USE OF DIGITAL IMAGES TO CLASSIFY LEAF PHOSPHORUS CONTENT IN GRAPE TOMATOES

Leonardo P. de Magalhães¹*, Lucas R. Trevisan¹, Tamara M. Gomes², Fabrício Rossi²

¹*Corresponding author. Universidade de São Paulo/ Piracicaba - SP, Brasil.
Email: leonardo.magalhaes@usp.br | ORCID ID: https://orcid.org/0000-0002-0014-0404

KEYWORDS
Artificial Neural Network, vegetation indices, DGCI, MPRI, Solanum lycopersicum

ABSTRACT
Leaf chemical analysis is one of the ways to assess plant development. However, this type of assessment is expensive and time-consuming. The variation of nutrient content in the leaves modifies the proportion of light reflected and absorbed by plants at different wavelengths. Being able to relate the color reflected by the leaves with their phosphorus (P) content and using this data as input into an artificial neural network (ANN) can be an alternative for its determination. For this, it is necessary to establish which colors are most correlated with the different nutrients. Therefore, the phosphorus content in tomato leaves was evaluated in this study, according to four treatments (0.25, 50, 75, and 100% of the P doses). Different vegetation indices were also evaluated using images of mini-tomato leaves through a principal component analysis to determine which ones would be suitable to serve as an input to an ANN (multilayer perceptron). DGCI (Dark Green Color Index) and Bn (Normalized Blue) were the indices most related to P content. The neural network obtained 90% accuracy in the classification after training using both sides of the leaves.

INTRODUCTION
Phosphorus (P) is one of the most extracted elements by plants among the nutrients needed for plant activities, such as protein production, productivity, and fruit quality (Kumar et al., 2015). Phosphorus plays an important role in the initial development of tomatoes and contributes to increasing the commercial fruit quality (Nowaki et al., 2017). Leaf chemical analysis is one of the most commonly used methods to control the nutritional status of plants, but the response is often time-consuming, hindering decision-making for efficient management (Meiqing et al., 2016). Alternatively, the light reflectance of leaves at different wavelengths can be evaluated. In the case of phosphorus, some authors have shown the useful band regions, which vary between the visible and the Short Wave Infrared (SWIR) (Li et al., 2018). The visible region is related to stress due to P deficiency, highlighting the bands in the blue, red, and red edge regions (Stein et al., 2014). This stress causes some changes in the leaves, such as the higher presence of anthocyanin and the resulting decrease in chlorophyll (Marschner, 2013).

Thus, the use of computer vision with RGB (Red, Blue, and Green) images associated with artificial neural networks arises as an alternative for evaluating and classifying the nutritional status of plants. The advantages of this method are its lower cost (compared to optical sensors), greater accuracy (compared to visual assessment), and reduced time compared to chemical analysis (Pir, 2016). Several vegetation indices have been described in the literature in recent years, such as NDVI (Normalized Difference Vegetation Index) (Rousse et al., 1973), MPRI (Modified Photochemical Reflectance Index) (Yang et al., 2008), and DGCI (Dark Green Color Index) (Karcher & Richardson, 2003).
Some authors have conducted studies on the subject to highlight the importance of vegetation indices in crop evaluation. Oliveira et al. (2019) evaluated different indices in the relationship between nitrogen doses and yield in tomato crops. Siedliska et al. (2021) used hyperspectral data from three different crops, while Yan Li et al. (2015) and Peng-Tao et al. (2018) evaluated orange and rubber tree leaves, respectively, to classify the amount of phosphorus. The application of neural networks to classify nutrients in leaves has been used in several crops, such as palm (Jayaselan et al., 2018) and tobacco (Backhaus et al., 2011). Furthermore, Barbudo (2019) published a review on this topic, with several studies using images and machine learning.

Therefore, this study aimed to evaluate, through digitalized images of grape tomato leaves, which colors and vegetation indices would have a higher correlation with phosphorus contents in leaves for their classification using an artificial neural network.

MATERIAL AND METHODS

Study area, crop, and substrate

The plants were drip irrigated using an automated system with a flow rate of 8 L h⁻¹, with each dripper tube providing 2 L h⁻¹ per plant. The pH was measured using a strip and the electrical conductivity was obtained using a portable pH meter (waterproof Pen mS/cm Tester), both at each preparation. Initially, irrigation was performed once a day at 9:30 am, with an irrigation pulse of four minutes, totaling an applied volume of 120 mL. The solution prepared in the water tank was made for 400 liters of water, according to Cunha et al. (2014).

The necessary phosphorus doses were applied manually according to the recommendation proposed in the experiment (0, 25, 50, 75, and 100% P₂O₅), with the 100% P₂O₅ dose consisting of 114 g 1000 L⁻¹ up to 22 DAT and 260 g 1000 L⁻¹ until the end of the experiment (64 DAT). The amount applied weekly was calculated as a function of other nutrients applied via irrigation.

Images, vegetation indices, and nutritional diagnosis

Tomato leaf images were taken at 22, 36, 50, and 64 DAT, thus allowing calculating the vegetation indices. The 4th fully expanded leaf from the apex was sampled, as determined by Malavolta et al. (1989), with four leaves being collected from each plot. The leaves from both the upper and the lower faces were scanned with a white stock card in the background on a flatbed scanner, resulting in 300 dpi images saved in TIFF format. A total of 80 samples were collected on each date, that is, four samples per plot, but the value of each index was obtained from the average of the samples per plot. Thus, 20 values of each index were obtained on each date, totaling n = 80 at the end of the collections.

### TABLE 1. Nutritional composition of the coconut fiber-based substrate.

| N  | P₂O₅ | K₂O | Ca | Mg | S   | B   | Cu | Fe | Mn | Mo | Zn |
|----|------|-----|----|----|-----|-----|----|----|----|----|----|
|    |      |     |    |    |     |     |    |    |    |    |    |
| 42 | 48   | 54  | 186| 39 | 37  | 1.09| 4.66| 14.07| 2.98| 1.6| 4.42|

Source: Amafibra³

³ Personal Communication
After the acquisition, the images were inserted in software developed in MATLAB (https://url.gratis/zRt6Cz) to obtain their RGB and HSV component values and calculate the vegetation indices of interest. The indices were MPRI, DGCI, Rn (Normalized Red), Gn (Normalized Green), and Bn (Normalized Blue), calculated from both sides of the leaves.

\[ MPRI = \frac{G-R}{G+R} \]  
\[ DGCI = \left( \frac{H+6}{60} \right) + \left( 1 - S \right) + \left( 1 - V \right) \]  
\[ Rn = \frac{R}{R+G+B} \]  
\[ Gn = \frac{G}{R+G+B} \]  
\[ Bn = \frac{B}{R+G+B} \]  

Where:
- R is the red;
- G is the green;
- B is the blue;
- H is the hue;
- S is the saturation;
- V is the value;
- Rn is the normalized red;
- Gn is the normalized green, and
- Bn is the normalized blue.

The leaves used for image acquisition were stored in a paper bag and oven-dried at 65 °C for 72 hours. Subsequently, the samples were ground and sent for analysis at the Laboratory of Analysis of the Agricultural Sciences sector of FZEA/USP.

Data analysis

Pearson’s correlation was used to assess the colors most related to the phosphorus content in the leaves, according to the following criteria: perfect correlation \((r=1)\), strong correlation \((r>0.75)\), moderate correlation \((r>0.5)\), weak correlation \((r<0.5)\), and non-existent correlation \((r=0)\) (NOGUEIRA et al., 2010). The indices Rn, Gn, and Bn were evaluated at this phase of the analysis. A principal component analysis (PCA) was performed using the statistical software R to determine the vegetation indices that most indicated the P variation in the leaves. The principal component analysis is a multivariate exploratory analysis tool, which allows revealing the existence of anomalous samples, relationships between measured variables, and relationships or groups between samples (Lyra et al., 2010).

The classification of the amount of phosphorus into distinct categories was performed after determining the indices. An artificial neural network (ANN) was developed using the indices as input data. The backpropagation Multilayer Perceptron (MLP) ANN was developed in MATLAB (https://url.gratis/8VLyIa). The intermediate layer had ten neurons, and the Levenberg-Marquardt training function (trainlm), in which the performance evaluation was based on the mean square error (MSE), was used. Two-thirds \((n = 53)\) of the data sample was used for network training, and 20 random samples were selected at the network validation stage, distributed throughout the four evaluation periods (22, 36, 50, and 64 DAT). The hit rate was used as a parameter for evaluating the efficiency of the artificial neural network. Its calculation is performed using the following equation:

\[ T = \frac{Vp}{Vp + Fp + Fn} \]  

Where:
- \(Vp\) is the true positive;
- \(Fp\) is the false positive, and
- \(Fn\) is the false negative. True positives are the ANN output data classified within the category to which the real data belongs, \(Fp\) is the data classified below this category, and \(Fn\) is the data classified as category II, with the real data belonging to category I. The P content categories were established according to Table 2. The same steps were performed with the values of P contents in the leaves obtained in the laboratory.

### TABLE 2. Categories of parameter classification.

| P content (g kg\(^{-1}\)) | Category |
|--------------------------|----------|
| < 4                      | Category I |
| 4–8                      | Category II |

Source: Trani et al. (2015)

RESULTS AND DISCUSSION

The statistical analysis of nutrient contents in the leaves showed that only phosphorus (P) and manganese (Mn) were significant relative to doses (Figure 1). Thus, the evaluation of the change in leaf color can be explained only by variations in phosphorus content.
FIGURE 1. Average phosphorus (P) (left) and manganese (Mn) (right) contents (DAT) in grape tomato leaves relative to P2O5 doses.

Pearson’s correlation (Figure 2) showed that the green (upper and lower) and blue colors (lower) showed moderate correlation with phosphorus contents in the leaf (0.628, 0.594, and -0.514, respectively). A reduction in the phosphorus content in leaves also leads to a decrease in chlorophyll efficiency due to its participation in energy production (ATP from ADP) for photosynthesis (Geiger & Servaites, 1994). Thus, adequate phosphorus contents lead to a higher green reflectance (wavelength of 550nm). This factor explains the higher correlation of normalized green with phosphorus contents. Figure 3 shows that lower phosphorus doses also had lower chlorophyll values, reducing the green color of the leaves. Costa et al. (2019) also observed a quadratic behavior of chlorophyll contents regarding the nitrogen variation in vine leaves.

Mulla (2013) also observed that normalized green is concentrated in absorption areas where there is the action of pigments and chlorophyll (such as anthocyanin), a fact that explains the changes in this color with the variation in P contents. Moreover, the reflection in the blue range in the leaf regions where less phosphorus is found increases due to the higher presence of anthocyanin.

These results coincide with those obtained in the principal component analysis (PCA), using the evaluated vegetation indices as input. The PCA shows that the two principal components explain 96.45% of the data. The largest eigenvectors were DGCI_upper (0.69) and DGCI_lower (0.66) in the first principal component and Bn_upper (-0.51) and Bn_lower (-0.50) in the second principal component. Figure 4 shows the contribution of each index obtained by PCA. Both leaf faces had a high value, indicated by the eigenvectors. Liu et al. (2015) observed that the upper leaf face had a better result to indicate N and P contents in citrus leaves.

According to Karcher & Richardson (2003), the higher the DGCI value, the darker the green tone measured by this index. Therefore, the positive eigenvector of this index indicates that P variation in tomato leaves more intensely alters the green color saturation. Healthy plants, which did not have their chlorophyll values changed by the deficient P contents (Figure 3), have darker green leaves. It occurs in opposition to plants where P decreases, thus increasing the amount of anthocyanins in the leaves, which leads to a higher presence of the blue color relative to green.
It can be observed when considering the Bn index, as it measures the amount of blue band reflected by leaves in contrast to the total RGB. The negative Bn value observed in the second principal component is explained by a decrease in the amount of P and the consequent purple color in the leaves. The purple color has a higher value of reflected blue in its composition, which would explain that a decrease in the amount of P in the leaves causes an increase in the Bn index. Bands reflected by plants in the visible range are more commonly associated with P deficiency, especially in the blue, red, and red-edge regions (Stein et al., 2014). It explains why the Bn index had a high negative correlation with the P data in the analyzed leaves for both Pearson’s correlation and PCA. Samples with lower P contents were distributed in the same quadrant (by PCA) as the Bn_U and Bn_L indices (Figure 5). Those that had higher nutrient contents (mostly samples from 1 to 20) are in the same quadrants as the DGCI_L, Gn_L, and Gn_U indices, which may indicate that a higher presence of P leads to a greener leaf than those with smaller P contents.

Sun et al. (2018) evaluated macronutrient deficiency in rice and observed a predominance of light green color, while dark green leaves could be observed in P-deficient plants, in which case the leaves changed from green to purplish gray. These authors evaluated different vegetation indices and demonstrated that the normalized green index using the first most expanded leaf of rice is among those most related to phosphorus deficiency, whereas the normalized green and MPRI are among the most related in the second leaf. However, only DGCI is related to the phosphorus content when using the third most expanded leaf. The indices cited by these authors are in line with those used in the present study, showing a higher relationship with P contents. Rice plants tend to form anthocyanin and may turn red or purplish in the sheaths (Chen et al., 2014). These authors also found that rice leaves are narrower, erect, and white with dark green spots under P deficiency, which is different from the result found in the present study for tomatoes, in which a higher P content in the leaves had a positive relationship (by PCA) with DGCI, that is, darker leaves.
Therefore, the neural network was trained with these four indices. The neural network had a hit rate (in classifying the samples within the defined categories) of 80.77% in training using 2/3 (53) of the 80 samples. This rate was obtained after 10 interactions of the system with a performance of 0.24 (mean squared error). The hit rate at this phase became higher with the data obtained after 36 DAT. It indicates that an analysis of leaves from this period of development would satisfy P classification in the leaves. The hit rate was 69% at 22 DAT and 76.9% at 36 and 50 DAT. Twenty random samples were selected from the remaining 1/3 of the dataset to validate the neural network. The hit rate at this phase was 90% to classify in the desired categories, with a performance of 1.45 (MSE).

Christensen et al. (2004) worked with a neural network to predict P and N contents in barley leaves using leaf reflectance data instead of vegetation indices and obtained hit rates of 74% (P) and 81% (N). Yanli et al. (2015) used hyperspectral images and obtained better results in the prediction of phosphorus rate using the adaxial face of orange leaves. Jayaselan et al. (2018) used a neural network to classify the nitrogen and potassium contents in palm leaves and obtained an overall hit rate of 85.3% for both elements. Aboukarima et al. (2020) applied a neural network to recognize five varieties of faba bean and obtained a hit rate of 77.5%. The hit rate obtained in the present study is, therefore, at a level close to that obtained by other authors, who used classifiers with neural networks. The P content in the leaves predicted by the ANN relative to the current values (Figure 6) showed a mean percentage error of 64.3% at the validation phase of the neural network.

Figure 7 shows the validation of the neural network. Among these samples, the error ranged from 6 to 413%. Four out of the 20 samples had errors with a variation from the actual value above 100%, three between 50 and 100%, and 13 below 50% (six having a variation below 6%). A larger number of samples for the neural network training phase can contribute to a decrease in the observed error.

About 10% of the ANN error (and not only in obtaining the P content) was classified in a category below the current value. In other words, both results were false positives for deficient nutrition if we consider category II as an ideal range of nutrients. No false negatives were observed, indicating that a sample from category I was within category II, the same occurring at the training phase of the network.
According to Withers et al. (2018), the use of phosphate fertilizers in Brazil is twice the demand of plants. This supply above demand is due to soil characteristics, in addition to the lack of leaf analysis. Thus, obtaining leaf analysis instantly using images and artificial neural networks can assist in faster handling and increase the efficiency in the use of this type of fertilization. The results presented in this study also indicate that the technique can be applied to other nutrients, making it a commercially viable alternative for leaf analysis.

FIGURE 7. Percent error of each sample at the ANN validation phase.

CONCLUSIONS

The analysis of the colors reflected in the visible range showed that the green and blue colors had a moderate correlation with phosphorus (P) content in leaves. This factor was corroborated by the results obtained after performing the principal component analysis with the vegetation indices, indicating that the DGCI index has a positive relationship with P content, while the opposite occurs with the Bn index. The developed neural network had a similar hit rate and, in some cases, superior to other studies with the same purpose of predicting or classifying nutrient contents in plant leaves. It indicates that the neural network, using the mentioned indices as input, can be a way to replace the current analysis in obtaining the P content in leaves. The results also indicate that the use of both leaf sides to measure the nutritional content increases the efficiency of the classification of P content.

ACKNOWLEDGMENTS

To the company Agristar, for the partnership and seed donation; to AMAFIBRA, for the coconut fiber donation. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

REFERENCES

Aboukarima A, El-Maragky M, Elsoury H, Zayed M, Minyawi M (2020) Artificial neural network-based method to identify five varieties of Egyptian Faba Bean according to seed morphological features. Engenharia Agrícola 40(6):791-799. DOI: https://doi.org/10.1590/1809-4430-Eng.Agric.v40n6p791-799/2020.

Backhaus A, Bollenbeck F, Seiffert U (2011) Robust classification of the nutrition state in crop plants by hyperspectral imaging and artificial neural networks. Biosystems Engineering. Fraunhofer Institute for Factory Operation and Automation (IFF) Sandtorstr 22: 39108.

Barbedo JGA (2019) Detection of nutrition deficiencies in plants using proximal images and machine learning: A review. Computers And Electronics In Agriculture 162:482-492. DOI: http://dx.doi.org/10.1016/j.compag.2019.04.035.

Chen LS, Lin L, Cai G, Sun Y, Huang T, Wang K, Deng J (2014). Identification of Nitrogen, Phosphorus, and Potassium deficiencies in rice based on static scanning technology and hierarchical identification method. PLoS ONE 9(11):e113200.

Christensen LK, Bennedsen BS, Jørgensen RN, Nielsen H (2004). Modelling Nitrogen and Phosphorus Content at Early Growth Stages in Spring Barley using Hyperspectral Line Scanning. Biosystems Engineering 88(1):19–24.

Costa BRS, Oldoni H, Silva WA, Martins RL, Bassoi LH (2019). Temporal variation and spatial distribution of relative indices of leaf chlorophyll in grapevine cv. Chardonnay. Engenharia Agrícola 39(n. spe):74-84. DOI: https://doi.org/10.1590/1809-4430-Eng.Agric.v39nep74-84/2019.

Cunha AHN, Sandri D, Vieira JA, Cortez TB, Oliveira TH (2014) Sweet grape mini tomato grown in culture substrates and effluent with nutrient complementation. Engenharia Agrícola 34(4):707-715.
Geiger DR, Servaites JC (1994) Diurnal regulation of photosynthetic carbon metabolism in C3 plants. Annual Review of Plant Physiology and Plant Molecular Biology 45:235–256.

Jayaselan HAJ, Nawi NM, Ismail WIW, Meh dizadeh SA, Shar iff ARM (2018) Application of artificial neural network classification to determine nutrient content in oil palm leaves. Applied Engineering in Agriculture 34(3).

Karcher DE, Richardson MD (2003) Quantifying turfgrass color using digital image analysis. Crop Science 43:943-951.

Kumar V, Singh TR, Hada A, Jolly M, Ganapathi A, Sachdev A (2015) Probing phosphorus efficient low phytic acid content soybean genotypes with phosphorus starvation in hydroponics growth system. Applied Biochemistry and Biotechnology 177:689-699.

Li D et al. (2018) Monitoring litchi canopy foliar phosphorus content using hyperspectral data. Computers And Electronics In Agriculture 154:176-186. DOI: http://dx.doi.org/10.1016/j.compag.2018.09.007.

Liu Y, Lyu Q, He S, Yi S, Liu X, Xie R, Zheng Y, Deng L (2015) Prediction of nitrogen and phosphorus in citrus leaves based on hyperspectral imaging. International Journal of Agricultural & Biology Engineering 8(2).

Lyra WS, Silva EC, Araújo MCU, Fragoso WD, Veras G (2010) Classificação periódica: um exemplo didático para ensinar análise de componentes principais. Química Nova 33(7):1594-1597.

Malavolta E, Vitti GC, Oliveira SA (1989) Avaliação do estado nutricional das plantas: princípios e aplicações. Piracicaba, Associação Brasileira de Potassa e do Fósforo. 319p.

Marschner P (2013) Mineral nutrition of higher plants. Beijing, Science Press, 651p.

Mei Qing L, Jinyang L, Hanping M, Yanyou W (2016) Diagnosis and detection of phosphorus nutrition level for Solanum lycopersicum based on electrical impedance spectroscopy. Biosyst Eng 143:108–118. https://doi.org/10.1016/j.biosystemseng.2016.01.005

Mulla DJ (2013) Monitoring litchi canopy foliar phosphorus content using hyperspectral data. Computers And Electronics In Agriculture 154:176-186. DOI: http://dx.doi.org/10.1016/j.compag.2018.09.007.

Nowaki RHD, Parent SÉ, Cecilio Filho AB, Rozane DE, Meneses NB, Silva JA, Parent LE (2017) Phosphorus over-fertilization and nutrient misbalance of irrigated tomato crops in Brazil. Frontiers in Plant Science 8:1-11.

Oliveira TF, Pinto FAC, Silva DJH (2019) Spectral vegetation indexes applied to nitrogen sufficiency index: a strategy with potential to increase nutrient use efficiency on tomato crop. Engenharia Agrícola 39(1):118-126. DOI: https://doi.org/10.1590/1809-4430-Eng.Agric.v39n1p118-126/2019.

Peng-Tao G, Zhou S, Mao-Fen L, Wei L, Zheng-Zao C (2018) A robust method to estimate foliar phosphorus of rubber trees with hyperspectral reflectance, Industrial Crops and Products 126:1-12, ISSN 0926-6690. DOI: https://doi.org/10.1016/j.indcrop.2018.09.055.

Pir RMS (2016) Nutrient Deficiency and Syndrome Recognition in both Mango Leaf and Cotton Plant using K-means Clustering and BPNN. International Journal of Engineering Development and Research 4(1).

Rousse JW, Haas RH, Schell JA, Deering DW (1973) Monitoring Vegetation Systems in the Great Plains with ERTS. In: A Symposium held. Washington, Goddard Space Flight Center, Proceedings...

Siedliska A, Baranowski P, Pastuszka-Woźniak J, Zubik M, Krzyszczak J (2021) Identification of plant leaf phosphorus content at different growth stages based on hyperspectral reflectance. BMC Plant Biol 21 28. DOI: https://doi.org/10.1186/s12870-020-02807-4

Stein BR, Thomas VA, Lorentz LJ, Strahm BD (2014) Predicting macronutrient concentrations from loblolly pine leaf reflectance across local and regional scales. Gisci. Remote Sensing 51:269–287.

Sun Y, Gao J, Wang K, Shen Z, Chen L (2018) Utilization of machine vision to monitor the dynamic responses of rice leaf morphology and colour to nitrogen, phosphorus, and potassium deficiencies. Journal of Spectroscopy, p1-13. DOI: http://dx.doi.org/10.1155/2018/1469314.

Trani PE, Kariya EA, Hanai SM, Anbo RH, Bassetto Júnior OB, Purquerio LFV, Trani AL (2015) Calagem e adubação do tomate de mesa [Boletim Técnico IAC 215]. Campinas, Instituto Agronômico. 35p.

Withers PJA, Rodrigues M, Soltangheisi A, Carvalho TS, Guilherme LRG, Benites VM, Gatiboni LC, Sousa DMG, Nunes RS, Rosolem CA, Andreote FD, Oliveira Júnior A, Coutinho ELM, Pavinato PS (2018) Transitions to sustainable management of phosphorus in Brazilian agriculture. Science Report 8(1):2537. DOI: http://dx.doi.org/10.1038/s41598-018-20887-z

Yan Li L, Lyu Q, He SL, Yi SL, Liu XF, Xie R, Zhong Y, Deng L. (2015) Prediction of nitrogen and phosphorus contents in citrus leaves based on hyperspectral imaging. International Journal Agriculture & Biology Engineering 8(2): 80-88.

Yang Z, Willis P, Mueller R (2008) Impact of band-ratio enhanced awifs image to crop classification accuracy. In: Pecora – The Future of Land Imaging… Going Operational. Colorado, Proceedings...