ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20241002104404

Original scientific paper
Received: 2024-10-02, Accepted: 2024-10-16

# The Multicollinearity Effect on Machine Learning Accuracy for Leaf Chlorophyll Content Prediction of Indoor Plants

Dorijan Radočaj*, Daria Galić Subašić, Ivan Plaščak, Mladen Jurišić

**Abstract:** Various factors that influence chlorophyll levels in indoor plants were analysed. The input dataset consisted of 52 samples, which represent 10 distinct plant types. To non-invasively measure the chlorophyll content in plant leaves, Soil-Plant Analysis Development (SPAD) sensor was used, measuring the absorbance of specific light wavelengths, allowing for the assessment of chlorophyll concentration. The dataset was supplemented by covariates from soil electrical conductivity (*EC*) sensing at depths of 5 cm, 10 cm, and 15 cm, along with the multispectral Plant-O-Meter sensor. Four covariates in the model, including plant type, *EC* (5 cm), *EC* (15 cm), and normalized difference red-edge index (*NDRE*), showed minimal correlation with other variables, highlighting their independence. To predict leaf chlorophyll content, Random Forest and Extreme Gradient Boosting machine learning models were utilized, with Random Forest achieving higher average coefficient of determination of 0.458. The study underscored the potential of a complementary dataset for evaluating the complex relationship among root-soil dynamics and leaf for optimizing indoor plant health.

**Keywords:** extreme gradient boosting; multispectral sensor; random forest; soil electroconductivity; Soil-Plant Analysis Development (SPAD)

## 1 INTRODUCTION

In recent years, there has been increasing interest in cultivating indoor plants effectively and sustainably, driven by recognition of their manifold benefits, which include positively impacting the environment, aesthetics, and human health [1]. These plants are not merely decorative additions to our living spaces, rather they have a crucial role in enhancing air quality, reducing stress, and beautifying indoor environment. The presence of chlorophyll in the leaves of indoor plants is fundamental to their health and vitality, as this pigment is essential for photosynthesis and overall growth [2]. Advanced machine learning algorithms have emerged as powerful tools in optimizing the care of indoor plants, enabling precise monitoring and fine-tuning of chlorophyll care plans. These algorithms, along with non-invasive estimation instruments, provide a method to comprehend and improve the well-being of our indoor plants [3]. Predicting the quantity of chlorophyll in indoor plants using machine learning algorithms poses challenges, including multicollinearity that occurs when predictor variables in the model have high correlations, leading to an impact on the accuracy and interpretability of the machine learning model [4]. Multicollinearity can introduce instability into coefficient estimations, making it difficult to discern the unique contributions of each predictor variable. In the context of predicting chlorophyll concentration, various physiological and environmental factors are at play, and their interrelationships can lead to multicollinearity, casting doubt on the validity of machine learning models. To tackle this issue, two significant sources of data play a vital role in machine-learning predictions of chlorophyll content in leaves: soil electrical conductivity (*EC*) sensing and leaf multispectral sensing [5]. Multispectral sensing of leaves offers a vast amount of information on the physiological state of the leaves, providing insights into pigment content, water stress, and the overall health of the plant. Soil *EC* sensing provides critical information about soil moisture levels and nutrient availability, both of which have a direct impact on plant growth and vitality. Integrating this data with chlorophyll predictions through machine learning models is expected to result in improved accuracy. In the context of predicting chlorophyll concentration in indoor plants, Monte Carlo cross-validation accounts for the inherent variability in factors such as plant growth, environmental conditions, and sensor data [6]. The care of indoor plants is influenced by numerous dynamic factors, and this approach ensures that the models can withstand the randomness in the training and test data. Additionally, it decreases the probability of overfitting to a particular dataset and encourages the creation of models that are robust.

To improve present knowledge of predicting indoor plant chlorophyll content, this study delves into the significant issue of multicollinearity using machine learning techniques, emphasizing the Random Forest (RF) and Extreme Gradient Boosting (XGB) methods. Monte Carlo cross-validation was employed to assess and mitigate the impact of multicollinearity on the forecasting accuracy of these two extensively adopted techniques. The consequences of multicollinearity on the effectiveness of both RF and XGB models were explored, alongside their predictive capabilities, and their reliability in accurately estimating chlorophyll content in leaves. Additionally, it will enhance the understanding of the challenges posed by multicollinearity. The aim of the evaluation of these models' performance is to assist practitioners in determining the most effective approach to handle multicollinearity.

## 2 MATERIALS AND METHODS

### 2.1 Indoor Plant Data Collection

Ten indoor plant types represented by 52 samples were evaluated in the study (Fig. 1). The evaluation of chlorophyll levels in indoor plant leaves was achieved using the portable Konica Minolta Soil-Plant Analysis Development (SPAD) sensor, in a non-invasive manner. This technology enables

measuring the absorbance of specific light wavelengths by the chlorophyll in leaf tissue, thereby ascertaining the relative chlorophyll concentration [7]. The SPAD sensor measures light absorbance, focusing on the decrease in red light absorption caused by chlorophyll's presence, resulting in a unitless SPAD reading that shows the leaf's relative chlorophyll concentration. Elevated SPAD levels are linked to higher chlorophyll content, indicating healthier and more photosynthetically active leaves. The gathered SPAD data enables comparing leaves or plants, facilitating evaluation of general plant health, monitoring changes over time, and informing care plans for indoor plant maintenance.

When combined, soil EC sensing at three soil depths (5, 10, and 15 cm), and the multispectral Plant-O-Meter sensor provide a dependable approach for obtaining complementary covariates for predicting chlorophyll content of indoor plants [5]. The Plant-O-Meter sensor was used to measure leaf physiology aspects like plant leaves' absorption and reflection characteristics at multiple wavelengths. Vegetation indices were calculated based on Plant-O-Meter measurements according to specifications by Kitić et al. [8], including red (*NDVIr*), green (*NDVIg*), blue (*NDVIb*), green-red (*GRNDVI*), green-blue (*GBNDVI*), red-blue (*RBNDVI*) and visible normalized difference vegetation index (*PNDVI*), simple ratio (*SR*) with its modification (*MSR*) and inversion (*ISR*), normalized difference red-edge index (*NDRE*) and enhanced vegetation index (*EVI*). Soil *EC* sensing at various depths enabled measuring the effects of soil moisture content. A total of 16 covariates were evaluated, including plant type, three soil *EC* covariates and 12 vegetation indices.

## 2.2 Multicollinearity Analysis

Three diagnostic measures were used in analysing multicollinearity, including tolerance (*TOL*), within-subject variance (*Wi*), and auxiliary F-test (Fi) [9]. *TOL* measured the percentage of variance in a predictor variable that isn't explained by other predictor variables in the model, with low tolerance levels indicating significant multicollinearity. Tolerance values ranged from 0 to 1, representing reciprocal values of the variance inflation factor. *TOL* values close to 1 suggest that it is relatively uncorrelated with other predictors, with values higher than 0.1 indicating multicollinearity. This indication implies that the variable can provide specific and unique information to the model. Farrar's F-test is a diagnostic metric for assessing multicollinearity, represented by *Wi*. This test examined the correlations between a specific predictor variable and all other variables in the model, indicating the presence and severity of multicollinearity. *Wi* demonstrated significant results when multicollinearity exists, indicating the dependent variable's linear predictability from the independent predictors. In the context of multicollinearity, the *Fi* assessed the relationship between the F-statistic and the coefficient of determination, allowing determination of whether high levels of multicollinearity decrease the overall goodness of fit of the regression model and how it affects the model's explanatory power.


**Figure 1** The display of representative indoor plants from ten species evaluated in the study

## 2.3 Machine Learning Prediction and Accuracy Assessment

The study employed the RF model to capture non-linear correlations between leaf chlorophyll concentration and chosen characteristics. The ensemble approach included several decision trees, which reduced overfitting and improved prediction accuracy [10]. In addition, the enhanced gradient boosting technique XGB was used to predict chlorophyll levels. This model is highly renowned for its durability and efficiency in processing high-dimensional information [11]. A feature selection procedure was conducted to enhance the predictive capacity of the models, combining RF and XGB models with all input covariates and with covariates for which multicollinearity was not detected based on *TOL*, *Wi* and *Fi* coefficients.

The effectiveness of the RF and XGB regression models in predicting leaf chlorophyll content was evaluated through the Monte Carlo cross-validation approach. To ensure robustness in the evaluation, the dataset per 20 repetitions was divided into an 80% training set and a 20% testing set. This repeated resampling strategy produced multiple distinct

train-test splits, enabling a thorough assessment of model performance. During each iteration, the RF and XGB models underwent training with the training data and evaluated their predictive performance using the testing data.

The coefficient of determination ($R^2$) indicated the proportion of the variance in the leaf chlorophyll content represented by *SPAD* values that the independent variables can predict. Root mean square error (*RMSE*) measured the average magnitude of the difference between the predicted and actual values. Mean absolute error (*MAE*) similarly measured the average absolute difference between predicted and actual values, being less susceptible to higher residuals between predicted and actual *SPAD* values. A higher $R^2$ and lower *RMSE* and *MAE* indicated more accurate predictive performance.

## 3 RESULTS AND DISCUSSION

Plant type, *EC* (5 cm), *EC* (15 cm), and *NDRE* exhibited limited correlation with other variables (Tab. 1). It is expected that plant type, being a categorical variable representing distinct types of plants, does not inherently exhibit multicollinearity with numerical variables. *EC* (5 cm) and *EC* (15 cm) represented distinct soil depths, and due to the sizable variance in soil properties at different depths, it is generally unlikely that they will exhibit a strong linear correlation. Moreover, *NDRE*, which is obtained from multispectral leaf sensing, characterized plant health in a unique manner based on red-edge spectral band and thus did not exhibit a strong correlation with other environmental variables or indices [12]. This lack of correlation between these particular variables indicates that they can be studied separately to determine their impact on the response variables, with minimal preoccupation for the interfering effects of collinearity.

**Table 1** The results of multicollinearity analysis for evaluated covariates.

| Covariates | *TOL* | *Wi* | *Fi* |
|---|---|---|---|
| Plant type | 0.311 | 2.495 | 2.700 |
| EC (5 cm) | 0.188 | 4.875 | 5.276 |
| EC (10 cm) | 0.067 | 15.666 | 16.953 |
| EC (15 cm) | 0.108 | 9.329 | 10.096 |
| EVI | 0.049 | 21.979 | 23.784 |
| GBNDVI | 0.000 | 3221.493 | 3486.060 |
| GRNDVI | 0.000 | 4352.326 | 4709.763 |
| ISR | 0.001 | 1868.367 | 2021.808 |
| MSR | 0.021 | 53.424 | 57.812 |
| NDRE | 0.301 | 2.618 | 2.833 |
| NDVIb | 0.002 | 639.444 | 691.959 |
| NDVIg | 0.000 | 6788.285 | 7345.777 |
| NDVIr | 0.000 | 2802.944 | 3033.137 |
| PNDVI | 0.000 | 5560.212 | 6016.848 |
| RBNDVI | 0.000 | 3395.957 | 3674.852 |
| SR | 0.011 | 105.751 | 114.436 |

A series of boxplots in Fig. 2 display the *SPAD* values of evaluated indoor plant species according to four covariates for which multicollinearity was not detected (Fig. 3), including plant type, soil *EC* at 5 cm and 15 cm soil depth, and *NDRE*. All four covariates indicate large heterogeneity of analysed individual indoor plants, which is otherwise not usually present in the same intensity for outdoor plants, especially crops and orchards [13].

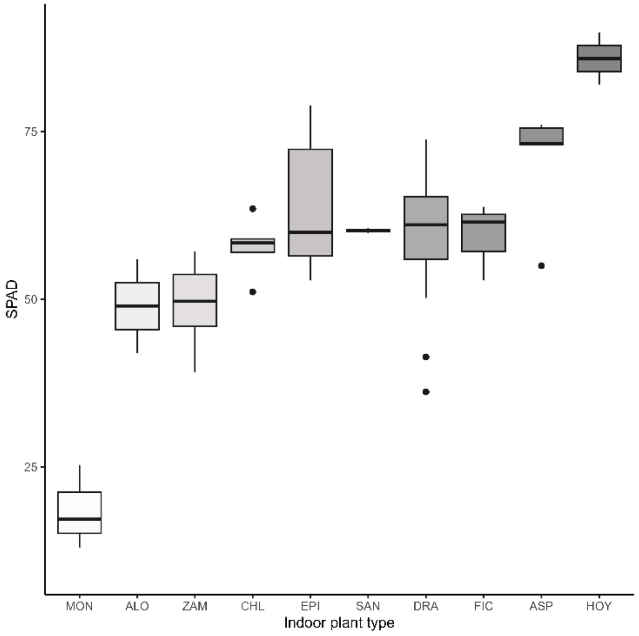


**Figure 2** Boxplots representing SPAD values per indoor plant type

Based on the Monte Carlo accuracy assessment, the RF model using all covariates achieved superior overall performance in 20 distinct data folds (Tab. 2). Specifically, the data produced a substantially greater average $R^2$ of 0.458, indicating its capacity to elucidate a larger portion of the dependent variable's variability, along with a lower average *RMSE* of 10.69 and *MAE* of 8.16. On the other hand, the RF model with filtered covariates for multicollinearity exhibited subpar relative performance with a lower average $R^2$ of 0.338 and higher average *RMSE* of 11.77 and *MAE* of 9.12, demonstrating a decrease in predictive accuracy. Additionally, the coefficient of variation indicated that the RF model demonstrated enhanced stability and consistency in performance across various data partitions, therefore presenting itself as a more dependable option within this experimental context.

The XGB model without multicollinearity filtering consistently demonstrated moderate performance across distinct data folds (Tab. 3). It achieved an average $R^2$ of 0.360, displaying moderately precise predictions, with an average *RMSE* of 13.50 and an average *MAE* of 10.77. Conversely, the XGB model with integrated multicollinearity filtering presented marginally improved performance, producing an average $R^2$ of 0.423 and a lower average *RMSE* of 13.28 and *MAE* of 10.45. These results suggest that the multicollinearity analysis had limited impact on predictive accuracy improvement in general, with only XGB having a slight benefit from it [14]. Additionally, the coefficient of variation values indicated that both models exhibited significant variability in performance across various data partitions, with the XGB model displaying slightly higher variability.

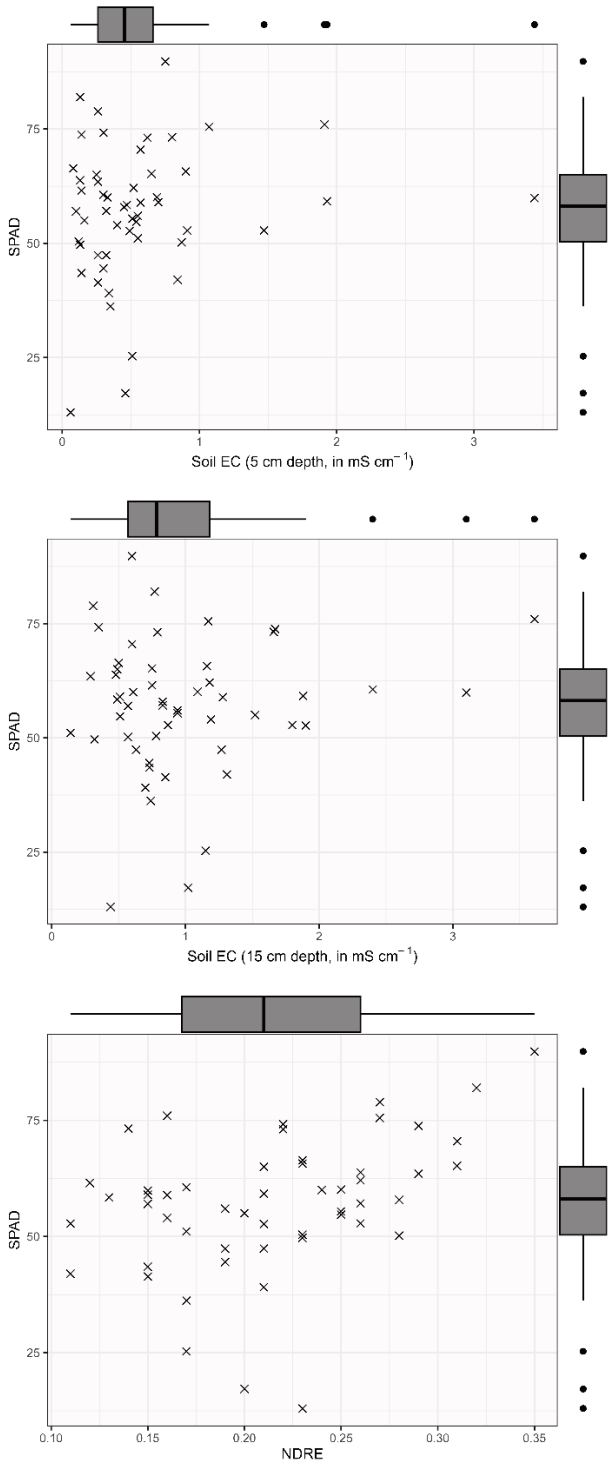


**Figure 3** Boxplots representing SPAD values and their relationship with covariates for which multicollinearity was not detected

By combining multispectral leaf sensing with soil *EC* data at various depths, a comprehensive dataset of covariates was generated. This facilitated the capturing of complex interactions between the below-ground root-soil interactions and above-ground leaf physiology by machine learning models, resulting in enhanced precision and resilience of predicting leaf chlorophyll content. Monte Carlo cross-validation aided in identifying the most effective algorithm for predicting chlorophyll levels in plant foliage by mitigating the impact of individual data subsets and allowing for a comprehensive evaluation of the models' capacity for generalization. The Monte Carlo cross-validation results exhibited notable differences in prediction accuracy according to properties of training and test sample data, strongly indicating that frequently used single split-sample approach might not provide adequate representation of prediction accuracy.

**Table 2** The accuracy assessment results for RF

| Fold | All covariates | | | Filtered covariates | | |
|---|---|---|---|---|---|---|
| | $R^2$ | *RMSE* | *MAE* | $R^2$ | *RMSE* | *MAE* |
| 1 | 0.435 | 15.30 | 9.49 | 0.110 | 11.34 | 7.94 |
| 2 | 0.395 | 15.00 | 9.84 | 0.534 | 7.65 | 6.27 |
| 3 | 0.428 | 8.64 | 7.36 | 0.445 | 9.90 | 9.14 |
| 4 | 0.659 | 7.32 | 6.03 | 0.205 | 12.03 | 10.91 |
| 5 | 0.704 | 12.89 | 10.26 | 0.558 | 8.58 | 7.15 |
| 6 | 0.639 | 15.74 | 11.60 | 0.575 | 9.59 | 7.72 |
| 7 | 0.213 | 10.47 | 8.88 | 0.441 | 9.92 | 8.44 |
| 8 | 0.430 | 9.33 | 8.02 | 0.000 | 20.07 | 14.43 |
| 9 | 0.748 | 10.62 | 7.60 | 0.147 | 11.75 | 8.75 |
| 10 | 0.889 | 5.67 | 4.41 | 0.179 | 19.19 | 12.82 |
| 11 | 0.497 | 11.39 | 9.58 | 0.347 | 18.35 | 12.36 |
| 12 | 0.032 | 11.20 | 8.56 | 0.141 | 10.23 | 6.06 |
| 13 | 0.270 | 8.65 | 6.14 | 0.828 | 10.20 | 7.83 |
| 14 | 0.584 | 7.31 | 6.29 | 0.654 | 7.77 | 5.83 |
| 15 | 0.027 | 14.65 | 9.15 | 0.018 | 13.49 | 11.24 |
| 16 | 0.018 | 15.97 | 10.18 | 0.083 | 11.44 | 9.74 |
| 17 | 0.880 | 6.74 | 6.02 | 0.632 | 5.74 | 4.18 |
| 18 | 0.322 | 10.63 | 9.15 | 0.592 | 11.56 | 9.60 |
| 19 | 0.787 | 5.95 | 5.51 | 0.278 | 9.92 | 8.35 |
| 20 | 0.208 | 10.43 | 9.07 | 0.000 | 16.72 | 13.61 |
| Average | 0.458 | 10.69 | 8.16 | 0.338 | 11.77 | 9.12 |
| CV | 0.603 | 0.311 | 0.234 | 0.751 | 0.334 | 0.302 |

**Table 3** The accuracy assessment results for XGB

| Fold | All covariates | | | Filtered covariates | | |
|---|---|---|---|---|---|---|
| | $R^2$ | *RMSE* | *MAE* | $R^2$ | *RMSE* | *MAE* |
| 1 | 0.731 | 11.07 | 8.30 | 0.545 | 11.33 | 9.04 |
| 2 | 0.291 | 14.59 | 9.20 | 0.271 | 19.06 | 15.82 |
| 3 | 0.492 | 16.24 | 11.29 | 0.676 | 7.67 | 6.11 |
| 4 | 0.298 | 12.53 | 11.23 | 0.399 | 15.17 | 14.26 |
| 5 | 0.034 | 15.14 | 11.99 | 0.517 | 11.42 | 9.87 |
| 6 | 0.625 | 10.28 | 8.68 | 0.465 | 11.96 | 9.42 |
| 7 | 0.456 | 13.46 | 10.83 | 0.527 | 9.47 | 7.60 |
| 8 | 0.272 | 10.47 | 7.86 | 0.067 | 12.81 | 10.63 |
| 9 | 0.598 | 8.93 | 6.80 | 0.296 | 16.70 | 11.84 |
| 10 | 0.252 | 12.66 | 10.15 | 0.410 | 14.06 | 11.91 |
| 11 | 0.002 | 25.53 | 21.10 | 0.225 | 11.01 | 8.41 |
| 12 | 0.243 | 13.64 | 11.16 | 0.526 | 13.90 | 9.15 |
| 13 | 0.553 | 11.10 | 9.38 | 0.342 | 13.36 | 10.45 |
| 14 | 0.517 | 8.60 | 7.27 | 0.679 | 11.91 | 9.47 |
| 15 | 0.074 | 12.89 | 11.17 | 0.586 | 12.81 | 11.04 |
| 16 | 0.580 | 13.28 | 10.36 | 0.691 | 12.04 | 7.87 |
| 17 | 0.444 | 6.82 | 5.38 | 0.578 | 11.30 | 9.54 |
| 18 | 0.001 | 25.65 | 21.41 | 0.280 | 18.10 | 13.06 |
| 19 | 0.677 | 14.29 | 10.58 | 0.320 | 14.73 | 11.84 |
| 20 | 0.056 | 12.84 | 11.24 | 0.064 | 16.85 | 11.56 |
| Average | 0.360 | 13.50 | 10.77 | 0.423 | 13.28 | 10.45 |
| CV | 0.665 | 0.351 | 0.371 | 0.444 | 0.215 | 0.222 |

To further improve the proposed approach, the addition of deep learning methods [11] and integration of additional sensors complementary to multispectral and soil *EC* devices,

such as insolation sensor [15], might provide more in-depth observations as an upgrade of this study.

## 4 CONCLUSIONS

A thorough analysis was carried out to investigate the various factors impacting chlorophyll levels in indoor plants. The study involved examining a dataset encompassing ten unique indoor plant species. Four factors, namely Plant Type, *EC* (5 cm), *EC* (15 cm), and *NDRE*, were singled out as displaying minimal correlation with other variables, indicating their independence within the model. As a categorical variable denoting individual plant species, plant type e does not exhibit multicollinearity with numerical variables. The soil electrical conductivity measurements at varying depths displayed significant diversity, while *NDRE*, obtained from multispectral leaf sensing, rendered a distinctive outlook on plant health, resulting in minimal correlation with other environmental variables or indices. Monte Carlo cross-validation was utilized to comprehensively evaluate the effectiveness of RF and XGB machine learning models in predicting leaf chlorophyll content. The RF model, which included all covariates, demonstrated superior overall performance by elucidating a significant portion of the dependent variable's variance. While the XGB model displayed moderately accurate predictions, it only slightly benefited from multicollinearity filtering. The RF model proved more stable and consistent in comparison, achieving average $R^2$ of 0.458. The Monte Carlo cross-validation results exhibited notable differences in prediction accuracy according to properties of training and test sample data, strongly indicating that frequently used single split-sample approach might not provide adequate representation of prediction accuracy.

This study indicated that a combination of multispectral leaf sensing and soil *EC* measurements at varying depths can establish a powerful tool for exploring the intricate relationship between root-soil dynamics below ground and leaf physiology above ground. Through the use of machine learning models and cross-validation techniques, moderately accurate predictive capabilities for indoor plant chlorophyll content were achieved. Future research could include the integration of complementary sensors and deep learning methods to advance present comprehension of indoor plant health and chlorophyll synthesis.

## 5 REFERENCES


[1] Yeo, L. B. (2021). Psychological and physiological benefits of plants in the indoor environment: A mini and in-depth review. *International Journal of Built Environment and Sustainability*, *8*(1), 57-67. https://doi.org/10.11113/ijbes.v8.n1.597

[2] AlFadhly, N. K., Alhelfi, N., Altemimi, A. B., Verma, D. K. & Cacciola, F. (2022). Tendencies affecting the growth and cultivation of genus Spirulina: An investigative review on current trends. *Plants*, *11*(22), 3063. https://doi.org/10.3390/plants11223063

[3] Yang, Y., Xu, F., Chen, J., Tao, C., Li, Y., Chen, Q., ... & Shen, W. (2023). Artificial intelligence-assisted smartphone-based sensing for bioanalytical applications: A review. *Biosensors and Bioelectronics*, 115233. https://doi.org/10.1016/j.bios.2023.115233

[4] Drobnič, F., Kos, A. & Pustišek, M. (2020). On the interpretability of machine learning models and experimental feature selection in case of multicollinear data. *Electronics*, *9*(5), 761. https://doi.org/10.3390/electronics9050761

[5] Khruschev, S. S., Plyusnina, T. Y., Antal, T. K., Pogosyan, S. I., Riznichenko, G. Y. & Rubin, A. B. (2022). Machine learning methods for assessing photosynthetic activity: environmental monitoring applications. *Biophysical Reviews*, *14*(4), 821-842. https://doi.org/10.1007/s12551-022-00982-2

[6] Phinzi, K., Abriha, D. & Szabó, S. (2021). Classification efficacy using k-fold cross-validation and bootstrapping resampling techniques on the example of mapping complex gully systems. *Remote Sensing*, *13*(15), 2980. https://doi.org/10.3390/rs13152980

[7] Křížová, K., Kadeřábek, J., Novák, V., Linda, R., Kurešová, G. & Šařec, P. (2022). Using a single-board computer as a low-cost instrument for SPAD value estimation through colour images and chlorophyll-related spectral indices. *Ecological Informatics*, *67*, 101496. https://doi.org/10.1016/j.ecoinf.2021.101496

[8] Kitić, G., Tagarakis, A., Cselyuszka, N., Panić, M., Birgermajer, S., Sakulski, D. & Matović, J. (2019). A new low-cost portable multispectral optical device for precise plant status assessment. *Computers and Electronics in Agriculture*, *162*, 300-308. https://doi.org/10.1016/j.compag.2019.04.021

[9] Corrêa, A. J. M., Alves, P. F., Cambuim, J., de Moraes, M. L. T. & Freitas, M. L. M. (2022). Climate fluctuation impacts in Astronium urundeuva (M. Allemão) Engl. silvicultural characters in the Brazilian Cerrado. *Environmental Research: Climate*, *1*(2), 025007. https://doi.org/10.1088/2752-5295/ac9695

[10] Radočaj, D., Jurišić, M., Antonić, O., Šiljeg, A., Cukrov, N., Rapčan, I., ... & Gašparović, M. (2022). A Multiscale Cost–Benefit Analysis of Digital Soil Mapping Methods for Sustainable Land Management. *Sustainability*, *14*(19), 12170. https://doi.org/10.3390/su141912170

[11] Hassan, S. M., Jasinski, M., Leonowicz, Z., Jasinska, E. & Maji, A. K. (2021). Plant disease identification using shallow convolutional neural network. *Agronomy*, *11*(12), 2388. https://doi.org/10.3390/agronomy11122388

[12] Prey, L., Von Bloh, M. & Schmidhalter, U. (2018). Evaluating RGB imaging and multispectral active and hyperspectral passive sensing for assessing early plant vigor in winter wheat. *Sensors*, *18*(9), 2931. https://doi.org/10.3390/s18092931

[13] Radočaj, D., Šiljeg, A., Plaščak, I., Marić, I. & Jurišić, M. (2023). A Micro-Scale Approach for Cropland Suitability Assessment of Permanent Crops Using Machine Learning and a Low-Cost UAV. *Agronomy*, *13*(2), 362. https://doi.org/10.3390/agronomy13020362

[14] Yan, H., He, Z., Gao, C., Xie, M., Sheng, H. & Chen, H. (2022). Investment estimation of prefabricated concrete buildings based on XGBoost machine learning algorithm. *Advanced Engineering Informatics*, *54*, 101789. https://doi.org/10.1016/j.aei.2022.101789

[15] Jung, C. & Arar, M. (2023). Natural vs. Artificial Light: A Study on the Influence of Light Source on Chlorophyll Content and Photosynthetic Rates on Indoor Plants. *Buildings*, *13*(6), 1482. https://doi.org/10.3390/buildings13061482



**Authors' contacts:**

**Dorijan Radočaj,** PhD
(Corresponding author)
Josip Juraj Strossmayer University of Osijek,
Faculty of Agrobiotechnical Sciences Osijek,
Vladimira Preloga 1, 31000 Osijek, Croatia
E-mail: dradocaj@fazos.hr

**Daria Galić Subašić,** PhD
Josip Juraj Strossmayer University of Osijek,
Faculty of Agrobiotechnical Sciences Osijek,
Vladimira Preloga 1, 31000 Osijek, Croatia
E-mail: dgsubasic@fazos.hr

**Ivan Plaščak,** PhD, Associate professor
Josip Juraj Strossmayer University of Osijek,
Faculty of Agrobiotechnical Sciences Osijek,
Vladimira Preloga 1, 31000 Osijek, Croatia
E-mail: iplascak@fazos.hr

**Mladen Jurišić,** PhD, Full professor
Josip Juraj Strossmayer University of Osijek,
Faculty of Agrobiotechnical Sciences Osijek,
Vladimira Preloga 1, 31000 Osijek, Croatia
E-mail: mjurisic@fazos.hr