Machine Learning for Soil Nutrients Prediction based on Optical and Electrical Impedance Spectroscopy

Janez Trontelj ml and Olga Chambers *

Faculty of Electrical Engineering, Trzaska cesta 25, Ljubljana 1000, Slovenia

***Corresponding author:** Olga Chambers, Faculty of Electrical Engineering, Trzaska cesta 25, Ljubljana 1000, Slovenia

Abstract

This paper addresses the problem of predicting soil properties using a combination of optical spectroscopic and electrical impedance methods. A comparative analysis between the most common machine learning methods such as Random Forest, Naive Bayes, Support Vector Machine, Decision Tree and Artificial Neural Network was performed using our research dataset consisting of 50 soil samples. The results indicate that none of the methods showed the best performance for nutrients prediction when only optical or electrical impedance spectroscopy measurements were used. Then, the influence of the principal components was validated   to improve the machine learning performance. Their negative influence on the overall accuracy was found. Finally, the influence of the nutrient category on the prediction was validated, where similar results were found for 3-level grade and 5-level grade systems indicating a possibility for more precise and accurate soil characterization. In addition, the work shows the importance of repeated measurements for each soil sample, which can improve the overall accuracy.

Keywords: Machine learning, Nutrient's prediction, Soil optical spectroscopy, Soil analysis, Artificial intelligence

Introduction

It is a significant interest in the development of a low-cost multi-sensor system for real-time measurements to monitor pH, soil salinity, moisture content, organic matter (OM), soil micronutrients and macronutrients such as nitrogen (N), phosphorus (P), potassium (K) and magnesium (M) in nowadays agriculture [1-3]. The Veris Technologies, Salina, KS, USA [4] equipped tractors with a multi-sensor system for on-the-going measurements in a field, where ion selective electrodes (ISE) are used. ISE sensors allow the most accurate prediction as they use chemical solutions that are sensitive to a particular soil component [5]. However, their high price, supplemental materials required for analysis and time required to get measurement make it unsuitable for real-time large field analysis.   In this paper, we present research that analyses machine learning performance when using measurements obtained simultaneously from two sensors, optical and electrical impedance spectroscopy, for phosphorus, potassium and magnesium prediction. These methods are the most common low-cost sensors showing promising results for implementation in a field for real-time characterization. Optical spectroscopy operates on the whole Ultraviolet-Visible-Near Infrared (UV-VIS-

NIR) range that enables rapid soil properties quantification. The spectroscopy principle is the interaction between incident light and soil surface properties, such that the reflected light varies as a function of soil physical and chemical properties [6]. The spectroscopic analysis of determining the nutrient content is not convenient. Some researchers indicate its unstable performance [7], showed the influence of the soil texture on the position of the potassium absorption center. Authors also reported that soil with high clay content results in a smaller change in absorption. The accuracy of the spectral analysis method is yet to be fully resolved. However, compared with traditional chemistry methods, it does not require chemical solutions and may provide reasonable measurements for rapid and non-destructive analysis. Another low-cost method for rapid data acquisition is electrical impedance spectroscopy. It is mainly used for soil moisture analysis or single-nutrient amount prediction when other components are neglected [8], reported their results for soil-nitrates detection in a real-time application indicating the electrical impedance spectroscopy's great potential for in-suite measurements. Authors showed that more information could be delivered from the real and imaginary part of the complex permittivity for several frequencies at the

same moisture value [9]. To the best of our knowledge, there is no literature reporting electrical impedance as a separate sensor for multi-componential soil characterization. It is commonly used as part of multi-sensor systems to get supplemented measurements for improving soil prediction in combination with other methods. For example, the commercial Agro Cares scanner [10], is based on reflectance spectroscopy and electric conductivity measurements to provide brief nutrients characterization. Machine learning is an advanced computer-based method that allows fast recognition of the invisible information and class prediction based on the obtained measurements and previous information. Although it is a well-known method, it is still unclear which classifier and its parameters best fit the research dataset. Therefore, for the accuracy of the analysis, several machine learning classifiers and parameters influencing its performance were validated. First, the five standard classifiers, Random Forest (RF), Naïve Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT) and Artificial Neural Network (ANN), were selected for comparative analysis. Then, the influence of the category selection was validated. This includes a comparison between 3-level and 5-level class labelling. Finally, a various number of Principal Components (PCs) were used to identify the best one for machine learning. As part of the machine learning strategy, the influence of the soil subsamples corresponding to the same soil on the overall performance was investigated by comparing corresponding results. The research analysis was performed and reported separately for electrical impedance spectroscopy, optical spectroscopy, and their combination to highlight their advantages and disadvantages.

Materials and Methods

Figure 1 shows a flowchart of the soil analysis using machine learning. It highlights the most important points of this research.

Dataset description

The research dataset of soil samples was collected over the whole Slovenia from 0-20 cm topsoil surface. It was naturally air-dried and then sieved with a 2 mm sieve. Dataset consists of 50 soil samples that were sent to a certified laboratory at Agriculture Institute of Slovenia [11], for chemical content characterization that also includes information about phosphorus, potassium, and magnesium. These nutrients were selected for analysis as they are the most common for vineyard and orchards. Three sub-samples randomly taken from the same soil were measured using optical and electrical impedance spectroscopy and joined in the research dataset. Because of the large variability of the soil sample's locations, it was called a Global dataset.

Optical spectroscopy data acquisition

Soil optical spectroscopy data were collected from air-dried soil samples with a flat surface. Measurements were performed within the UV-VIS-NIR range, i.e., 200 - 2500 nm, and joined into a research dataset of measurements. A deuterium-halogen light box was used as the light source. The light reflectance from the sample was measured by placing 5 g of air-dried sieved sample into a quartz glass petri dish three mm-diameter, as shown in Figure 2. The setup includes a fiber-coupled spectrometer FCR-7UV200-2-ME from Avantes that is fixed perpendicularly to have a 3 cm distance between the probe and samples. The light from a light source is sent through six illumination fibers to the sample, and the reflection is measured by a seventh fiber in the center of the reflection probe tip. The AvaSpec-ULS2048CL-EVO-RS and Ava Spec-HSC-TEC perform the light measurement in the UV-VIS-NIR range of the electromagnetic spectrum, i.e., 200-2500 nm. Spectra normalization was performed by dividing soil reflectance spectra by the white body reflectance spectra used here as a reference.
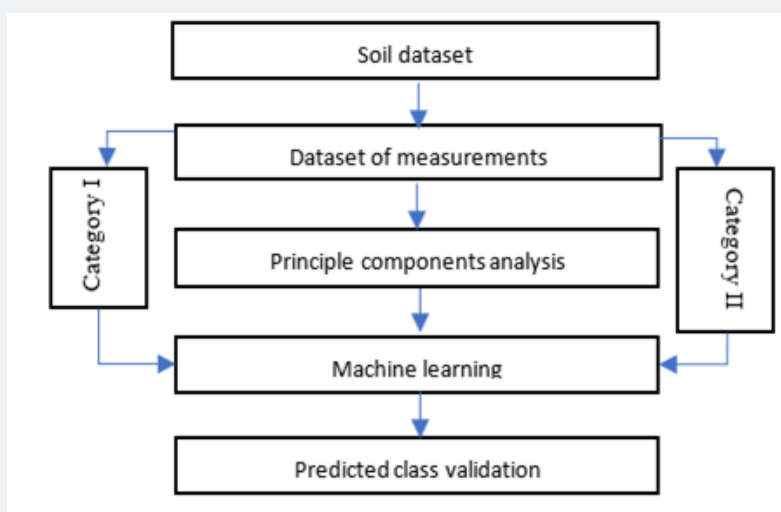

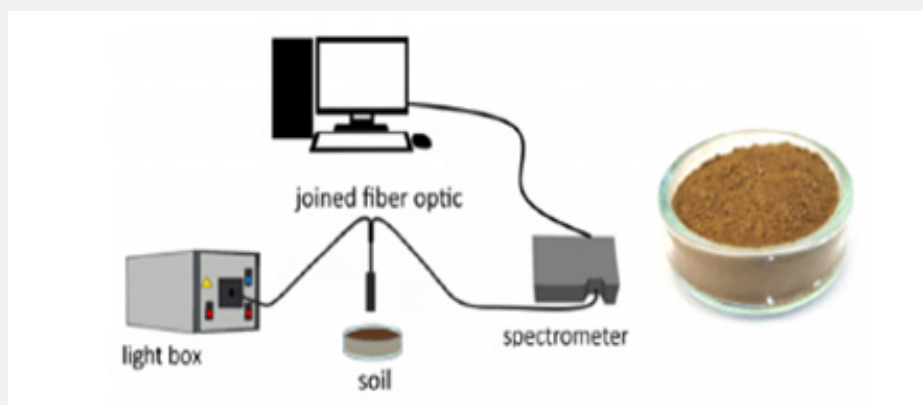**Figure 1:** Flowchart of the research analysis.

Figure 2: The experimental set-up for optical spectroscopy data acquisition.

Electrical impedance spectroscopy data acquisition

The laboratory setup for electrical impedance data acquisition is shown in Figure 3 that includes an impedance sensor connected to the personal computer for data processing and data storage. The sample holder is designed to hold near 5 g of the soil. The measuring process was controlled using a graphical user interface developed in MATLAB software [12]. Electrical impedance spectroscopy measures resultant voltages when a constant current is applied at different frequencies. The 122 frequencies selected between 30 kHz and 14 MHz enable a good fit of the impedance signal over the whole frequency domain. The laboratory developed a portable impedance spectrometer is shown in (Figure 4). It consists of analog, processing and sensor sections. The spectrometer generates AC current with user-defined frequency and sends it

in the soil through sensor electrodes. The resulting imaginary and real components of the impedance are then digitized and send to a personal computer. Electrical impedance spectroscopy measures resultant voltages when injecting a constant current into the sample at different frequencies over a selected range of interest. The 122 frequencies selected between 30 kHz and 14 MHz provided a good fit of the impedance signal over the whole frequency domain. As the frequency increased, the impedance of the samples dropped obviously. The lower fertilizer content is associated with higher impedance amplitude. Our earlier research indicates that the main information about soil solutions can be extracted from impedance magnitudes alone [14,15] that is also observed for bulk soil characterization. Therefore, only impedance magnitudes are used in the analysis.
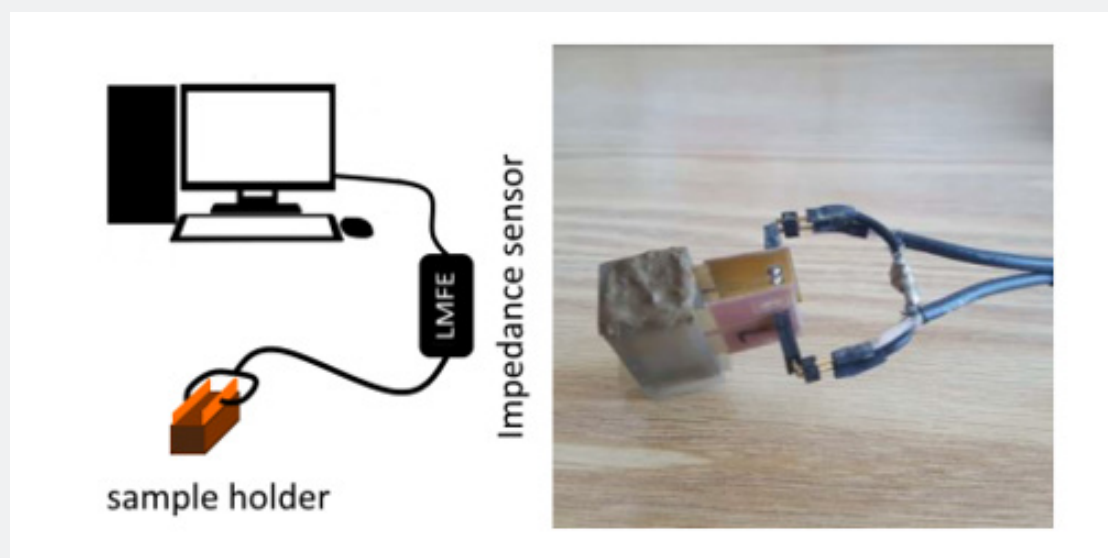

Figure 3: Left image: the measuring setup. Right image: impedance sensor designed for soil impedance measurement.
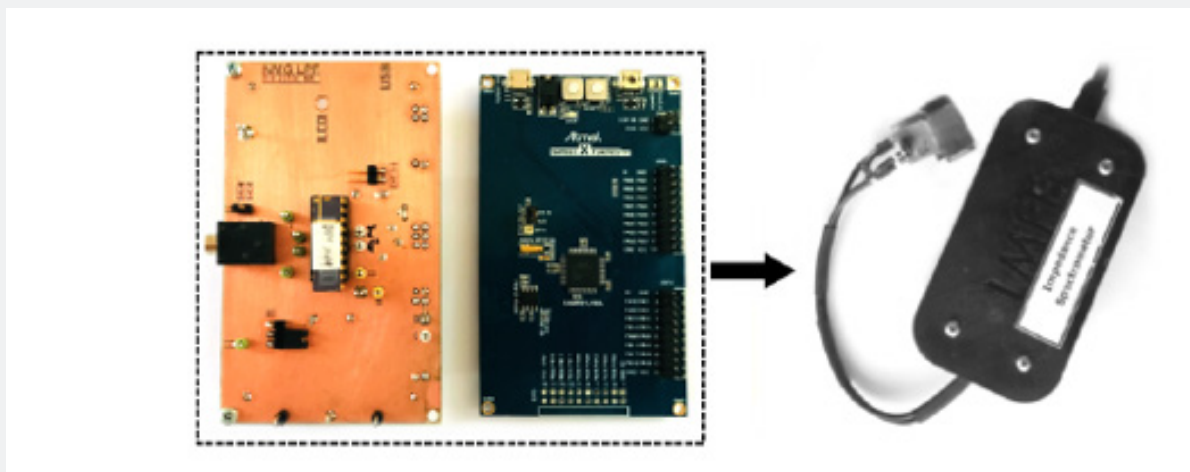
Figure 4: The laboratory developed a portable impedance sensor with its microcontroller [13].

Result

This paragraph shows the results of the measuring methods applied separately to the optical spectroscopy measurements, electrical impedance spectroscopy measurements and their combination. The influence of the following parameters was investigated:

a.   The category of soil nutrients level

b.   The research dataset

c.   The machine learning classifier

d.   The principal components

Table 1: Category range specifications.

Score, mg/100g	Category I	Category II
0 – 3	1	1
5 – 7		
8 – 10		
11 – 13		2
14 – 17		
18 – 20	2	
21 – 23		3
24 – 27		
28 – 30		
31 – 33	3	4
34 – 37		
38 – 40		
>40		5

(Table 1) shows the Category I and Category II details for data class labelling, where the first column provides information about nutrients in a soil sample. For example, if soil sample contains 2 mg of phosphorus, 15 mg of potassium and 32 mg of magnesium, the class label for this sample is "1-5-10" for Category I and "1-2-4" for Category II, where the first position corresponds to phosphorus, second – potassium and third – magnesium. First, a comparison between machine learning classifiers was performed (Figure 5) shows results corresponding to five classifiers when using measurements performed with optical spectroscopy, electrical impedance spectroscopy and their combination. It can be seen that the best results were obtained when using both, optical and electrical impedance spectroscopy together (Figure 5), indicates a different degree of accuracy, where ANN in most cases outperforms other classifiers providing a more accurate prediction of potassium. At the same time, magnesium and phosphorus were predicted less accurately. Next, the principal components' influence was investigated. The first 5 and 20 principal components were used to reduce the data dimensionality and highlight the most relevant information. Their influence was validated over classification and comparison with results obtained for data without applying principal components. The obtained results for each case may be seen in Figures 6-8 show precisions corresponding to optical spectroscopy, electrical impedance spectroscopy measurements and their combination, respectively. It can be seen that there was no positive influence of the PCs on the overall performance. Moreover, the results corresponding to electrical impedance spectroscopy were obtained with decreased accuracy. Results corresponding to optical spectroscopy measurements were slightly improved for RF and DT, but not for ANN. This is an important observation indicating a different influence of the PCs on the multi-sensor data analysis. (Table 2) shows results corresponding to Category I and Category II. It can be seen that the decrease of the nutrients levels not always increase the overall prediction accuracy. For example,

it can be observed some increase in potassium prediction when using the ANN. Nevertheless, the overall result indicates better performance of the machine learning for Category I, where larger ranges for nutrients characterization are used. Comparative analysis was performed using Category II for data labelling and ANN for classification. Table 3 provides results for a Global dataset with three and seven sub-samples corresponding to the same soil, respectively. The positive influence of the sub-samples on the prediction accuracy may be seen for all nutrients predictions.
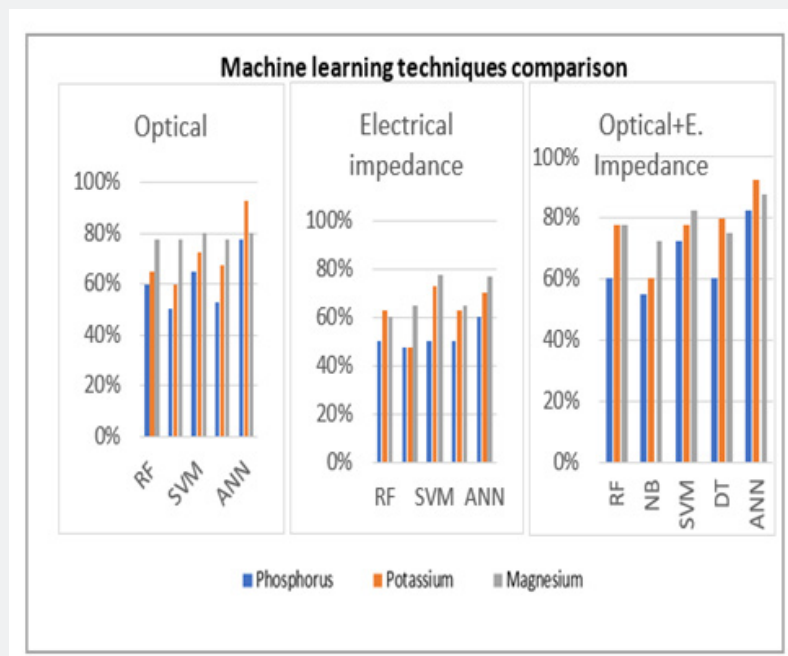

Figure 5: Comparison of the results corresponding different machine learning (ML) algorithms for research dataset using Grade II system.
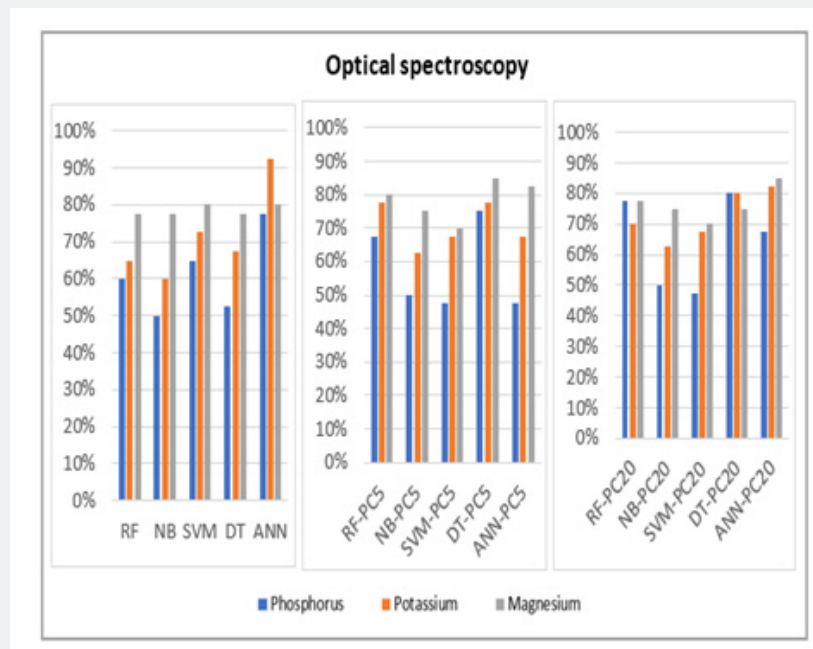

Figure 6: Results of the machine learning for nutrients prediction based on optical spectroscopy using Category II for labelling and Global research dataset.
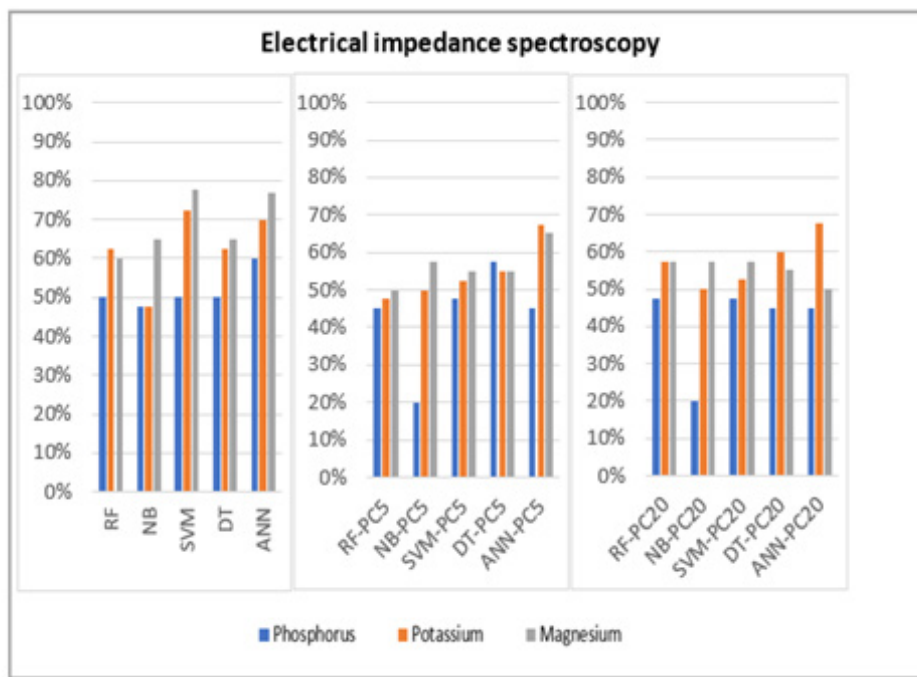
Figure 7: Results of the machine learning for nutrients prediction based on electrical impedance spectroscopy using Category II for labelling and Global research dataset.
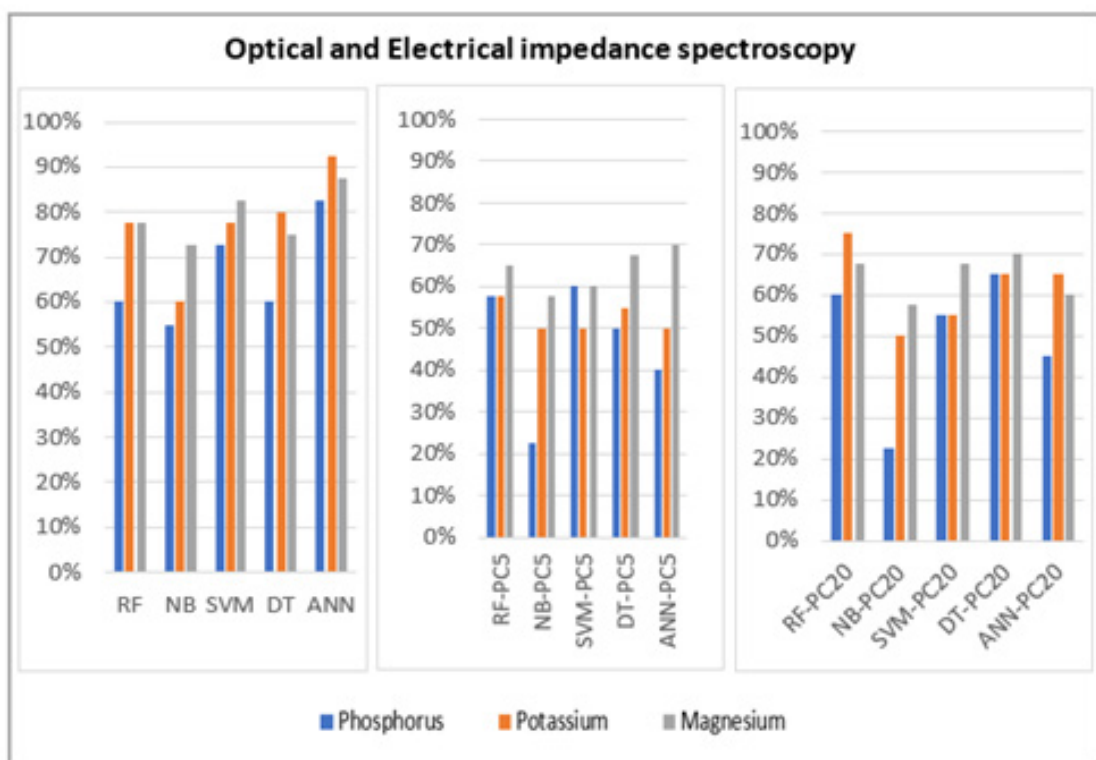

Figure 8: Results of the machine learning for nutrients prediction based on optical and electrical impedance spectroscopy using Category II for labelling and Global research dataset.

Table 2: Results of the machine learning using both optical and electrical impedance spectroscopy when using different categories for data labelling.

ML method	Phosphorus	Potassium	Magnesium
Category I (3 levels)			
RF	75%	82.50%	75%
NB	65%	65%	82.50%
SVM	75%	77.50%	87.50%
DT	87.50%	70%	85%
ANN	85%	90%	100%
Category II (5 levels)			
RF	60%	77.50%	77.50%
NB	55%	60%	72.50%
SVM	72.50%	77.50%	82.50%
DT	60%	80%	75%
ANN	82.50%	92.50%	87.50%

Table 3: Results of the ANN machine learning corresponding Global dataset, when using optical spectroscopy with Category II of data labelling.

	Phosphorus	Potassium	Magnesium
Global dataset with three sub-samples	67.50%	80.20%	77.50%
Global dataset with seven sub-samples	77.50%	92.50%	80.00%

Discussion and Conclusion

The results obtained in this study may defer from those of similar analyses due to the high influence of the dataset on the overall performance. The measurement set-up, chemical content and physical properties of the soil, instrument specifications, etc., are critical factors that must also be considered in the analysis. Our research indicates a significant difference between machine learning methods performance, where ANN in most cases outperforms other methods. Comparative analysis of the first 5 and 20 principal components was performed to investigate their influence on prediction accuracy. The Results reported in Figures 6-8 indicate better performance of the method without using PCs applied to the data. It tends to be seen that the more PCs are used, the more accurate performance can be obtained. Moreover, the negative effect of PCs can be observed in electrical impedance spectroscopy. Next, the influence of the category on the machine learning performance was investigated. Researchers mainly divide fertility levels into three main categories: below optimum, optimum, and above optimum [16,17], where the "low" or "high" category can increase or decrease the fertilizer recommendation by 25% or 30% of a general recommendation. The selection of nutrient categories is generally based on the requirements of agriculture, the yield level, and the type and nature of the soil in question [18]. The results obtained in our research indicate a very small difference between results obtained for 3-level and 5-level nutrients characterization. The DT for potassium and phosphorus prediction showed better results for 5-level class labelling than 3-level class labelling. This is an exciting observation because more precise soil characterization may be obtained without accuracy loss.

The following observations have been made based on the research results presented in this paper:

a) The best prediction accuracy was obtained using both optical and electrical impedance spectroscopy

b) The principal components extraction for dimensionality reduction, in general, did not show prediction accuracy improvement

c) The best effect of the principal components on the soil prediction was observed for DT and RF when using optical spectroscopy

d) In most cases, the phosphorus was the most difficult to predict with high accuracy, when potassium and magnesium predictions were the most accurate

e) Optical spectroscopy for soil prediction alone performs better than electrical impedance spectroscopy

f) Electrical impedance spectroscopy alone was the most successful for magnesium prediction

g) The increase of the number of the sub-samples in a dataset may increase the overall prediction accuracy

h) The increase of the category levels does not necessarily lead to the prediction accuracy decrease.

It should be noted that the observations are based on the results corresponding to our research dataset. Due to the considerable variation in the factors and their combinations, results for another research dataset may vary. Nevertheless, the correlations obtained in this research are important for

understanding the overall strategy building for soil properties prediction using optical and electrical impedance spectroscopy sensors. The obtained results can be improved when performing averaging between corresponding measurements. This will illuminate the interclass variation and influence of the common artifacts in the natural environment. Our research indicates that a combination of optical and electrical impedance spectroscopy provides different complementary information that enables accurate nutrients prediction in a field. We believe that this analysis may be helpful and provide vital information to improve low-cost multi-sensor system analysis for precise and accurate soil nutrients prediction.

References

1. L Burton, K Jayachandran, S Bhansali (2020) Review the "real-time" revolution for in situ soil nutrient sensing. Journal of The Electrochemical Society 167(3): 037569.

2. G Pandey, R Kumar, RJ Weber (2014) A low rf-band impedance spectroscopy based sensor for in situ, wireless soil sensing. IEEE Sensors Journal 14 (6): 1997-2005.

3. K Andreas, Z Angelos, P Panagiotis, K Georgios, Tommy, et al. (2018) A versatile electronic interface for soil quality P. 1-8.

4. Veris(r) technologies Pp. 05-17.

5. V Adamchuk, E Lund, B Sethuramasamyraja, M Morgan, A Dobermann, D Marx (2005) Direct measurement of soil chemical properties on-the-go using ion-selective electrodes. Computers and Electronics in Agriculture 48: 272-294.

6. M Luleva, H van der Wer, V Jetten, F Meer (2011) Can infrared spectroscopy be used to measure change in potassium nitrate concentration as a proxy for soil particle movement. Sensors 11: 4188-4206.

7. AM Mouazen, J De Baerdemaeker, H Ramon (2005) Towards development of online soil moisture content sensor using a fibre-type nir spectrophotometer. Soil and Tillage Research 80 (1): 71-183.

8. G Pandey, R Kumar, RJ Weber (2013) Real time detection of soil moisture and nitrates using on-board in-situ impedance spectroscopy, In: 2013 IEEE International Conference on Systems, Man, and Cybernetics Ppp. 1081–1086.

9. L Umar, R Setiadi (1656) Low-cost soil sensor based on impedance spectroscopy for in-situ measurement.

10. (2018) Agrocares nutrient intelligence.

11. (2018) Agriculture Institute of Slovenia.

12. (2020) MATLAB, 9.7.0.1190202 (R2019b), The MathWorks Inc., Natick, Massachusetts, R Razman, Neinvazijno merjenje krvnega sladkorja, Ph D. thesis, University of Ljubljana, Slovenia

13. O Chambers, A Sesek, R Razman, JF Tasic, J Trontelj (2018) Fertiliser characterisation using optical and electrical impedance methods. Computers and Electronics in Agriculture 155: 69-75.

14. TJ Chambers, Olga (2020) Comparison between electrical impedance and optical spectroscopy for a field soil analysis. in sensor devices: The eleventh international conference on sensor device technologies and applications 21-25.

15. A Morellos, XE Pantazi, D Moshou T, Alexandridis R, Whetton G, et al. (2016) Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using vis-nir spectroscopy. Biosystems Engineering 152: 104-116.

16. W Ng, Husnain L, Anggria AF, Siregar W, Hartatik Y, Minasny, et al. (2020) Developing a soil spectral library using a lowcost nir spectrometer for precision fertilization in Indonesia. Geoderma Regional 22: e00319.

17. X Pantazi, D Moshou, T Alexandridis, R Whetton, A Mouazen (2016) Wheat yield prediction using machine learning and advanced sensing methods. Computers and Electronics in Agriculture 121: 57-65.

18. A Srivastava, S Singh (2008) Dris norms and their field validation in nagpur mandarin. Journal of Plant Nutrition 31:1091-1107.

Your next submission with Juniper Publishers will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
  (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
https://juniperpublishers.com/online-submission.php