto camadas ocultas adicionais não proporcionaram vantagem consistente. Da mesma forma, sob as condições estudadas, estender o treinamento de 10.000 para 30.000 épocas melhorou a acurácia da predição, mas aumentos adicionais produziram apenas ganhos marginais. Os efeitos dos parâmetros de arquitetura e treinamento também dependeram da natureza biológica do caráter-alvo, com PH e FPIH apresentando desempenho superior a DM e GY. De modo geral, os resultados demonstram que a otimização da arquitetura e dos parâmetros de treinamento pode melhorar o desempenho dos modelos MLP, mas seu potencial é condicionado pela capacidade dos índices de vegetação de representar o caráter fenotípico avaliado. Esses resultados fornecem orientações práticas para a configuração de modelos MLP na fenotipagem de alto rendimento da soja baseada em UAV.

**Palavras-chave:** Redes neurais artificiais; Fenotipagem de alto rendimento; Imagens multiespectrais; Índices de vegetação; Sensoriamento remoto.

## INTRODUCTION

The increasing global demand for food, combined with the need to improve the efficiency and sustainability of agricultural systems, has intensified the search for technologies capable of accelerating genetic gains in plant breeding (Hickey *et al.*, 2019). Soybean (*Glycine max* (L.) Merr.) is one of the world's most economically important crops, serving as a major source of food, feed, and industrial raw materials. Brazil, the United States, and Argentina currently account for most of the global soybean production (FAO, 2025; Modgil *et al.*, 2021).

Despite substantial breeding advances over recent decades, the development of cultivars with higher yield, broader environmental adaptation, and improved tolerance to biotic and abiotic stresses remains a major challenge (Fodor *et al.*, 2017). A key limitation is the acquisition of reliable large-scale phenotypic data, as conventional phenotyping often relies on manual, destructive, and labor-intensive evaluations performed across numerous experimental plots (Santana *et al.*, 2022).

High-throughput phenotyping (HTP) has emerged as an effective approach to overcome this bottleneck by enabling rapid, non-destructive, and large-scale phenotypic assessments. The integration of sensors mounted on ground-based and aerial platforms has substantially increased the capacity to monitor plant development under diverse environmental conditions (Yang *et al.*, 2020; Song *et al.*, 2021; Teodoro *et al.*, 2024). In particular, multispectral sensors deployed on unmanned aerial vehicles (UAVs) allow the generation of vegetation indices (VIs), which provide indirect measurements of canopy architecture, physiological status, biomass accumulation, and chlorophyll content. These indices have been widely explored as predictors of agronomic traits in crop breeding programs (Xue & Su, 2017; Tayade *et al.*, 2022).

The increasing adoption of UAV-based phenotyping has generated large and complex datasets that require advanced computational approaches for extracting biologically meaningful information. Machine learning methods, particularly artificial neural networks (ANNs), have therefore become increasingly important in agricultural applications involving classification, prediction, and pattern recognition (Gill *et al.*, 2022). Among these methods, the multilayer perceptron (MLP) has attracted considerable attention because of its ability

to model complex nonlinear relationships between predictor and response variables (Kruse *et al.*, 2022). This capability is particularly relevant for HTP applications, where phenotypic expression results from complex interactions among genetic, physiological, and environmental factors.

Despite their growing use in agricultural research, most studies employing MLPs have focused primarily on comparing machine learning algorithms or reporting the performance of a single network configuration. Consequently, relatively little attention has been given to understanding how architectural and training hyperparameters influence predictive performance in HTP applications. This limitation is particularly relevant because hyperparameters such as the number of hidden layers, the number of neurons per layer, the number of training epochs, and the cross-validation strategy directly influence model complexity, convergence, computational cost, and generalization ability (Vujicic *et al.*, 2016; El-Hassani *et al.*, 2024). As a result, the selection of an appropriate network configuration remains largely empirical, making it difficult to establish methodological guidelines for applying MLPs to real-world phenotyping datasets.

This knowledge gap is especially important for UAV-derived multispectral data, in which predictor variables capture only part of the biological processes underlying agronomic traits (Gill *et al.*, 2022). Under these conditions, increasing network complexity does not necessarily improve predictive performance. The interactions among architectural and training hyperparameters remain poorly understood. A better understanding of these interactions is therefore essential for developing efficient, robust, and computationally feasible prediction models for high-throughput phenotyping.

Accordingly, the objective of this study was to investigate the effects of the number of neurons, hidden layers, training epochs, folds used in cross-validation. These factors and their interactions were evaluated in terms of the predictive performance of multilayer perceptron models for soybean phenotypic traits using vegetation indices derived from UAV-based multispectral imagery.

## MATERIAL AND METHODS

Field experiments were conducted at the experimental station of the Federal University of Mato Grosso do Sul, located in Chapadão do Sul,

Mato Grosso do Sul, Brazil (18°46'26" S, 52°37'28" W; 810 m above sea level), under a conventional tillage system (Figure 1).

A total of 32 soybean genotypes were evaluated in a randomized complete block design with four replications across three growing seasons (2019/2020, 2020/2021, and 2021/2022), resulting in 384 experimental plots ( $n = 384$ ). The same agronomic management practices were maintained across all three growing seasons, including fertilization, weed and pest control, and other crop management practices, following recommendations appropriate for the regional conditions and soil characteristics. Each growing season was considered a distinct environment. Experimental plots consisted

of three-meter-long rows spaced 0.45 m apart with a planting density of 15 plants  $m^{-1}$ . The agronomic traits evaluated were days to maturity (DM), first pod insertion height (FPIH, cm), plant height (PH, cm), and grain yield (GY,  $kg\ ha^{-1}$ ). Days to maturity was determined as the number of days from emergence until more than 50% of the plants within each plot reached physiological maturity. First pod insertion height and plant height were measured on five randomly selected plants per plot using a measuring tape. The mean value was used in the subsequent analyses. Grain yield was determined by harvesting the central 2 m of each plot and adjusting grain moisture to 13%.

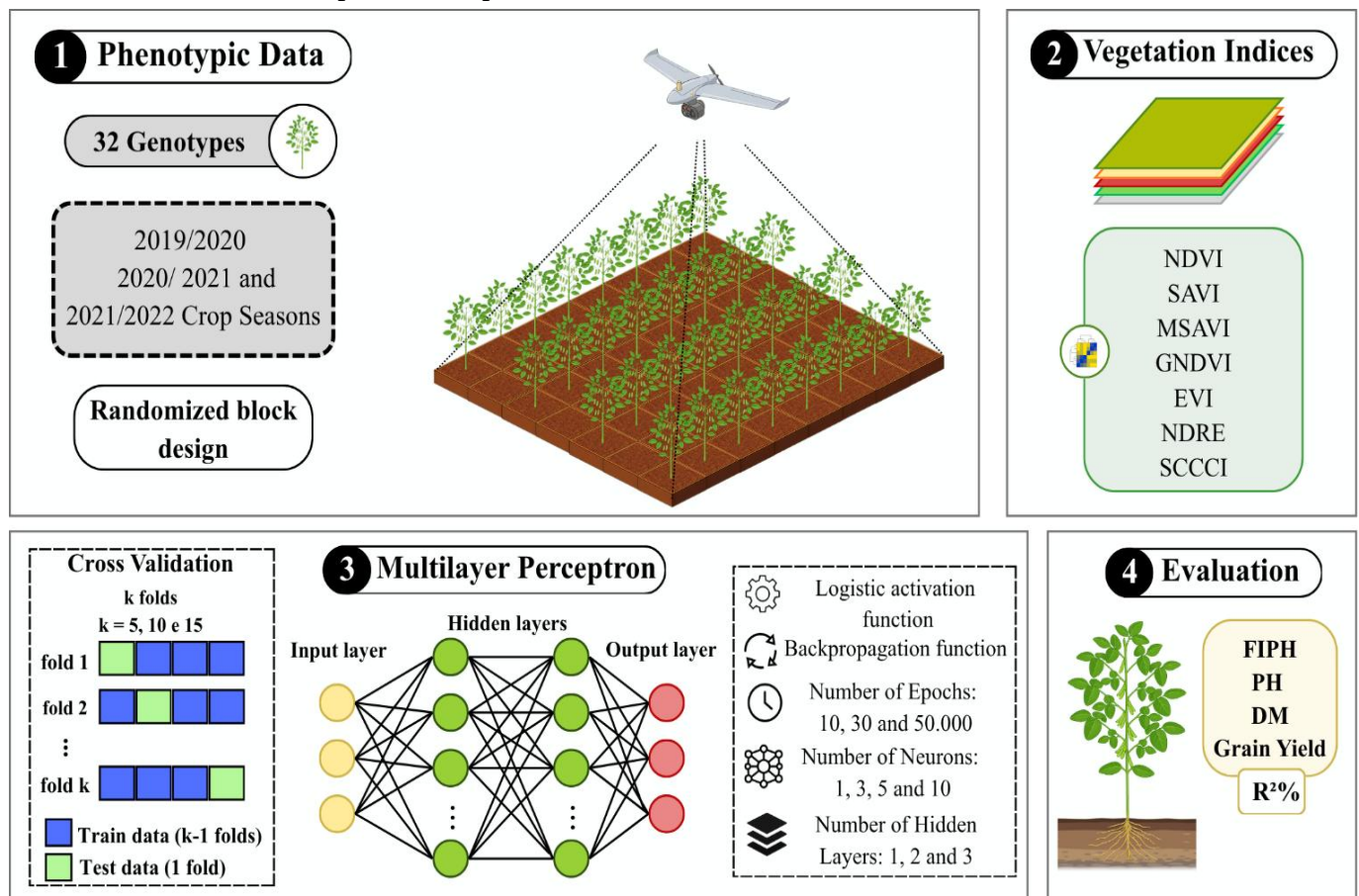


**Figure 1.** Overview of the experimental workflow used to evaluate the performance of multilayer perceptron (MLP) models under different architectural and training configurations. (1) Field experiment comprising 32 soybean genotypes evaluated in a randomized complete block design with four replications across three growing seasons (2019/2020, 2020/2021, and 2021/2022). Fixed-wing unmanned aerial vehicle (UAV) flights were conducted during the flowering stage to acquire multispectral images. (2) Multispectral images were processed to generate vegetation indices (VIs). (3) Multilayer perceptron (MLP) models were evaluated under different configurations of hidden-layer architecture, varying the number of neurons and training epochs. (4) Model performance was evaluated using the coefficient of determination ( $R^2$ ) for first pod insertion height (FPIH), plant height (PH), days to maturity (DM), and grain yield (GY).

Spectral data were acquired during the full flowering stage (R2) using a fixed-wing unmanned aerial vehicle (UAV; SenseFly eBee RTK) equipped with a SenseFly Sequoia multispectral sensor and operated with a fixed RTK base station. Images were captured in the blue, green, red, red-edge, and near-infrared (NIR) spectral bands. Flights were conducted at approximately 09:00 h (local time) at an altitude of 40 m, with 80% forward overlap and 75% side overlap, resulting in a ground sampling distance (GSD) of approximately 7 cm. RTK-assisted navigation provided positioning accuracy of up to 2.5 cm. Image processing was performed using PIX4Dmapper software, where radiometric calibration and conversion of digital numbers to surface reflectance were applied to generate georeferenced orthomosaics for each experimental trial.

Seven vegetation indices (Table 1) were calculated from the orthomosaics and the mean pixel value within each experimental plot was extracted to represent its spectral response. These plot-level spectral data were then associated with the corresponding plot-level agronomic values, calculated as the mean of the five sampled plants for first pod insertion height and plant height.

Phenotypic traits were predicted using multilayer perceptron (MLP) models (Table 2).

Model architectures with one, two, or three hidden layers were evaluated, with each hidden layer comprising 1, 3, 5, or 10 neurons. Cross-validation was performed using 5-, 10-, and 15-fold schemes, combined with training durations of 10,000, 30,000, and 50,000 epochs. Models were trained using the backpropagation algorithm with a logistic activation function for the hidden layers through the `neuralnet()` function, and predictions were obtained using the `predict()` function, both implemented in the `neuralnet` R package (Fritsch *et al.*, 2019).

Vegetation indices were used as predictor (input) variables. Whereas each phenotypic trait was considered the response (output) variable. Prior to model training, the phenotypic response variables were normalized using Min–Max scaling (Cabello-Solorzano *et al.*, 2023) within each cross-validation iteration. The minimum and maximum values were calculated exclusively from the training subset and then used to normalize both the training and corresponding validation subsets. Thus, information from the validation data was not used in the normalization procedure, preventing data leakage. The normalized phenotypic data were then used to train and evaluate the MLP models.

**Table 1.** Vegetation indices included in the analyses.

| Vegetation Index | Equation                                                                        | Reference                     |
|------------------|---------------------------------------------------------------------------------|-------------------------------|
| NDVI             | $\frac{Nir - Red}{Nir + Red}$                                                   | Xue & Su (2017)               |
| SAVI             | $(1 + L) \frac{NIR - Red}{NIR + Red + L}$                                       | Liu & Huete (1995)            |
| MSAVI            | $2 \times NIR + 1 - \sqrt{(2 \times NIR + 1)^2 - 8 \times \frac{NIR - Red}{2}}$ | Xue & Su (2017)               |
| GNDVI            | $\frac{Nir - Green}{Nir + Green}$                                               | Gitelson <i>et al.</i> (1996) |
| EVI              | $G \times \frac{NIR - Red}{NIR + C_1 \times Red - C_2 \times Blue + L}$         | Liu & Huete (1995)            |
| NDRE             | $\frac{Nir - Rededge}{Nir + Rededge}$                                           | Li <i>et al.</i> (2014)       |
| SCCCI            | $\frac{NDRE}{NDVI}$                                                             | Sumner <i>et al.</i> (2021)   |

NDVI: normalized difference vegetation index; SAVI: soil-adjusted vegetation index; MSAVI: modified soil-adjusted vegetation index; GNDVI: green normalized difference vegetation index; EVI: enhanced vegetation index; NDRE: normalized difference red edge; SCCCI: simplified canopy chlorophyll content index; NIR: near-infrared band; Red: red band; Green: green band; Blue: blue band; RedEdge: red edge band; L: soil adjustment factor (1.0); G: gain factor (2.5); C1 and C2: atmospheric correction coefficients (6.0 and 7.5, respectively).

**Table 2.** General configuration of the multilayer perceptron (MLP) models used

| Parameter                   | Values                       |
|-----------------------------|------------------------------|
| Number of layers            | 1, 2 and 3                   |
| Number of neurons per layer | 1, 3, 5 and 10               |
| Activation function         | Logistic                     |
| Training function           | Backpropagation              |
| Evaluation method           | R <sup>2</sup>               |
| Number of training epochs   | 10.000, 30.000 and 50.000    |
| Number of data points       | 384 rows                     |
| Training type               | Cross-validation             |
| Number of folds/repetitions | 5, 10 and 15 / 2 repetitions |

Model performance was evaluated using 5-, 10-, and 15-fold cross-validation, with two independent repetitions for each combination of network architecture and training hyperparameters. In each run, the dataset was partitioned into training and validation subsets, ensuring that every observation was included in the validation process across the cross-validation iterations. For each repetition, a new random partition of the dataset was generated. In each cross-validation fold, the model was trained using the remaining folds and evaluated on the held-out fold, yielding one R<sup>2</sup> value per fold. The mean and standard deviation were then calculated from the R<sup>2</sup> values obtained across all folds and repetitions for each model configuration.

For each phenotypic trait and combination of network architecture, cross-validation scheme, and training epochs, predictive performance was assessed using the coefficient of determination (R<sup>2</sup>), calculated from the observed and predicted values according to the following equation:

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

To assess the effects of genotype and environment on the evaluated traits, an analysis of variance (ANOVA) was performed using the R statistical environment (version 4.5.2). The statistical

model included genotype, growing season, block nested within growing season, and the genotype × growing season interaction as fixed effects and was expressed as follows:

$$Y_{ijk} = \mu + G_i + S_j + B_{k(j)} + (GS)_{ij} + \varepsilon_{ijk}$$

Where  $Y_{ijk}$  is the observed value of the trait for genotype  $i$ , growing season  $j$ , and block  $k$  nested within growing season  $j$ ;  $\mu$  is the overall mean;  $G_i$  is the fixed effect of genotype  $i$ ;  $S_j$  is the fixed effect of growing season  $j$ ;  $B_{k(j)}$  is the fixed effect of block  $k$  nested within growing season  $j$ ;  $(GS)_{ij}$  is the genotype × growing season interaction; and  $\varepsilon_{ijk}$  is the experimental error.

## RESULTS AND DISCUSSION

The analysis of variance revealed significant effects ( $p < 0.01$ ) of genotype, growing season, and genotype × season interaction for all evaluated traits, namely plant height (PH), first pod insertion height (FPIH), days to maturity (DM), and grain yield (GY) (Table 3). These results confirm substantial genetic variability among soybean genotypes and highlight the strong influence of environmental conditions on phenotypic expression, providing the variability

required for machine learning models to identify relationships between spectral predictors and target traits.

**Table 3.** Analysis of variance for plant height, first pod insertion height, days to maturity, and grain yield, showing the effects of genotype, season, block within season, and genotype  $\times$  season interaction, along with their respective degrees of freedom (Df), sum of squares, mean squares, and F-values.

| Factor                            | Df  | Sum of squares | Mean squares | F value    |
|-----------------------------------|-----|----------------|--------------|------------|
| <i>Plant height</i>               |     |                |              |            |
| Genotype                          | 31  | 8,415          | 271.5        | 4.024*     |
| Season                            | 2   | 4,556          | 2,278.2      | 33.768*    |
| Season/block                      | 9   | 2,170          | 241.1        | 3.574*     |
| Genotype $\times$ season          | 62  | 13,356         | 215.4        | 3.193*     |
| Residuals                         | 279 | 18,823         | 67.5         |            |
| <i>First pod insertion height</i> |     |                |              |            |
| Genotype                          | 31  | 1,019          | 33.0         | 5.670***   |
| Season                            | 2   | 7,989          | 3,995.0      | 689.240*** |
| Season/block                      | 9   | 273            | 30.0         | 5.227***   |
| Genotype $\times$ season          | 62  | 2,207          | 36.0         | 6.141***   |
| Residuals                         | 279 | 1,617          | 6.0          |            |
| <i>Days to maturity</i>           |     |                |              |            |
| Genotype                          | 31  | 501            | 16.2         | 2.463***   |
| Season                            | 2   | 856            | 428.2        | 65.195***  |
| Season/block                      | 9   | 92             | 10.3         | 1.562ns    |
| Genotype $\times$ season          | 62  | 1,020          | 16.5         | 2.506***   |
| Residuals                         | 279 | 1,832          | 6.6          |            |
| <i>Grain yield</i>                |     |                |              |            |
| Genotype                          | 31  | 73,150,545     | 2,359,695    | 2.231***   |
| Season                            | 2   | 230,627,726    | 115,313,863  | 109.017*** |
| Season/block                      | 9   | 23,801,059     | 2,644,562    | 2.500**    |
| Genotype $\times$ season          | 62  | 130,918,299    | 2,111,585    | 1.996***   |
| Residuals                         | 279 | 295,115,474    | 1,057,762    |            |

Significance levels: ns = non-significant; \*, \*\* and \*\*\* significant at 5%, 1% and 0.1% probability levels, respectively.

The significant genotype  $\times$  season interaction further indicates that genotype performance varied across growing seasons, increasing the complexity of the prediction problem because similar spectral responses may correspond to different phenotypic outcomes under different environmental conditions (Araus *et al.*, 2018). Consequently, predictive models must identify relationships between vegetation

indices and phenotypes that remain robust across contrasting environments. Similar challenges have been widely reported in high-throughput phenotyping studies, where genotype  $\times$  environment interactions often reduce prediction accuracy, particularly for complex agronomic traits (Araus *et al.*, 2018; Yang *et al.*, 2020; Gill *et al.*, 2022).

Against this background, differences in predictive performance among the evaluated traits were expected. Across all tested MLP architectures and training configurations, PH consistently achieved the highest predictive accuracy, frequently exceeding  $R^2$  values of 80 %, whereas FPIH reached intermediate values, with maximum  $R^2$  approaching 70 %. Conversely, DM and GY consistently exhibited

lower predictive performance regardless of the evaluated hyperparameter combination (Figure 2). Regardless of the architecture evaluated, structural traits were consistently predicted more accurately than integrative traits, demonstrating that predictive performance was primarily constrained by the biological characteristics of the target trait.

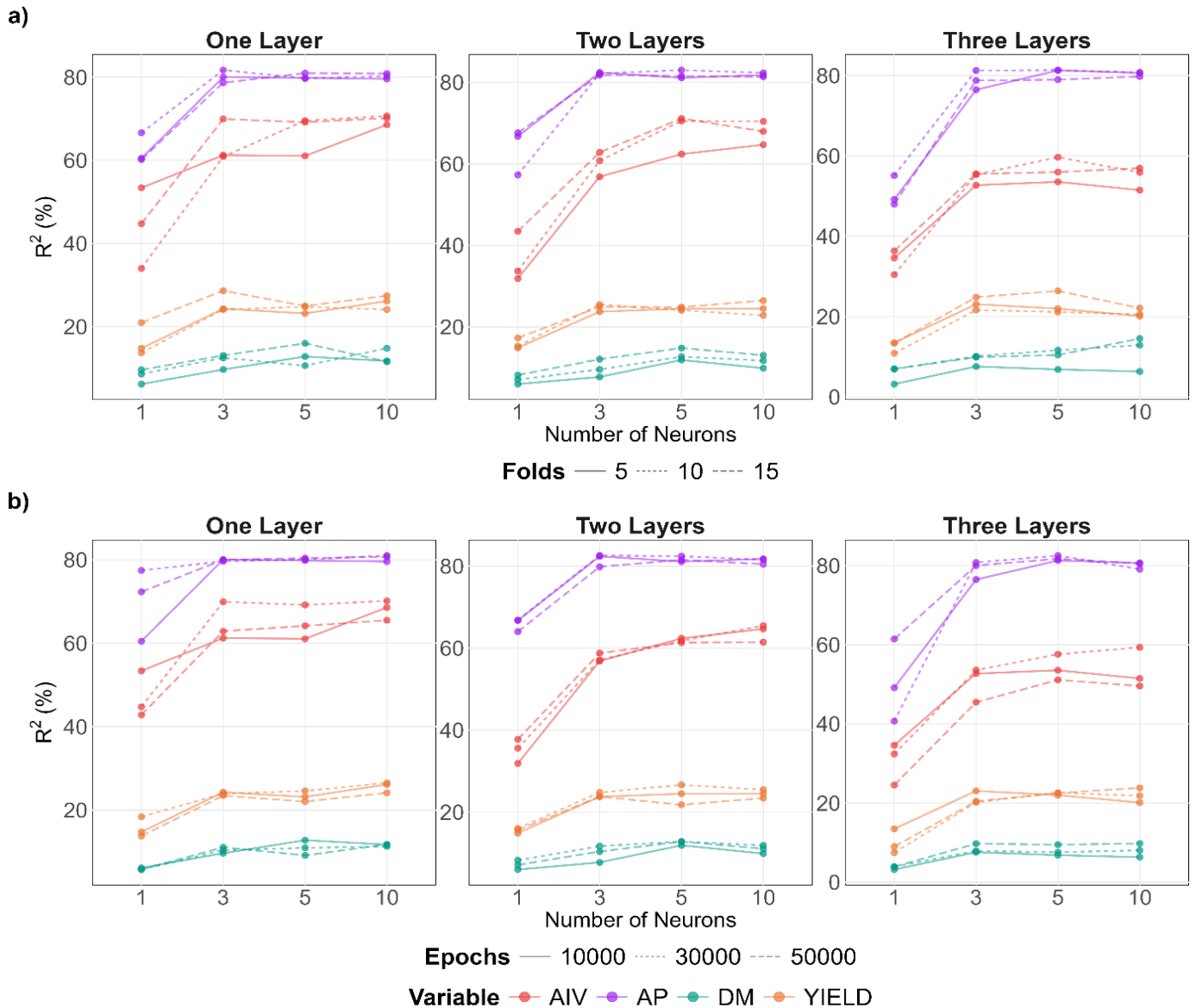


**Figure 2.** Predictive performance ( $R^2\%$ ) of MLP. Panels show (a) the relationship between folds, number of neurons, and layers and (b) the relationship between epochs, number of neurons, and layers.

The superior prediction of PH and FPIH reflects their close relationship with canopy structural attributes. As discussed by Xue and Su (2017), Yang *et al.* (2020), Tayade *et al.* (2022), and Gill *et al.* (2022), multispectral vegetation indices effectively capture canopy architecture, biomass accumulation, and chlorophyll content, providing a biologically

meaningful representation of structural traits. Consistent with these observations, the MLP models achieved higher predictive performance for these structural traits.

Although both structural traits were predicted with satisfactory accuracy, PH consistently outperformed FPIH. Unlike PH, FPIH depends on

reproductive structures located in the lower canopy, which become increasingly obscured as canopy closure progresses. As a result, multispectral sensors estimate this trait only indirectly through its association with overall plant architecture, reducing prediction accuracy. Similar limitations have been reported for passive optical sensing in soybean, where lower-canopy traits are consistently more difficult to predict than canopy structural traits using machine learning models (Vieira Netto *et al.*, 2026).

High-throughput phenotyping studies have consistently shown that spectral predictors achieve higher accuracies for structural traits than for complex integrative traits such as grain yield (Yang *et al.*, 2020; Gill *et al.*, 2022). The same pattern was observed in the present study, where DM and GY consistently exhibited lower predictive performance than PH and FPIH. Unlike structural traits, DM and GY depend on biological processes extending beyond the spectral information available at a single phenological stage. Days to maturity is primarily determined by developmental and genetic mechanisms that are only indirectly reflected in canopy reflectance, whereas grain yield integrates the cumulative effects of genotype, environment, and multiple physiological processes throughout the crop cycle. Consequently, vegetation indices acquired during flowering provide only a partial representation of these traits, inherently limiting predictive performance.

An MLP performance depends not only on network architecture but also on the biological information contained in the predictor variables. Consequently, the effects of neurons, hidden layers, training epochs, and cross-validation strategies should be interpreted as mechanisms for improving the extraction of information already present in the spectral data, rather than overcoming the inherent biological complexity of each target trait.

Network topology affected predictive performance, although the benefits depended on how model complexity was increased. While increasing network width improved prediction accuracy up to a certain point, greater network depth provided no consistent advantage, indicating that additional complexity did not necessarily improve model generalization.

Increasing the number of neurons enhanced predictive performance across all evaluated traits, with the greatest gains occurring between one and three neurons. Beyond five neurons, however, model

performance stabilized, with only marginal differences between architectures containing five and ten neurons. This plateau indicates that moderate network width was sufficient to capture the dominant nonlinear relationships between vegetation indices and phenotypic variation. Similar responses have been reported for MLP applications in agriculture, where increasing network width improves performance only until diminishing returns are reached (Kruse *et al.*, 2022; El-Hassani *et al.*, 2024).

The extent of these gains, however, depended on the target trait. Structural traits (PH and FPIH) benefited more from increasing network width than DM and GY, reinforcing that architectural optimization can only improve the extraction of information already represented by the predictor variables. For traits weakly represented by multispectral vegetation indices, increasing the number of neurons alone was insufficient to substantially improve predictive performance.

In contrast, increasing network depth produced no systematic improvements. Architectures with one and two hidden layers showed similar predictive performance, whereas three-layer networks consistently resulted in lower coefficients of determination across the evaluated traits. Although isolated hyperparameter combinations deviated from this trend, deeper architectures provided no consistent advantage.

Unlike wider networks, deeper architectures substantially increase the number of trainable parameters and are generally more advantageous for large datasets or highly complex learning tasks. However, studies by Vujicic *et al.* (2016), Ibnu *et al.* (2019), and Raut and Dani (2020) showed that shallow MLPs often match or even outperform deeper architectures when applied to moderate-sized agricultural datasets. This explains why, in the present study, the relationships between multispectral vegetation indices and soybean phenotypes were adequately represented by shallow architectures, making additional hidden layers unnecessary.

While wider architectures improved prediction until reaching a performance plateau, deeper networks introduced additional parameters without consistently improving generalization. This suggests that the complexity of the prediction problem did not justify deeper architectures under the conditions evaluated. Therefore, for moderate-sized UAV-based phenotyping datasets, modest increases

in network width appear to be a more effective strategy than increasing network depth.

Among the evaluated training parameters, both the number of training epochs and the cross-validation scheme influenced model performance, although through distinct mechanisms. Predictive performance improved consistently as the number of training epochs increased from 10,000 to 30,000, but extending training to 50,000 epochs produced only marginal gains and, in some configurations, slight reductions in predictive accuracy. Kruse *et al.* (2022) described a similar convergence pattern for MLPs applied to agricultural datasets, in which rapid improvements during the early stages of training are followed by a stabilization of predictive performance.

Although some architectures exhibited slight reductions in predictive performance after prolonged training, these findings should be interpreted cautiously. Because training and validation loss curves were not evaluated, the observed reductions cannot be unequivocally attributed to overfitting. Instead, they suggest that extending optimization beyond the convergence point did not further improve model generalization. Consistent with the architectural results, increasing computational effort alone was insufficient to improve predictive performance once the information contained in the predictor variables had been effectively exploited.

A complementary effect was observed for the cross-validation strategy. Unlike network architecture and training duration, the number of folds does not modify the learning process itself but influences how the available observations are allocated during model calibration and evaluation. Across the evaluated architectures, 15-fold cross-validation consistently produced the highest coefficients of determination, which may be related to the greater proportion of observations available for training in each iteration.

This effect is likely related to the characteristics of the dataset. With 384 field observations, the 15-fold scheme allocated approximately 93% of the data for training in each iteration, exposing the models to a larger proportion of the biological variability present in the experiment. In contrast, validation schemes with fewer folds reduced the amount of information available for model calibration, likely contributing to their lower predictive performance.

The results obtained for training epochs and cross-validation reinforce the conclusions drawn

from the architectural analyses. Model performance depended less on increasing computational complexity than on efficiently exploiting the available information. Under the conditions evaluated here, approximately 30,000 training epochs combined with 15-fold cross-validation resulted in the highest predictive performance among the configurations evaluated. Collectively, our results show that the predictive performance of MLPs in UAV-based high-throughput phenotyping depends less on maximizing network complexity than on matching model capacity to the amount of biologically relevant information available in the predictor variables. While architectural and training hyperparameters influenced model performance, their effects were consistently constrained by the information content of the multispectral data.

## CONCLUSION

This study demonstrated that the effects of network architecture and training hyperparameters on MLP predictive performance are neither equivalent nor independent. Network width, training duration, and cross-validation strategy were associated with differences in prediction accuracy, whereas increasing network depth provided little practical benefit under the evaluated conditions. Moreover, predictive performance differed among the target traits, with structural traits responding more positively to architectural optimization than complex integrative traits. Optimizing MLP performance for UAV-based soybean phenotyping depends less on increasing network complexity than on considering the relationship between the vegetation indices used as predictors and the target phenotypic traits.

## ACKNOWLEDGMENTS

This work was supported by the National Council for Scientific and Technological Development (CNPq) – grant number 310610/2021-4, the Research Support Foundation of the State of Minas Gerais (FAPEMIG), Federal University of Mato Grosso do Sul and the Postgraduate Program in Genetics and Breeding of the Federal University of Viçosa (UFV). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) – Financial Code 001.

## REFERENCES

- Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., & Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. **Trends in Plant Science**, 23(5), 451-466. <https://doi.org/10.1016/j.tplants.2018.02.001>
- Cabello-Solorzano, K.; Ortigosa de Araujo, I.; Peña, M.; Correia, L. & Tallón-Ballesteros, A.J. (2023). The impact of data normalization on the accuracy of machine learning algorithms: A comparative analysis. In: García Bringas, P.; Pérez García, H.; de Pisón, F. J. M.; Álvarez, F. M.; Lora, A. T.; Herrero, A.; Rolle, J. L. C.; Quintián, H.; Corchado, E. (eds.). 18th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2023). Lecture Notes in Networks and Systems. Vol. 750. **Springer**, Cham. [https://doi.org/10.1007/978-3-031-42536-3\\_33](https://doi.org/10.1007/978-3-031-42536-3_33)
- El-hassani, F. Z.; Amri, M.; Joudar, N. E. & Haddouch, K. (2024). A new optimization model for MLP hyperparameter tuning: modeling and resolution by real-coded genetic algorithm. **Neural processing letters**, 56(2), 105. <https://doi.org/10.1007/s11063-024-11578-0>
- Fodor, N.; Challinor, A.; Droutsas, I.; Ramirez-Villegas, J.; Zabel, F.; Koehler, A.-K. & Foyer, C.H. (2017). Integrating plant science and crop modeling: Assessment of the impact of climate change on soybean and maize production. **Plant Cell Physiology** 58(11): 1833-1847. <https://doi.org/10.1093/pcp/pcx141>
- Food and Agriculture Organization of the United Nations – FAO. (2025). **FAOSTAT**. Available at: <https://www.fao.org/faostat/>
- Fritsch, S.; Guenther, F; Wright, M. N. Neuralnet: Training of Neural Networks. R package version 1.44.6, 2019. <https://doi.org/10.32614/CRAN.package.neuralnet>
- Gitelson, A.A.; Kaufman, Y.J. & Merzlyak, M.N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote Sensing of Environment** 58(3): 289-298. [https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7)
- Gill, T.; Gill, S. K.; Saini, D. K.; Chopra, Y.; De koff, J. P. (2022). A Comprehensive Review of High Throughput Phenotyping and Machine Learning for Plant Stress Phenotyping. **Phenomics**, 2(3), 156-183. <https://doi.org/10.1007/s43657-022-00048-z>
- Hickey, L.T.; Hafeez, A.N.; Robinson, H.; Jackson, S.A.; Leal-Bertioli, S.C.M.; Tester, M.; Gao, C.; Godwin, I.D.; Hayes, B.J. & Wulff, B.B.H. (2019). Breeding crops to feed 10 billion. **Nature Biotechnology** 37(7): 744-754. <https://doi.org/10.1038/s41587-019-0152-9>
- Kruse, R.; Borgelt, C.; Braune, C.; Mostaghim, S. & Steinbrecher, M. (2022). **Computational Intelligence: A Methodological Introduction**. Springer. <https://doi.org/10.1007/978-3-030-42227-1>
- Li, F.; Miao, Y.; Feng, G.; Yuan, F.; Yue, S.; Gao, X.; Liu, Y.; Liu, B.; Ustin, S.L. & Chen, X. (2014). Improving estimation of summer maize nitrogen status with red edge-based spectral vegetation indices. **Field Crops Research** 157: 111-123. <https://doi.org/10.1016/j.fcr.2013.12.018>
- Liu, H.Q. & Huete, A.A. (1995). A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. **IEEE Transactions on Geoscience and Remote Sensing** 33(2): 457-465. <https://doi.org/10.1109/TGRS.1995.8746027>
- Modgil, R.; Tanwar, B.; Goyal, A. & Kumar, V. (2021). Soybean (Glycine max). In: Tanwar, B. & Goyal, A. (eds.). **Oilseeds: Health Attributes and Food Applications**. Singapore: Springer. [https://doi.org/10.1007/978-981-15-4194-0\\_1](https://doi.org/10.1007/978-981-15-4194-0_1)
- R Core Team (2026). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.R-project.org/>
- Raut, P. & Dani, A. (2020). **Correlation between number of hidden layers and accuracy of artificial neural network**. In: Vasudevan, H.; Michalas, A.; Shekokar, N. & Narvekar, M. (eds.). **Advanced Computing Technologies and Applications. Algorithms for Intelligent Systems**. Singapore: Springer. [https://doi.org/10.1007/978-981-15-3242-9\\_49](https://doi.org/10.1007/978-981-15-3242-9_49)

- Santana, D.C.; Oliveira Cunha, M.P.; 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(1): 13. <https://doi.org/10.1186/s13007-022-00848-4>
- Song, P.; Wang, J.; Guo, X.; Yang, W. & Zhao, C. (2021). High-throughput phenotyping: Breaking through the bottleneck in future crop breeding. **The Crop Journal** 9(3): 633-645. <https://doi.org/10.1016/j.cj.2021.03.015>
- Sumner, Z.; Varco, J.J.; Dhillon, J.S.; Fox, A.A.; Czarnecki, J. & Henry, W.B. (2021). Ground versus aerial canopy reflectance of corn: Red-edge and non-red edge vegetation indices. **Agronomy Journal** 113: 2773-2788. <https://doi.org/10.1002/agj2.20657>
- Tayade, R.; Yoon, J.; Lay, L.; Khan, A.L.; Yoon, Y. & Kim, Y. (2022). Utilization of spectral indices for high-throughput phenotyping. **Plants** 11: 1712. <https://doi.org/10.3390/plants11131712>
- Teodoro, P.E.; Teodoro, L.P.; Baio, F.H.; Silva Junior, C.A.; Santana, D.C. & Bhering, L.L. (2024). High-throughput phenotyping in maize and soybean genotypes using vegetation indices and computational intelligence. **Plant Methods** 20(1): 164. <https://doi.org/10.1186/s13007-024-01294-0>
- Vieira Netto, J.A.F.; Panichi, H.P.; Teodoro, P.E. & Bhering, L.L. (2026). Machine learning for unmanned aerial vehicles-based soybean phenotyping: Limits of cross-environment transfer and opportunities to reduce field measurements. **Pesquisa Agropecuária Tropical** 56: e84508. <https://doi.org/10.1590/1983-40632026v5684508>
- Vujicic, T.; Matijevic, T.; Ljucovic, J.; Balota, A. & Sevarac, Z. (2016). Comparative analysis of methods for determining number of hidden neurons in artificial neural network. In: **Central European Conference on Information and Intelligent Systems**. Varazdin: Faculty of Organization and Informatics. p. 219.
- Xue, J. & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. **Journal of Sensors** 1-17. <https://doi.org/10.1155/2017/1353691>
- Yang, W.; Feng, H.; Zhang, X.; Zhang, J.; Doonan, J.H.; Batchelor, W.D.; Xiong, L. & Yan, J. (2020). Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. **Molecular Plant** 13: 187-214. <https://doi.org/10.1016/j.molp.2020.01.008>