Smart Agricultural Technology 4 (2023) 100213

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Predicting coffee water potential from spectral reflectance indices with neural networks

Pedro Henrique Nunes [a], Eduardo Vilela Pierangeli [a], Meline Oliveira Santos [b], Helbert Rezende Oliveira Silveira [c], Christiano Sousa Machado de Matos [c], Alessandro Botelho Pereira [d], Helena Maria Ramos Alves [e], Margarete Marin Lordelo Volpato [f], Vânia Aparecida Silva [f], Danton Diego Ferreira [a,*]

[a] *Department of Automatic, Federal University of Lavras, Lavras, MG, Brazil*
[b] *DTI-B INCT Café/EPAMIG Sul, Lavras, MG, Brazil*
[c] *Consórcio Pesquisa Café/EPAMIG Sul, Lavras, MG, Brazil*
[d] *Técnico EPAMIG ITAC, Lavras, MG, Brazil*
[e] *EMBRAPA, Brasília, DF, Brazil*
[f] *EPAMIG Sul, Lavras, MG, Brazil*

ARTICLE INFO

Editor: Stephen Symons

Keywords:
Artificial intelligence
Artificial neural networks
Coffee trees
Decision trees
Water potential

ABSTRACT

Leaf water potential is one of the main parameters used to assess water relations in plants by revealing levels of tissue hydration. It is commonly measured with the Scholander pressure chamber; which demands hard work and a time-consuming process. On the other hand, there is a diversified literature demonstrating the assessments of several plant variables via indices of leaf reflectance, that also present direct and indirect relationships with water potential. The aim of this work is to exploit spectral variables to estimate the water potential of coffee plants by using computational intelligence approaches. Data was collected in the cities of Santo Antônio do Amparo and Diamantina, Brazil, from 2014 to 2018. Two neural networks (Multi-Layer Perceptron) were designed to estimate and classify leaf water potential based on spectral variables. Moreover, a classifier and an estimator based on decision tree were also developed. The results showed that the artificial neural network model was superior as an estimator when compared with the decision tree model, with an average confidence index of 0.8550. On the other hand, decision trees showed a slightly higher performance as a classifier, with an overall accuracy of 88.8% and a *Kappa* index of 70.07%. We concluded that the leaf reflectance indices may be properly used to build accurate models for estimating coffee water potential. The indices PRI, NDVI, CRI1 and SIPI were the most relevant ones for estimating and classifying the coffee water potential.

1. Introduction

The rational use of water resources in agriculture, both for saving water and improving plant yield, is an aspect of great relevance for sustainable production systems. Overall, such demands make it highly necessary to seek alternative technologies that benefit the water management in croplands.

Knowledge of the plant water relations is essential to establish suitable agricultural practices. In this vein, there are several variables able to describe the plant water *status*; among them, the leaf water potential (ψ_{am}) is the most highlighted. This index is often directly measured via Scholander pressure chamber, in which leaf samples collected from plants are submitted to different pressures to assess ψ_{am} [1]. However, this is a hard labor and time-consuming method which demands a longer period of analyses and raises safety concerns, such as the risk of explosion. Moreover, it is also classified as a destructive method and thus can cause damage to the plant tissue. Hence, some studies have addressed the indirect assessment of ψ_{am} by using spectral indices related to the leaf reflectance of targeted plants with promising results [25].

* Corresponding author.
E-mail addresses: phnunes95@gmail.com (P.H. Nunes), dudupierangeli@gmail.com (E.V. Pierangeli), melineoli@hotmail.com (M.O. Santos), herosrezende@yahoo.com.br (H.R.O. Silveira), cmatosepamig@gmail.com (C.S.M. de Matos), alessandro.pereira@epamig.br (A.B. Pereira), helena.alves@embrapa.br (H.M.R. Alves), margarete@epamig.br (M.M.L. Volpato), vania.silva@epamig.br (V.A. Silva), danton@ufla.br (D.D. Ferreira).

https://doi.org/10.1016/j.atech.2023.100213
Received 3 October 2022; Received in revised form 2 March 2023; Accepted 3 March 2023
Available online 8 March 2023
2772-3755/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

One of the ways to obtain spectral indices is by applying remote sensing and field spectrometers. These methods work by measuring the percentage of light reflected by the plant leaves as function of different light wavelengths. Based on these values, the indices are calculated via equations that already exist in the literature.

The foliar reflectance indices combined with one or more bands can provide information about water status due to changes in spectral reflectance properties occasioned by variations in plant water content, antioxidants levels, photosynthetic pigments content, and photosynthetic rates [8]. There is a correlation between visible and red-edge, near-infrared spectral indices, and crop water content [18]. Reflectance spectral presents many possibilities of indices because there are several patterns of absorption by water or pigments in those regions in response to plant dehydration. Among these reflectance indices, the PRI (Photochemical Reflectance Index) is correlated with the photosynthetic and fluorescence parameters of chlorophyll, and can be used as a reliable indicator of water deficit - an abiotic limiting factor of photosynthesis. This index is sensitive to the oxidation of pigments from the xanthophyll cycle, which is one of the mechanisms to dissipate the energy of radiation under water deficit conditions [25]. Although PRI characterizes photosynthetic efficiency, the plant senescence reflectance index PSRI (Plant Senescence Reflectance Index) and the SIPI (Structure Insensitive Pigment index) respond sensitive to changes in the proportion of carotenoids and chlorophyll and were used as a quantitative measure of leaf senescence events [19]. The NDVI (Normalized Difference Vegetation Index) is more closely sensitive to the presence of chlorophylls and is used as input variable in models and algorithms to estimate the water potential by remote sensing in coffee [15]. By using different water absorption characteristics regardless of chlorophyll concentration, the WBI (water band index) has been reported to be an effective indicator of water stress, thereby indicating changes in the relative water content [4]. Other indices - such as ARI1 (anthocyanin reflectance index) and CRI1 (carotenoid reflectance index) - are associated with pigment absorption bands related to antioxidant responses, mechanisms of photoprotection and adaptation to light, as well as for the early diagnosis of water stress in plants [24].

The aforementioned papers have investigated the relations between spectral indices and the water potential, but for different cultivars, except the work of Maciel et al. [15] that focused on satellite images to estimate the leaf water potential of coffee. Unlike these, the proposed work aimed to exploit spectral variables captured by a mini-spectrometer to estimate the water potential of coffee plants by using computational intelligence approaches. To do this, we employed artificial neural networks (ANN) [13] and decision trees [23]. Due to their high generalization capacity, ANN are able to map the behavior of functions, even if they are non-linear and/or discontinuous [13]. The decision trees are capable of classifying samples into groups, and estimating their behavior according to characteristics values. In addition, they allow mapping data in a simple and easy way to understand, even of high dimensions, which is very useful to pattern recognition systems [12]. In this paper, these tools are used to classify the water potential of coffee trees, thereby identifying whether their water *status* is adequate, without requiring Scholander pressure chambers.

Coffee cultivation plays a fundamental role in terms of social and economic development in Brazil. Responsible for the generation of jobs, taxes and formation of the Brazilian foreign exchange earnings, coffee production is well-known as an important activity of the agricultural and economic sector of the country [9].

2. Methodology

2.1. Database

This work used a spectral database built by the team of researchers from the Empresa de Pesquisa Agropecuária de Minas Gerais (EPAMIG) and Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA). The samples were collected in the 2014-2018 timeframe, in arabica coffee fields located in the cities of Santo Antônio do Amparo and Diamantina, Minas Gerais, Brazil.

Table 1
Division of leaf water potential data classes.

Class	Break
C1	0 to -0.9429
C2	-0.9429 to -1.8857
C3	-1.8857 to -2.8286
C4	-2.8286 to -3.7715
C5	-3.7715 to -4.7144
C6	-4.7144 to -5.6573
C7	-5.6573 to -6.6002

The data set consists of 1280 events with 9 characteristics, as follows:

- *Photochemical Reflectance Index* - PRI (Reflectance ratio: R531-R570/R531 + R570) [10];
- *Plant Senescence Reflectance Index* - PSRI (Reflectance ratio: R680-R500/R570) [16];
- *Normalized Difference Vegetation Index* - NDVI (Reflectance ratio: R800-R680/R800 + R680) [21];
- *Water Band Index* - WBU (Reflectance ratio: R900 /R970) [19];
- *Anthocyanin Reflectance Index* - ARI1 (Reflectance ratio: 1/R550 - 1/R700) [20];
- *Carotenoid Reflectance Index* - CRI1 (Reflectance ratio: 1/R510) [14];
- *Structure Insensitive Pigment Index* - SIPI (Reflectance ratio: R800-R445/R800 + R680) [21];
- *Flavonol Reflectance Index* - FRI (Reflectance ratio: R800/R410-R800/R460) [17]

The variables were obtained by a mini-spectrometer, while the corresponding water potential (ψ_{am}) was measured with a Scholander pressure chamber. The data were collected on eleven different periods (dates), seeking to capture the effect of seasonal climatic variations in the region. This variable was converted into numerical values, which ranged from 1 to 4 according to the season, respectively, as follows: 1, spring; 2, summer; 3, autumn; and 4, winter.

For data estimation, normalized values of water potential were considered. Moreover, for classification purposes, the targets were divided into classes, according to the water potential value, which totalled 7 classes (Table 1).

2.2. Pre-processing

The outliers were manually eliminated, and then the samples normalized, to avoid any data biases [13]. The set scale of [0 1] was adopted as a rule of normalization for all data. Equation (1) was used to accomplish the normalization.

$$p_n = \frac{(p - p_{min})}{(p_{max}) - (p_{min})} \tag{1}$$

where p_n is the normalized value; p is the original value of a given variable; p_{min} is the lowest value; and p_{max} is the highest value.

The target values were multiplied by -1, thereby turning them positive, which makes it better of being visualized in a graphic format. After normalization, the data were shuffled at each interaction of the algorithm, and divided into 70% for training, 20% for validation (used for cross-validation) and 10% for testing.

Finally, an analysis of correlation was carried out to select the most relevant variables and to identify redundancies. For that, a linear correlation between variables and targets was performed, thereby eliminating those that showed either low (low relevance) or high correlation

Table 2
Kappa classification.

Kappa Value	Performance
0.81 to 1.00	Great
0.61 to 0.80	Very Good
0.41 to 0.60	Good
0.21 to 0.40	Reasonable
0.00 to 0.20	Bad
< 0.0	Terrible

Table 3
id value classification.

id value	Performance
> 0.85	Great
0.76 to 0.85	Very Good
0.66 to 0.75	Good
0.61 to 0.65	Median
0.51 to 0.60	Sufferable
0.41 to 0.50	Bad
< 0.40	Terrible

(high redundancy) with each other, reducing thus the dimensionality of the problem [13].

2.3. Performance evaluation metrics

Seeking to evaluate the performance of the developed classifiers, confusion matrices were set up. In addition, the *Kappa* index, proposed by Cohen [6], was used as a parameter.

The *Kappa* index is obtained by the analysis of the confusion matrix and is determined based on Equation (2). It is a suitable measure to assess the accuracy of a classifier, because it considers all the elements of the confusion matrix, differently from the global accuracy which uses only the main diagonal.

$$K = \frac{(P_o - P_e)}{1 - P_e} \quad (2)$$

where K is the coefficient of *Kappa*, P_o and P_e are defined by Equations (3) and (4), respectively.

$$P_o = \sum \frac{n_{ii}}{n} \quad (3)$$

$$P_e = \sum \frac{(n_i . n_j)}{n^2} \quad (4)$$

where n_{ii} is the number of elements correctly classified, n is the total of samples, n_i is the sum of the i-th line elements; and n_j the sum of the j-th column, considering $i = j$.

Once the *Kappa* coefficient was determined, the results can be classified according to Table 2, which provides the measure of the system's performance.

Seeking to assess estimators' performance, the average relative error, overall mean and standard deviation for test interactions were initially determined. In addition, the average confidence index, best confidence index and standard deviation were also calculated.

The performance analysis of the estimators was carried out by calculating the correlation (ρ), concordance (d), and the confidence (id) indices, which represent the accuracy, reliability and confidence of the model, respectively.

The correlation index was obtained using the *Pearson* coefficient, according to Equation (5).

$$\rho = \frac{cov(X, Y)}{\sqrt{var(X).var(Y)}} \quad (5)$$

where X and Y are the vectors to be correlated, and ρ is the correlation coefficient of *Pearson*.

The concordance index is given by Equation (6), where P_i is the estimated value; O_i is the observed value; and O is the mean of observed values [22].

$$d = 1 - \frac{\sum(P_i - O_i)^2}{\sum(|P_i - O| + |O_i - O|)^2} \quad (6)$$

The confidence index (id) was accomplished by the product of correlation (ρ) and concordance (d), and can be used as an indicator of estimators' performance. Camargo and Sentelhas [3] proposed the classification of performance of estimators based on Table 3, which was also used by Soares et al. [22] and by Batista et al. [2] as a methodological strategy to analyze their results.

Table 4
Correlation matrix.

	$\Psi_a m$	PRI	PSRI	NDVI	WBU	ARI1	CRI1	SIPI	FRI	Data
$\Psi_a m$	1									
PRI	0.27	1								
PSRI	-0.04	-0.19	1							
NDVI	0.20	0.49	-0.24	1						
WBU	0.04	-0.03	0.02	0.07	1					
ARI1	-0.04	0.34	0.03	0.56	-0.11	1				
CRI1	0.20	0.42	0.04	0.61	-0.18	0.53	1			
SIPI	0.27	0.30	-0.08	0.60	0.08	0.07	0.46	1		
FRI	0.10	0.05	0.02	0.25	-0.10	0.09	0.55	0.58	1	
Date	0.22	0.02	-0.09	0.02	-0.38	0.01	0.13	0.09	0.08	1

Table 5
The range for each region of the electromagnetic spectrum.

Color	Break
Violet	400 to 450 nm
Blue	451 to 520 nm
Green	521 to 570 nm
Red	571 to 700 nm
Yellow	681 to 740 nm
Orange	740 to 800 nm

3. Results and discussions

This section presents the results and discussions of the proposed classification and regression systems, and is divided into four subsections: feature selection; classifiers, estimators and comparative results.

3.1. Feature selection

Based on the procedures of performance analysis proposed in Section 2.2, a correlation matrix is presented in Table 4. Overall, it represents the correlation matrix between all attributes including target (ψ_{am}). Firstly, indices presenting the highest correlation values with the target were selected. The highest correlation values among attributes were identified as redundancy occurrences and, therefore, they were discarded.

Based on the correlation results, PSRI, ARI1 and FRI indexes were eliminated, as they presented high redundancy and low relevance. The reflectance indices combine the hyper spectral reflectance into two or three wavelengths, which is simultaneously using a common range of the spectrum. The range considered for each region of the electromagnetic spectrum is shown in Table 5.

Despite do not presenting redundancy with other characteristics, the WBU index was also removed, because it showed a low correlation with the target (ψ_{am}). The water band index reflects the water absorption in the mesophyll and increases as a function of decreases in water content. However, coffee displays a high relative water content in the leaf (RWC) to retain a high volume of water under dehydration conditions, which impossibilities a correlation with the WBI index [19]. Therefore, the selected indices were PRI, NDVI, CRI1, and SIPI. The variable Date was also selected.

Table 6
Confusion matrix of the best result with 100 interactions for ANN in %.

	C1	C2	C3	C4	C5	C6	C7
C1	90.85	4.96	1.89	1.37	0.90	0.02	0.00
C2	2.18	95.51	0.38	1.54	0.26	0.13	0.00
C3	3.58	2.68	81.07	7.80	4.09	0.25	0.51
C4	15.35	4.18	22.32	23.72	31.16	3.25	0.00
C5	15.14	4.59	5.4	26.60	33.95	13.76	0.92
C6	1.92	3.84	1.92	13.46	30.76	26.92	21.15
C7	0.00	0.00	0.00	0.00	5.62	18.47	75.90

Table 7
Confusion matrix of the best result with 100 interactions for Decision Tree in %.

	C1	C2	C3	C4	C5	C6	C7
C1	91.57	4.82	1.64	1.20	0.76	0.00	0.00
C2	5.00	93.11	1.44	0.00	0.44	0.00	0.00
C3	2.69	3.09	87.33	3.50	3.36	0.00	0.00
C4	9.42	0.67	14.81	39.39	32.99	2.69	0.00
C5	12.30	3.17	9.52	31.34	35.71	6.74	1.19
C6	0.00	0.00	0.00	0.00	17.54	59.64	22.80
C7	0.00	0.00	0.00	0.00	0.00	17.62	82.38

The indices PRI, SIPI and NDNI offer ways to quantify vegetation's ability to use incident light for photosynthesis. The water deficit directly affects the reflectance properties of leaves, changes often occur in the visible spectral region rather than in the infrared because of the sensitivity of chlorophyll to physiological disturbances. The correlations of NDVI, CRI and SIPI with ψ_{am} may indicate that variations in water potential cause changes in the contents of chlorophyll, carotenoids and in the relationship between these pigments, respectively. These changes may contribute to photoprotection, photosynthetic acclimation and photosynthetic efficiency in coffee plants, because the chlorophylls and carotenoids are part of the essential structures of the photosynthetic antenna, and help stabilize chlorophyll-protein complexes. Furthermore, the PRI index indicates that changes in water potential in coffee plants are associated with the photosynthetic and fluorescence parameters of chlorophyll. Therefore, PRI, SIPI and NDNI could be used as a reliable indicator of water deficit as an abiotic factor limiting photosynthesis.

3.2. Classifiers

In this section, the results obtained with the proposed classifiers are presented and discussed.

The number of neurons in each layer of the artificial neural network, and the corresponding activation function was found experimentally by using the training data set. It was found that the best neural network configuration has five input nodes, five neurons in the hidden layer and seven output neurons, each one representing a class. The confusion matrix shown in Table 6 was created containing the best-defined architecture resulted of 100 interactions. It is possible to observe that classes C1, C2, C3 and C7, that are the classes with higher number of data, presented the best hits.

A solution to improve the ANN performance for classes C4, C5 and C6, which present a lower number of occurrences, is to balance the data distribution between classes. The lower number of occurrences in classes C4, C5 and C6 is justified by the fact that there was less occurrence of water potential lower than -3 MPa under field conditions during the studied period (year). Except for 2014, when there was a low rainfall, which contributed to enriching the number of occurrences in class 7. In addition, another alternative would be the samples quantity rebalance by using other approaches to perform data augmentation, like the Synthetic Minority Oversampling Technique (SMOTE) [5].

Considering the confusion matrix of the best result for the ANN model (Table 6), the index *Kappa* was calculated according to Equation (2). This presented a value of 67.21%, thereby labeling the classifier as "very good" (Table 2).

For decision tree construction, 100 interactions were accomplished for each pruning configuration. The confusion matrix of the best result is shown in Table 7. In this case, the *Kappa* index presented a value of 71.07%, which categorize the classifier as "very good" (Table 2). Thus, it places in the same category as the classifier based on artificial neural networks, despite the higher *Kappa* coefficient.

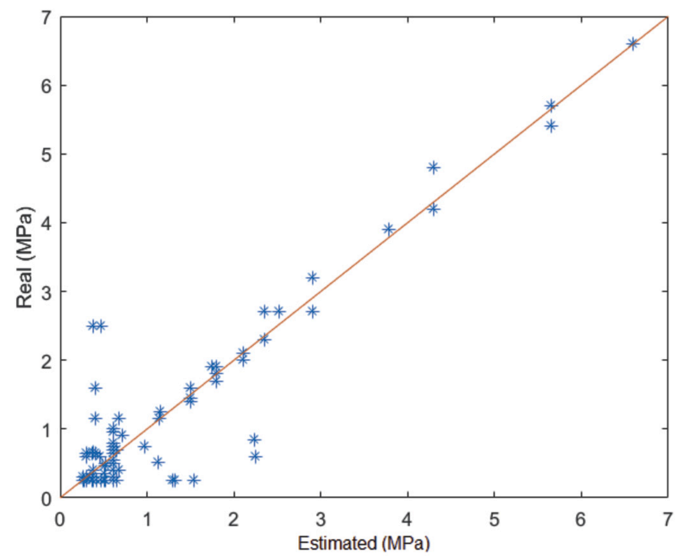

Fig. 1. Dispersion of the estimated data in the ANN model.

3.3. Estimators

To define the number of neurons in the layers of the ANN used for regression, besides determining the best activation functions, a similar procedure to the previously used for the classifier was performed, in which 100 runnings were carried out. For the construction of decision tree regression, 100 runnings were also performed for each pruning configuration.

Figs. 1 and 2 present the real water potential values versus the estimated ones (for one of the 100 runnings) and illustrate the data dispersion in relation to the ideal line, for the ANN and the decision tree models, respectively. Note that there are more values of water potential in the range 0 to 3 MPa. Also, it is in this range that the greatest dispersions between the actual and estimated values occur. The more distant from the line, the greater the errors associated with the observed points. It is noted that the ANN model presents more data dispersion for small values of water potential while the decision tree model presents a relatively constant dispersion for different values of water potential.

Overall, the performance of both estimators (ANN and decision tree) can be considered satisfactory, since the ANN model was classified as "Great" and the decision tree as "Very Good" by the criterion of Camargo and Sentelhas [3], with a low standard deviation in 100 interactions, which demonstrates the constancy of the results (see Table 8). Moreover, the mean relative error (ERM) of 7.27 ± 4.44% and 3.85 ± 4.27%, for the ANN and decision tree models, respectively, reinforced the integrity of the results.

3.4. Comparative results

In order to define the best technique, Tables 8 and 9 were set up, which expose the best results obtained in the regression and classification systems, respectively. σ_{ERM} in Table 8 represents the standard

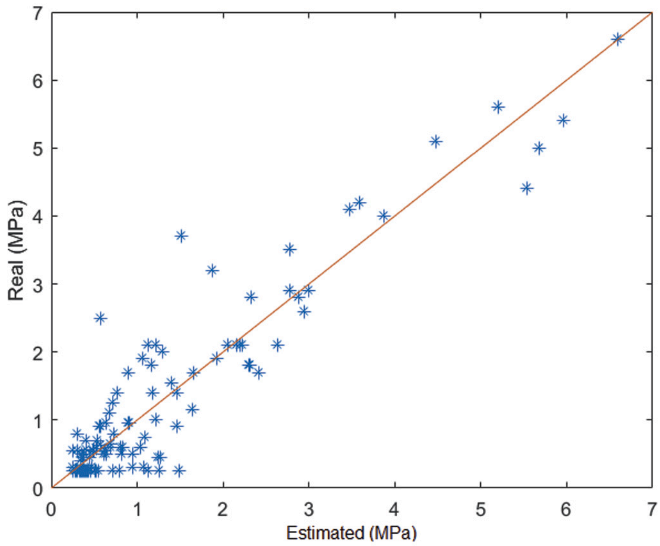

Fig. 2. Dispersion of the estimated data in the decision tree model.

Table 8
Results for the regression systems based on Artificial Neural Networks (ANN) and Decision Trees (DT).

Technique	ERM	σ_{ERM}	σ_{i_d}	μ_{id}	id_{max}
ANN	7.27	4.44	**0.0739**	**0.8550**	**0.9544**
DT	**3.85**	**4.27**	0.0909	0.8299	0.9443

Table 9
Results for the classification systems based on Artificial Neural Networks (ANN) and Decision Trees (DT).

	EQM	σ_{EQM}	$\downarrow EQM$	Hit	P.C.	P.C. (%)
ANN	0.1203	0.0523	**0.0055**	87.9%	4	23.7%
DT	**0.1113**	**0.0255**	0.0547	**88.8%**	5	**35.7%**

deviation of the ERM, μ_{id} the average confidence index, σ_{id} its standard deviation, and id_{max} represents the highest confidence index obtained.

Based on the analysis presented in Table 8, it is observed that the decision tree model displayed lower values of both mean relative error and relative error deviation, and it can be justified by the fact that decision trees only return fixed values, and its estimation is performed via classification into many classes, thus preventing intermediate values, which lead to smaller deviations in the estimated values. However, the artificial neural network showed a higher average confidence index, as compared to observed in decision tree, which demonstrates its better constancy. Finally, the ANN model presented a higher confidence index in its best interaction as compared to that accomplished by the decision tree at the same circumstances. These results demonstrate a greater capacity for generalization of ANN, for the regression problem, listing it as the best tool for this scenario, considering the confidence index as the performance evaluation criterion.

For the classification results shown in Table 9, EQM represents the mean square error, $\downarrow EQM$ is the smallest EQM obtained; σ_{EQM} is the standard deviation of EQM, PC represents the "Worst Class", PC (%) is its corresponding percentage of hits, and Hit refers to the global hit percentage. Based on Table 9, it is possible to infer that the decision tree was superior to the neural network in almost every aspects, except for the smallest mean square error of a single interaction, which corresponded to 0.0055 and 0.0547 for ANN and decision tree, respectively. This is likely justified by the singularities of the samples used in the tests of this specific analysis, not necessarily reflecting the superiority of the neural network, since, as already mentioned, its performance was inferior regarding both the average value and standard deviation of the EQM.

Another point considered when these results are analyzed is the percentage of correct answers in the worst class. As observed, this value was higher for the decision tree, evidencing their greater generalization capacity for this case, which corroborates other results presented in Table 9.

Considering the percentage of correct answers, the average of EQM in the interactions, and its standard deviation, no significant differences were observed, but the decision tree remained also slightly better in such aspects. Finally, when *Kappa* coefficient is observed, it is possible to note that the decision tree maintained as superior (71.07% against 67.21% of the ANN classifier), justifying its choice as the best classification methodology in comparison with the ANN-based one.

4. Conclusions

Leaf water potential is one of the main parameters used to assess water relations in plants by revealing levels of tissue hydration. In this paper, we exploited indices of leaf reflectance extracted form a mini-spectrometer to estimate the water potential of coffee plants by using artificial neural networks and decision trees approaches. In addition, we also investigated the capability of these approaches in acting as classifiers, by changing the problem from estimation to classification. To do this, the leaf water potential values were segmented into clusters.

As the study was carried out in different places in terms of climatic conditions (Santo Antônio do Amparo city – region with a shorter dry period and Diamantina city – a region with a longer dry period), the data include wide ranges of water potential that allow the creation of generic models for different regions. In addition, the study period comprised varied climatic conditions (from 2014 to 2018 – a long and representative set of coffee crops), including an extreme drought event that occurred in 2014. In addition, the proposed models take advantage of being simple, since only four variables are required (PRI, NDVI, CRI1, SIPI and Date) to perform the water potential estimation/classification.

The results showed the slightly superiority of the neural network model for the estimation task and of the decision tree model for the classification task. Thus, it allows their application in similar problems (as in this study) according to the adopted approach (classification or estimation/regression).

It is worth considering that decision trees facilitate the understanding of operations that underlie the results, since they are composed of "*If, Then*" rules. On the other hand, the artificial neural networks can even be easily replicated using a sequence of sums and multiplications after being trained. They are considered a black box system in which it is unknown the pathway to the results. Therefore, such characteristics must be considered during the choice of the methodology to be used, as the results did not show major differences in performance.

For future works, the authors intend to implement the developed approaches in embedded systems in association with a mini-spectrometer, which may provide an indirect measure of water potential of coffee plants based only on the observed spectral indices, thus not requiring Scholander pressure chambers. Also, we intend to exploit other machine learning methods, like support vector machines, gaussian process regression, and deep learning-based algorithms, which have been extensively used for classification and regression purposes [7,11].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

The project team thanks National Council for Scientific and Technological Development – CNPq, FAPEMIG, CAPES, Consórcio Pesquisa Café and INCT Café for their support.

References

[1] J. Barnes, L. Balaguer, E. Manrique, S. Elvira, A. Davison, A reappraisal of the use of dmso for the extraction and determination of chlorophylls a and b in lichens and higher plants, Environ. Exp. Bot. 32 (2) (1992) 85–100.

[2] L.A. Batista, R.J. Guimarães, F.J. Pereira, G.R. Carvalho, E.M.d. Castro, Leaf anatomy and water potential in the coffee cultivars tolerance to water stress, Rev. Ciênc. Agron. 41 (09 2010) 475–481.

[3] A. Camargo, P. Sentelhas, Performance evaluation of different potential evapotranspiration estimating methods in the state of São Paulo, Brazil, Rev. Bras. Agrometeorol. 5 (01 1997) 89–97.

[4] L. Caturegli, S. Matteoli, M. Gaetani, N. Grossi, S. Magni, A. Minelli, G. Corsini, D. Remorini, M. Volterrani, Effects of water stress on spectral reflectance of bermudagrass, Sci. Rep. 10 (1) (2020) 15055.

[5] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, Smote: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.

[6] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1) (1960) 37–46.

[7] B. Daryayehsalameh, M.A. Ayari, A. Tounsi, A. Khandakar, B. Vaferi, Differentiation among stability regimes of alumina-water nanofluids using smart classifiers, Adv. Nano Res. 12 (5) (2022) 489–499.

[8] S. Elsayed, S. El-Hendawy, M. Khadr, O. Elsherbiny, N. Al-Suhaibani, Y.H. Dewir, M.U. Tahir, M. Mubushar, W. Darwish, Integration of spectral reflectance indices and adaptive neuro-fuzzy inference system for assessing the growth performance and yield of potato under different drip irrigation regimes, Chemosensors 9 (3) (2021).

[9] EMBRAPA, Café, https://www.embrapa.br/busca-de-noticias/-/noticia/77921739/faturamento-das-lavouras-dos-cafes-do-brasil-alcanca-r-56-bilhoes-em-2022, 2023. (Accessed 9 March 2023).

[10] J. Gamon, J. Peñuelas, C. Field, A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency, Remote Sens. Environ. 41 (1) (1992) 35–44, https://doi.org/10.1016/0034-4257(92)90059-S, https://www.sciencedirect.com/science/article/pii/003442579290059S.

[11] A. Garg, P. Aggarwal, Y. Aggarwal, M. Belarbi, H. Chalak, A. Tounsi, R. Gulia, Machine learning models for predicting the compressive strength of concrete containing nano silica, Comput. Concr. 30 (1) (2022) 33–42.

[12] L. Genc, M. Inalpulat, U. Kizil, M. Merik, S.E. Smith, M. Mendes, Determination of water stress with spectral reflectance on sweet corn (Zea mays L.) using classification tree (CT) analysis, Photochem. Photobiol. 100 (1) (10 2013) 81–90.

[13] S. Haykin, Neural Networks and Learning Machines, 3rd edición, Prentice Hall, New York, 2008.

[14] W. Kong, W. Huang, X. Zhou, Remote estimation of carotenoid/chlorophyll ratio in vertical layers using canopy multi-angle spectral data, in: 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics), July 2016, pp. 1–5.

[15] D.A. Maciel, V.A. Silva, H.M.R. Alves, M.M.L. Volpato, J.P.R.A.d. Barbosa, V.C.O.d. Souza, M.O. Santos, H.R.d.O. Silveira, M.F. Dantas, A.F.d. Freitas, G.R. Carvalho, J. Oliveira dos Santos, Leaf water potential of coffee estimated by landsat-8 images, PLoS ONE 15 (3) (03 2020) 1–13.

[16] M.N. Merzlyak, A.A. Gitelson, O.B. Chivkunova, V.Y. Rakitin, Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening, Physiol. Plant. 106 (1) (1999) 135–141.

[17] M.N. Merzlyak, A.E. Solovchenko, A.I. Smagin, A.A. Gitelson, Apple flavonols during fruit adaptation to solar radiation: spectral features and technique for non-destructive assessment, J. Plant Physiol. 162 (2) (2005) 151–160.

[18] I. Pôças, T.A. Paço, P. Paredes, M. Cunha, L.S. Pereira, Estimation of actual crop coefficients using remotely sensed vegetation indices and soil water balance modelled data, Remote Sens. 7 (3) (2015) 2373–2400.

[19] J. Penuelas, J. Gamon, A. Fredeen, J. Merino, C. Field, Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves, Remote Sens. Environ. 48 (2) (1994) 135–146.

[20] J.R. Reyes, J.S. Bohórquez, W.I. Alama, Hyperspectral analysis based anthocyanin index (ari2) during cocoa bean fermentation process, in: En: 2015 Asia-Pacific Conference on Computer Aided System Engineering, July 2015, pp. 169–172.

[21] L. Serrano, C. González-Flor, G. Gorchs, Assessment of grape yield and composition using the reflectance based water index in Mediterranean rainfed vineyards, Remote Sens. Environ. 118 (2012) 249–258.

[22] F.C. Soares, A.D. Robaina, M.X. Peiter, J.L. Russi, G.A. Vivian, Artificial neural networks to estimate soil water retention, Rev. Ciênc. Rural 44 (2) (2014) 293–300.

[23] S. Theodoridis, K. Koutroumbas, Pattern Recognition, fourth edition, Academic Press, 2009.

[24] R. Tosin, I. Pôças, H. Novo, J. Teixeira, N. Fontes, A. Graça, M. Cunha, Assessing predawn leaf water potential based on hyperspectral data and pigment's concentration of vitis vinifera l. In the Douro wine region, Sci. Hortic. 278 (2021) 109860.

[25] C. Zhang, E. Pattey, J. Liu, H. Cai, J. Shang, T. Dong, Retrieving leaf and canopy water content of winter wheat using vegetation water indices, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (1) (2018) 112–126.