An Interpretable Machine Learning Framework for Soil Nutrient Assessment Based on pH and Electrical Conductivity

Taciano Ares Rodolfo[1,2,3], Oscar J. González Zárate[1,2,4], Carlos J. Gonzalez Aguilera[1,2]

[1] Center for Embedded Devices and Research in Digital Agriculture (CEDRA) of SENAI-RS, São Leopoldo, Brazil
[2] SENAI Innovation Institute for Sensing Systems (ISI-SIM), São Leopoldo, Brazil
[3] Federal University of Rio Grande do Sul (UFRGS) - Porto Alegre – Brazil
[4] Universidad de los Andes, Bogotá – Colombia

A paper from the Proceedings of the
17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture
13-15 July 2026
Porto Alegre, Brazil

Abstract.
Understanding how the physical and chemical properties of soil influence nutrient availability is fundamental for advancing precision agriculture, as these properties directly affect the efficiency of macro- and micronutrient absorption by plants. In recent years, the increasing availability of open agricultural datasets has created new opportunities for developing data-driven frameworks capable of supporting large-scale soil assessment and decision-making. However, the effective integration of heterogeneous soil attributes into interpretable and scalable analytical tools remains a significant challenge. In this work, we propose and validate a comprehensive analytical framework for soil characterization and nutrient availability assessment using an open-access dataset containing physical and chemical soil properties. The dataset comprises textural fractions of sand, silt, and clay, as well as chemical parameters including pH, electrical conductivity, calcium carbonate ($CaCO_3$), organic matter, and concentrations of multiple macro- and micronutrients. The proposed framework integrates Principal Component Analysis (PCA) for dimensionality reduction, K-means clustering for soil profile identification, and decision tree–based models for rule extraction and knowledge representation. All processing stages were implemented to ensure reproducibility and adaptability to different agricultural contexts. The experimental results demonstrate that pH is the primary factor governing micronutrient dynamics, particularly for manganese (Mn) and iron (Fe), by directly influencing their solubility and bioavailability. Electrical conductivity acts as a secondary modulating variable, especially under high-salinity conditions, where it may enhance or restrict the availability of specific nutrients. The clustering process consistently identified three dominant soil profiles, providing a structured representation of soil variability across the studied region. The proposed framework does not replace traditional laboratory analyses but provides a complementary, scalable, and cost-effective solution for preliminary soil diagnostics. It enables rapid assessment of large agricultural areas, supports optimized soil sampling strategies, and delivers interpretable, data-driven recommendations for nutrient management. Furthermore, by leveraging open datasets and models, the framework facilitates integration into digital agriculture platforms through low-cost sensing and geospatial systems. This approach promotes transparency, fosters adoption by agronomists, and contributes to sustainable soil management, optimized fertilization practices, and improved decision-making under real-world agricultural constraints.

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture. EXAMPLE: Last Name, A. B. & Coauthor, C. D. (2026). Title of paper. In Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture and Brazilian Congress on Precision and Digital Agriculture.

Keywords.
Soil diagnosis, pH, electrical conductivity, nutrient availability, machine learning.

Introduction

Understanding how soil physicochemical properties influence nutrient availability is fundamental for supporting sustainable crop management and precision agriculture (Najdenko et al., 2024; Dewangan et al., 2024; Vullaganti et al., 2025). Soil chemical conditions directly affect nutrient solubility, mobility, and plant uptake, making them critical factors in determining soil fertility and crop productivity. Among these properties, soil pH and electrical conductivity (EC) are widely used indicators of soil chemical conditions and overall fertility (Xia et al., 2024; Almaw, 2022; Kim and Park, 2024). Although laboratory analyses provide reliable insights into soil nutrient dynamics, they are often costly, time-consuming, and impractical for large-scale or real-time monitoring. The increasing availability of portable soil sensors, particularly those capable of measuring pH and EC, has created new opportunities for rapid field diagnostics. However, translating these easily measurable indicators into meaningful assessments of soil nutrient dynamics remains a significant challenge.

In this context, data-driven approaches offer promising alternatives for interpreting complex relationships between soil attributes and nutrient availability. This study proposes an interpretable machine learning framework that leverages pH and EC measurements to characterize soil physicochemical profiles and support data-driven assessments of nutrient availability. The framework is developed using an open-access soil dataset containing both physical and chemical attributes, including textural fractions and multiple nutrient indicators. The proposed approach integrates Principal Component Analysis (PCA) for dimensionality reduction, K-Means clustering for soil profile identification, and decision tree models to extract interpretable rule-based relationships between soil properties and nutrient dynamics. Through this framework, micronutrients such as manganese (Mn), iron (Fe), and zinc (Zn), as well as key soil chemical indicators including magnesium (Mg), phosphorus (P), and calcium carbonate ($CaCO_3$), can be analyzed in relation to variations in pH and EC.

By combining interpretable machine learning with widely measurable soil indicators, the proposed framework provides a scalable and low-cost complementary tool for preliminary soil diagnostics. This approach enables rapid assessment of large agricultural areas, supports improved soil sampling strategies, and contributes to data-driven decision-making in precision agriculture while reducing the dependence on frequent laboratory analyses.

Soil Attributes

Soil attributes consist of a combination of physical and chemical properties that together determine soil behavior, fertility, and its capacity to sustain plant growth. Among the physical properties, soil texture (defined by the relative proportions of sand, silt, and clay) plays an important role in water retention, drainage, aeration, and nutrient storage within the soil matrix (Weil and Brady, 2017). Texture also influences several chemical processes by affecting the retention of organic matter and the soil's ability to retain and exchange ions.

Chemical properties exert a more direct influence on nutrient availability. Among these, soil pH and electrical conductivity (EC) are widely recognized as key indicators of soil chemical conditions. Soil pH strongly affects the solubility and bioavailability of several micronutrients, particularly iron (Fe) and manganese (Mn), whose availability decreases significantly under alkaline conditions. Electrical conductivity reflects the concentration of dissolved ions in the soil solution and provides an indirect indication of salinity and ionic balance, both of which influence nutrient mobility and plant uptake.

In addition to pH and EC, soil studies commonly consider other chemical attributes such as nutrient concentrations (e.g., P, K, Mg, Fe, Mn, and Zn), organic matter content, and chemical indicators such as calcium carbonate ($CaCO_3$). These attributes contribute to describing soil fertility conditions and help explain variations in nutrient dynamics across different soil environments (Weil and Brady, 2017).

Soil physicochemical properties are intrinsically interrelated due to shared pedogenetic and geochemical processes. For instance, soil texture influences organic matter retention, which in turn affects nutrient cycling and the soil's capacity to retain ions (Kogel-Knabner and Amelung, 2021; Catoni et al., 2016). Consequently, many soil attributes exhibit strong correlations and dependencies.

Understanding these relationships is essential when developing data-driven models for soil characterization. The presence of correlated attributes motivates the use of dimensionality reduction techniques and interpretable machine learning models capable of identifying patterns and extracting meaningful relationships between soil properties and nutrient availability.

Materials and Methods

This section describes the dataset, the attributes used, and the methodological approach adopted in this study. The workflow includes data preprocessing, unsupervised clustering, and the development of predictive models using interpretable machine learning techniques. The objective is to infer the soil physicochemical profile and nutrient availability based solely on pH and EC measurements. All steps were designed to ensure both predictive accuracy and model transparency, supporting the development of practical tools for real-time soil diagnostics.

All analyses conducted in this study adhere to a fully reproducible computational workflow. The entire pipeline was implemented in Python using open-source libraries such as pandas, scikit-learn, matplotlib, and seaborn. The code (Rodolfo, 2025) was developed and executed in Google Colab, ensuring a cloud-based, hardware-agnostic environment that facilitates replication without the need for local software installations.

Original Dataset

This study uses a dataset titled “Soil Data Grevena” (Tziachris et al., 2022; Tziachris, 2022), which contains 780 soil samples characterized by physicochemical laboratory analyses. The dataset comprises 17 columns, including one identifier column (ID) and 16 numerical attributes representing the soil’s physical, chemical, and textural properties: pH, Sand, Clay, Silt, EC, OM, $CaCO_3$, N-NO_3, P, K, Mg, Fe, Zn, Mn, Cu, and B.

Exploratory Data Analysis

An exploratory data analysis (EDA) (Tukey, 1977) was conducted to assess the statistical behavior, variability, and interrelationships among the soil’s physical and chemical attributes, supporting variable selection for subsequent modeling.

1. Descriptive Statistics and Outlier Detection

Descriptive analyses revealed high variability across key attributes, particularly Mn, Fe, K, and EC. Distribution plots and boxplots were used to inspect the data and identify potential outliers. No data points were removed, as all values were deemed agronomically plausible, reflecting natural variability arising from diverse geological origins and soil management practices. Retaining these extreme values was considered essential to capture real-world dynamics, particularly regarding nutrient availability and salinity effects.

2. Correlation Analysis

A Pearson correlation matrix (Pearson, 1901; Benesty et al., 2009) was constructed to examine the linear relationships between variables. The correlation heatmap revealed several key patterns that align with established agronomic principles:

- pH exhibited strong negative correlations with Fe ($r \approx -0.65$) and Mn ($r \approx -0.70$), confirming that acidic conditions favor their solubility, while alkaline conditions reduce their availability.
- pH was strongly correlated with $CaCO_3$ ($r \approx +0.80$), characteristic of calcareous soils.
- EC showed moderate positive correlations with K, Mg, and Na, reflecting its association with ionic concentrations in saline soils.
- Texture, particularly clay, correlated with nutrients like Mg and K, consistent with higher retention in fine-textured soils.

Data Preprocessing

Before the development of the machine learning pipeline, dataset was preprocessed to ensure data quality and improve model performance. First, the missing values were evaluated on the dataset. As no missing entries were detected, no imputation or data removal was necessary.

All numerical variables were then standardized using z-score normalization (Hastie et al., 2009), which centers the data by subtracting the mean and scales it by the standard deviation. This procedure was applied to ensure that variables with different units and scales (such as pH, EC, and nutrient concentrations) contributed equally to the modeling process, preventing bias toward features with larger numerical ranges.

Since the study focuses on regression tasks related to predicting nutrient availability rather than categorical classification, class balancing techniques were not required. However, to support the clustering step (K-Means), standardized data were essential to prevent dominance of high-magnitude features in distance calculations.

This preprocessing step ensured that the subsequent clustering and machine learning models were trained on consistent and properly scaled data.

Machine Learning Pipeline

The machine learning pipeline developed in this study was designed to analyze the influence of pH and EC on nutrient availability, while also enabling the characterization of distinct soil physicochemical profiles. The pipeline consists of three main stages: exploratory data analysis using Principal Component Analysis (PCA), unsupervised clustering with K-Means, and the construction of interpretable predictive models using decision trees.

1. PCA-Based Variance Analysis

Principal Component Analysis (PCA) (Jolliffe, 2002; Jolliffe and Cadima, 2016) was applied to explore the main sources of variability within the dataset, which includes physical (sand, silt, clay) and chemical attributes (pH, EC, $CaCO_3$, OM, and nutrients). The first two components (PC1 and PC2) jointly explain 42.71% of the total variance, insufficient for strict dimensionality reduction but adequate for structural analysis.

As shown in Figure 1, the first principal component (PC1) is strongly associated with chemical variables linked to acidity and alkalinity processes. Specifically, pH (0.47) and $CaCO_3$ (0.42) exhibit the highest positive loadings, whereas Mn (-0.40) and Fe (-0.43) demonstrate strong negative loadings, consistent with their tendency to precipitate under alkaline conditions. Hence, PC1 represents a chemical gradient controlled primarily by soil acidity and the availability of pH-sensitive micronutrients.

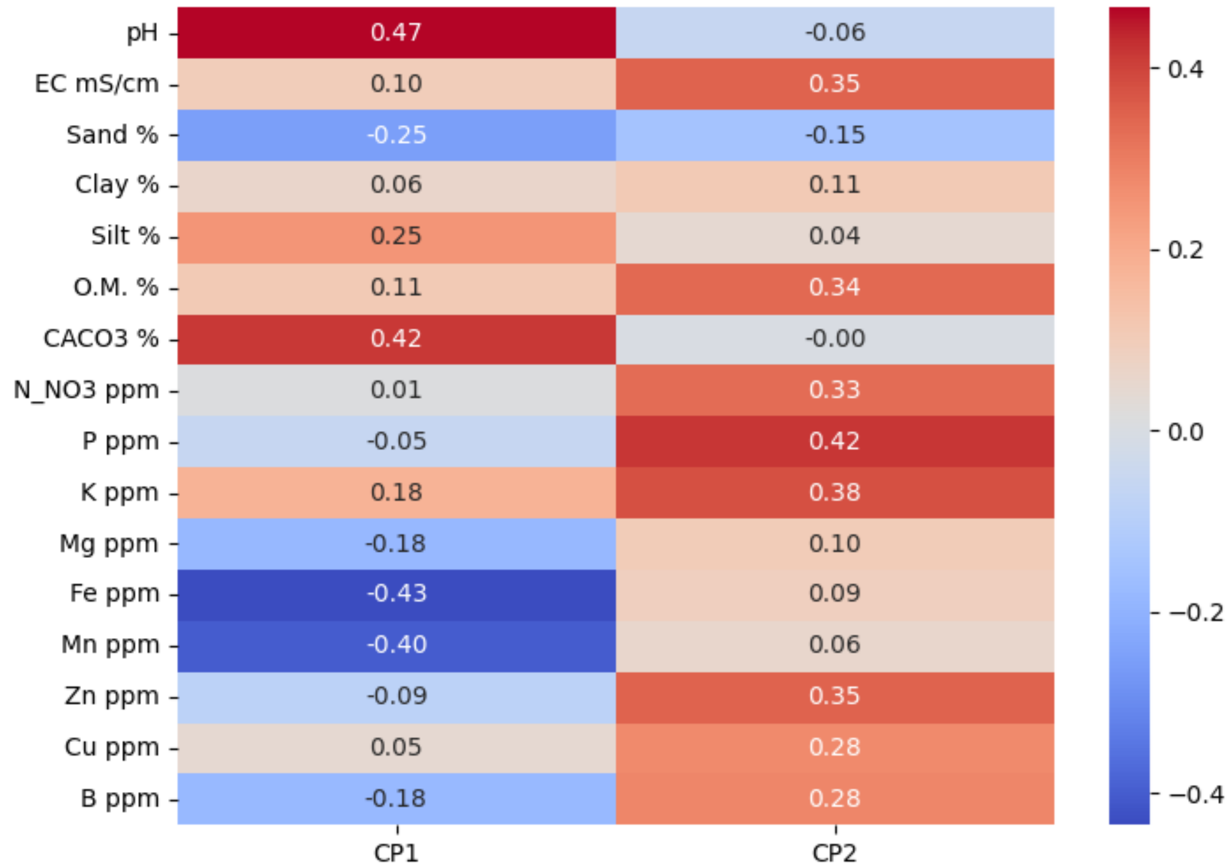

Figure 1 - PCA loading matrix.

The second principal component (PC2) captures variations mainly associated with EC (0.35), OM (0.34), N-NO_3 (0.33), as well as nutrients such as P (0.42), K (0.38), and Zn (0.35), elements frequently correlated with the immediate chemical fertility of soils. Although texture attributes showed minor contributions (sand -0.15; clay 0.11), their influence was secondary compared to the chemical variables, which exhibited stronger loadings. Consequently, texture was not a primary driver for clustering or rule extraction, which focused on chemical proxies measurable by in-field sensors (specifically pH and EC) and their relationships with nutrient dynamics.

Thus, PCA was primarily employed for visual validation and understanding of clustering patterns, serving as a precursor for subsequent unsupervised clustering. The choice of three clusters (k=3) was justified both visually, through clear data separation observed in the PCA space, and quantitatively, supported by the silhouette index (Rousseeuw, 1987) of 0.49. Testing alternative cluster numbers revealed that lower values (k=2) were insufficient for capturing transitional soil profiles, while higher values (k=4 or 5) led to fragmented clusters with decreased agronomic interpretability and lower silhouette scores. Therefore, k=3 represented the optimal balance between statistical coherence, clear differentiation of soil profiles, and practical agronomic utility.

2. Soil Profile Clustering with K-Means

Subsequently, the K-Means clustering algorithm (MacQueen, 1967; Jain, 2010) was applied to segment the data into distinct soil profiles. This method was selected due to its computational efficiency and its suitability for use with standardized data. Based on the PCA exploratory analysis and the silhouette index (0.49), the K-Means clustering algorithm was applied with k = 3, effectively segmenting the data into three distinct soil profiles. This choice balances statistical coherence and agronomic interpretability. Figure 2 presents the projection of the data onto the plane formed by the first two principal components, with points colored according to the clusters identified by K-Means. A clear separation between the groups is observed:

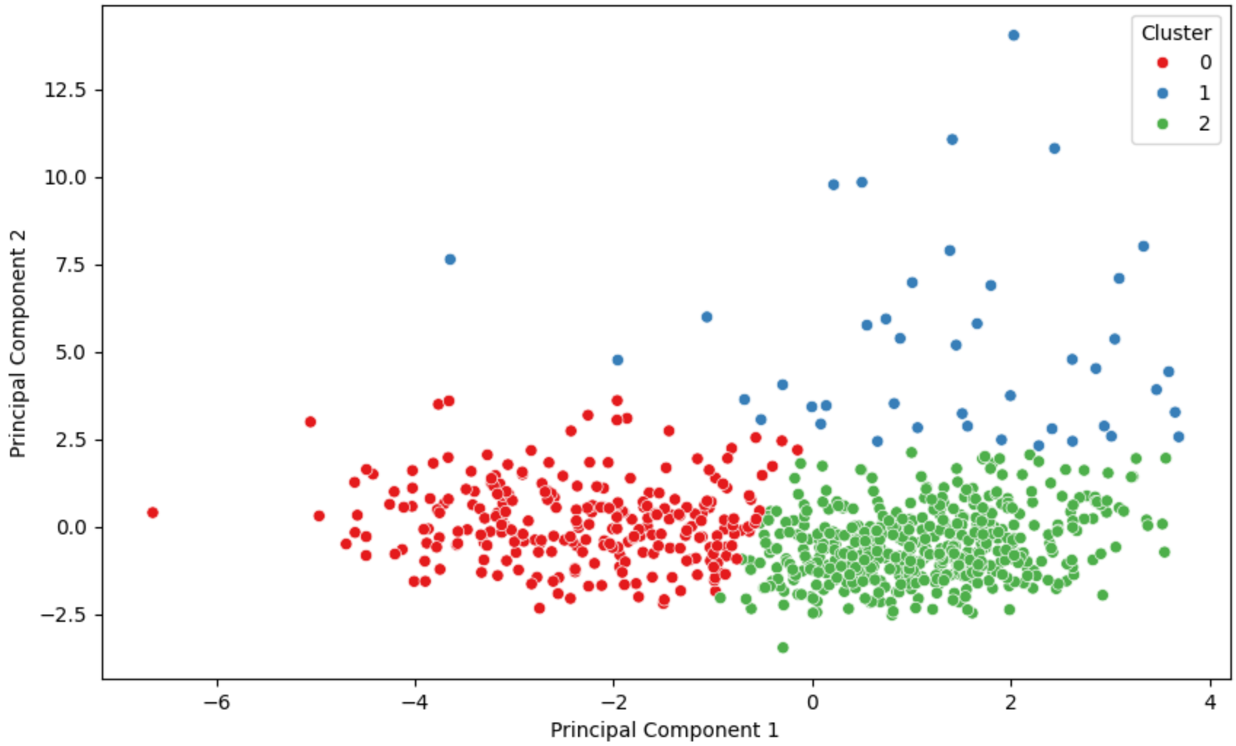

Figure 2 - PCA - Soil Clustering (K-Means).

- **Cluster 0** (**red**) includes samples from soils with lower pH and lower $CaCO_3$ content, but with higher concentrations of Fe and Mn, reflecting more acidic environments where these micronutrients are more available.
- **Cluster 2** (**green**) represents soils with higher pH, higher $CaCO_3$ content, and lower availability of Fe and Mn, which is consistent with alkaline conditions where these elements tend to precipitate.
- **Cluster 1** (**blue**) occupies a transitional region, exhibiting intermediate characteristics in terms of pH, micronutrient concentrations, and overall fertility.

This approach demonstrates that the data related to pH, EC, and nutrient concentrations are coherently organized in the PCA space, reinforcing the robustness of the clusters obtained. Therefore, the combination of PCA and K-Means not only provides a meaningful visual understanding of soil profiles but also establishes a solid foundation for subsequent modeling steps and the extraction of interpretable agronomic rules through decision trees.

3. Rule-Based Modeling with Decision Trees

Following the clustering stage, decision tree models were developed individually for each identified cluster in order to model the relationships between soil physicochemical attributes and the availability of key nutrients. This step enables the extraction of transparent and interpretable decision rules that describe how different soil properties influence nutrient dynamics within each soil profile.

The selected targets for modeling include essential nutrients and soil chemical indicators with agronomic relevance: Fe (ppm), Mn (ppm), P (ppm), Mg (ppm), Zn (ppm), and $CaCO_3$ (%). These elements were chosen due to their critical roles in soil fertility, plant nutrition, and chemical soil balance.

The input features for the decision tree models consist of the full set of physicochemical attributes, with the exception of pH and EC. These two variables were deliberately excluded because they were the primary drivers used in the clustering stage. Including them in the decision tree models could introduce redundancy and circular reasoning, as the clusters themselves were defined based on pH and EC ranges. By excluding these attributes, the trees are designed to capture how the remaining soil properties, such as texture components (sand, clay, silt), organic matter (OM), and other macro- and micronutrients ($N\text{-}NO_3$, K, Cu, and B), influence nutrient availability within each soil profile, independently of the clustering criteria. This methodological choice ensures that the decision rules extracted are not trivially dependent on pH and EC but rather reflect deeper relationships among the other soil characteristics.

The decision trees were built using the Classification and Regression Tree (CART) algorithm (Breiman et al., 1984; Quinlan, 1993; Zhogolev and Savin, 2020; Kebonye et al., 2023), with the maximum depth constrained to three levels. This constraint was intentionally applied to maintain simplicity, ensure model interpretability, and reduce the risk of overfitting. It represents a trade-off between capturing meaningful patterns and generating models that are practical and easily understandable for agronomic applications.

Each tree was trained separately within each cluster, allowing the models to reflect the unique relationships between soil attributes and nutrient availability specific to each soil profile. The resulting decision rules provide clear insights into which variables most strongly influence each nutrient and how these influences vary across different clusters.

This rule-based modeling approach directly supports agronomic decision-making by enabling the prediction of nutrient availability based on easily measurable soil properties. Furthermore, it facilitates the development of practical, data-driven recommendations for soil management tailored to the specific physicochemical contexts represented by each cluster.

Results

Based on the results obtained from the decision tree models for each cluster, heatmaps were generated to visualize the distribution of key nutrients, Fe, Mn, Zn, Mg, P, and $CaCO_3$, according to pH and EC ranges. These visualizations synthesize the decision paths identified by the trees, highlighting how nutrient availability varies across different soil physicochemical conditions. This approach was instrumental in formulating diagnostic rules for soil fertility management.

The heatmaps, exemplified in Figure 3 for Fe, clearly show how lower pH conditions are associated with higher Fe availability, while alkaline conditions tend to reduce it. Similar patterns were observed for Mn and Zn, whose availability declines in soils with higher pH and EC levels. Conversely, elements like P and Mg showed different dynamics, often associated with interactions involving soil texture and other nutrient levels.

Therefore, the combination of decision trees and heatmaps not only enhanced the interpretability of the models but also provided a robust basis for deriving agronomically meaningful rules, supporting soil management strategies without the exclusive reliance on laboratory testing.

Based on the combined analysis of the decision tree models and the heatmaps generated for each nutrient, it was possible to derive a set of robust diagnostic rules directly grounded in the patterns observed in the data. These rules summarize how different ranges of pH and EC influence the availability of key soil nutrients, enabling the generation of agronomic alerts that are both interpretable and practically applicable. Below are the validated rules, accompanied by the evidence derived directly from the dataset.

- If (pH < 5.5) and (EC < 0.3) → Likely deficiency of P, Ca, and Mg. Evidence: The mean values of Mg (912 ppm), P (14.77 ppm), and $CaCO_3$ (1.07%) are the lowest in the entire dataset. In contrast, there is an accumulation of Fe (74.9 ppm) and Mn (46.68 ppm), which is typical of acidic soils.
- If (pH > 7.5) and (EC > 1.0) → Possible deficiency of Fe, Mn, and Zn. Evidence: The mean values of Fe (15.77 ppm), Mn (6.64 ppm), and Zn (1.23 ppm) are observed in the

subgroup with (pH > 7.5) and EC between 1.0 and 2.5. In the few cases where EC exceeds 2.5, the values are even lower (Fe = 11.28 ppm, Mn = 8.89 ppm, Zn = 2.29 ppm).

- If (EC > 2.5) and (any pH) → Risk of salinization and nutritional imbalance. Evidence: Although rare (only 2 samples with medium pH and 1 with high pH), this group shows clear signs of imbalance, such as extremely high Mg levels (up to 1868 ppm) and potential impacts on the uptake of other nutrients.

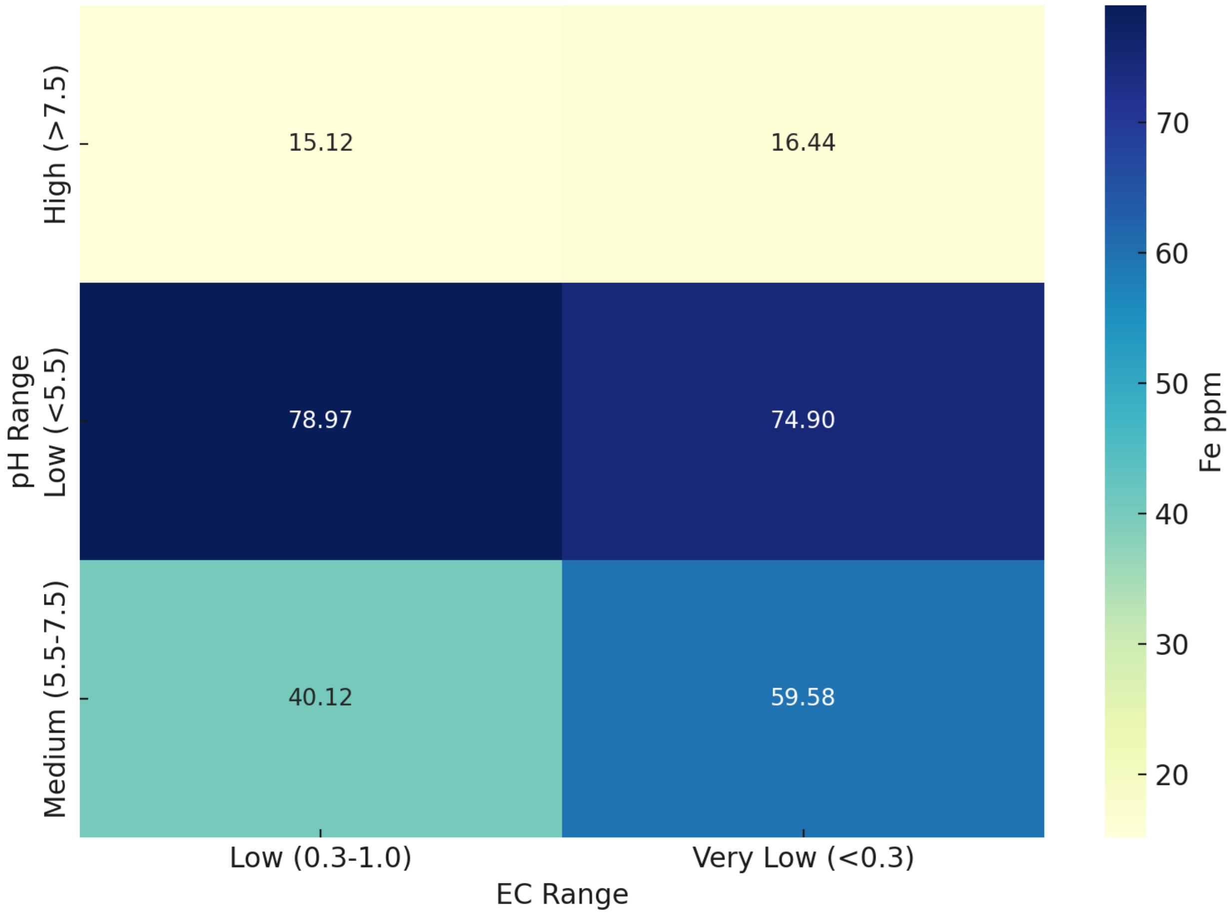

Figure 3 - Fe ppm distribution by pH e EC range.

It is important to note that the attributes N-NO_3, K, Cu, and B were not selected in the final decision rules. This outcome is consistent with the patterns identified throughout the modeling pipeline. In the PCA stage, these attributes exhibited low loadings on the principal components, indicating minimal contribution to the main sources of variance in the dataset. Subsequently, during the clustering process, the soil profiles were primarily defined by chemical gradients related to pH, EC, $CaCO_3$, and key micronutrients, further reducing the relative importance of N-NO_3, K, Cu, and B. Finally, in the decision tree modeling stage, these attributes did not provide sufficient impurity reduction to be selected as splitting nodes, especially considering the model constraint of a maximum depth of three levels.

Conclusion

This study demonstrated the effectiveness of combining pH and electrical conductivity (EC) sensors with interpretable machine learning techniques to profile soil physicochemical conditions and predict nutrient availability. The integration of Principal Component Analysis (PCA) with K-Means clustering successfully identified three distinct soil profiles, driven primarily by variations in pH, EC, $CaCO_3$, and key micronutrients such as Fe and Mn. Decision tree models provided transparent, interpretable rules clearly describing how different pH and EC ranges influence nutrient dynamics.

Rather than fully replacing laboratory-based analyses, the proposed approach serves as a powerful complementary tool for rapid, in-field soil assessment. It allows efficient characterization of large agricultural areas, enabling scalable and cost-effective diagnostics and providing actionable insights for precision agriculture. This sensor-driven, rule-based framework thus enhances soil monitoring, reduces the dependency on frequent laboratory testing, and supports more efficient and sustainable crop management decisions.

Acknowledgments

This work has been fully funded by the project IA na Borda supported by the Center for Embedded Devices and Research in Digital Agriculture (CEDRA) of SENAI-RS, with financial resources from the PPI IoT/Manufatura 4.0 / PPI HardwareBR of the MCTI, grant number 056/2023, signed with EMBRAPII.

References

ALMAW, C. D. Determination and correlation of pH and electrical conductivity of Assosa Agricultural Research Center sites soil. **Asian Soil Research Journal**, v. 6, n. 3, p. 6–10, 2022.

BENESTY, J.; CHEN, J.; HUANG, Y.; COHEN, I. Pearson correlation coefficient. In: Noise Reduction in Speech Processing. **Springer**, 2009. p. 1–4.

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. Classification and Regression Trees. Belmont, CA, USA: **Chapman & Hall**, 1984.

CATONI, M.; D'AMICO, M. E.; ZANINI, E.; BONIFACIO, E. Effect of pedogenic processes and formation factors on organic matter stabilization in alpine forest soils. **Geoderma**, v. 263, p. 151–160, 2016.

DEWANGAN, S. K.; KASHYAP, N.; TIGGA, D.; TIRKEY, N. The role of physico-chemical properties in soil functionality: a literature review. **EPRA International Journal of Research & Development (IJRD)**, v. 9, p. 167–173, 2024.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The Elements of Statistical Learning. New York, USA: **Springer**, 2009.

JAIN, A. K. Data clustering: 50 years beyond K-means. **Pattern Recognition Letters**, v. 31, n. 8, p. 651–666, 2010.

JOLLIFFE, I. T. Principal Component Analysis. 2. ed. New York, USA: **Springer**, 2002.

JOLLIFFE, I. T.; CADIMA, J. Principal component analysis: A review and recent developments. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 374, n. 2065, p. 20150202, 2016.

KEBONYE, N. M.; AGYEMAN, P. C.; BINEY, J. K. M. **Optimized modelling of countrywide soil organic carbon levels via an interpretable decision tree**, 2023.

KIM, H. N.; PARK, J. H. Monitoring of soil EC for the prediction of soil nutrient regime under different soil water and organic matter contents. **Applied Biological Chemistry**, v. 67, n. 1, 2024.

KOGEL-KNABNER, I.; AMELUNG, W. Soil organic matter in major pedogenic soil groups. **Geoderma**, v. 384, p. 114785, 2021.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: **Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability**. Berkeley, CA, USA: University of California Press, 1967. p. 281–297.

NAJDENKO, E.; LORENZ, F.; DITTERT, K.; et al. Rapid in-field soil analysis of plant-available nutrients and pH for precision agriculture: a review. **Precision Agriculture**, v. 25, p. 3189–3218, 2024.

PEARSON, K. On lines and planes of closest fit to systems of points in space. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, v. 2, n. 11, p. 559–572, 1901.

QUINLAN, J. R. C4.5: Programs for Machine Learning. San Mateo, CA, USA: **Morgan Kaufmann**, 1993.

RODOlFO, T. Sensor-Based Soil Profiling with Machine Learning. GitHub Repository. Disponível em: https://github.com/tacrodolfo/Sensor-Based-Soil-Profiling-with-ML. Acesso em: 20 jun. 2025.

ROUSSEEUW, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, v. 20, p. 53–65, 1987.

TZIACHRIS, P. Soil data Grevena. Mendeley Data, v. 1, 2022. Disponível em: https://doi.org/10.17632/r7tjn68rmw.1.

TZIACHRIS, P.; ASCHONITIS, V.; METAXA, E.; BOUNTLA, A. A soil parameter dataset collected by agricultural farms in northern Greece. **Data in Brief**, v. 43, p. 108408, 2022.

VULLAGANTI, N.; RAM, B.; SUN, X. Precision agriculture technologies for soil site-specific nutrient management: a comprehensive review. **Artificial Intelligence in Agriculture**, v. 15, 2025.

WEIL, R. R.; BRADY, N. C. The Nature and Properties of Soils. 15. ed. Upper Saddle River, NJ, USA: **Pearson**, 2017.

XIA, X.; FENG, J.; ZHANG, H.; XIONG, D.; KONG, L.; SEVIOUR, R.; KONG, Y. Effects of soil pH on the growth, soil nutrient composition, and rhizosphere microbiome of Ageratina adenophora. **PeerJ**, v. 12, e17231, 2024.

ZHOGOLEV, A.; SAVIN, I. Soil mapping based on globally optimal decision trees and digital imitations of traditional approaches. **ISPRS International Journal of Geo-Information**, v. 9, n. 11, p. 664, 2020.

KEBONYE, N. M.; AGYEMAN, P. C.; BINEY, J. K. M. Optimized modelling of countrywide soil organic carbon levels via an interpretable decision tree. **Smart Agricultural Technology**, v. 3, art. 100106, 2023.