

Analyzing Soil Health Indicators for Agricultural...



Analyzing Soil Health Indicators for Agricultural Sustainability Using Machine Learning

Prepared By

Name: Alaa Mohammed Jawish

ID No.: 221002192

Name: Khadija Ehab Sadeq

ID No.: 221002423

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

At

College of Informatics

Midocean University

UAE

Supervisor Name: Dr. Mohamed Eassa

College of Informatics, Midocean University, Comoros

Email:

Alaa0emad@gmail.com

kad2005.ias@gmail.com

Acknowledgements

We would like to thank our supervisor, Dr. Mohamed Essa, for his continuous guidance and valuable feedback throughout this research. His expertise in both artificial intelligence and Soil health studies has greatly influenced the direction of this work. His comments challenged us to think more critically and approach the research from a scientific perspective rather than a purely technical one.

We are also thankful to the faculty members at the College of Informatics, Midocean University, for providing a supportive academic environment that inspired independent thinking and research development.

Finally, we would like to express our appreciation to our families for their patience, encouragement, and support during this journey.

Abstract

Due to the rising need for food and the developing problems brought on by climate change and the depletion of natural resources, agricultural sustainability has emerged as a global issue. In order to sustain agricultural productivity, soil health is crucial because it promotes plant growth, nutrient availability, water retention, and ecosystem stability. Traditional techniques for evaluating soil health are frequently costly, time-consuming, and challenging to consistently implement over broad agricultural regions. In addition, agricultural systems produce a growing volume of data that cannot be effectively analysed using traditional methods as digital technologies continue to advance. Because of this, machine learning has emerged as a useful tool for evaluating complicated soil data and assisting with more trustworthy agricultural decision-making. This study aims to analyze multiple soil health indicators using machine learning techniques in order to identify the most influential indicators related to agricultural sustainability and provide data-driven support for sustainable soil management.

Contents

Acknowledgements	1
Abstract	2
Chapter 1: Introduction	4
1.1 Overview	4
1.2 Problem Definition	5
1.3 Thesis Objectives	6
1.4 Thesis Contribution	7
1.5 Thesis Organization	7
Chapter 2: Background and Related Work	8
2.1 Introduction	8
2.2 Background	9
2.3 Related work	10
2.4 Related Work Summary	13
References	15

Chapter 1: Introduction

1.1 Overview

The need for agricultural sustainability has grown to a significant issue globally as food demand has risen and climate change and natural resource depletion have become a concern. Soil health is one of the most crucial aspects to ensure the productivity of agriculture, as it affects various processes such as crop growth, nutrient availability, water retention and ecosystem stability. Hence soil health is a key ingredient of sustainable agriculture over the years. With growing demands for sustainable agriculture, the need for ongoing soil monitoring and soil management has become even more critical over the past few years(1).

Soil health is the capacity of soil to sustain plants, animals and humans, and function in an environmentally healthy way as a living biota(2). Commonly used measures of soil health include: soil moisture, pH, organic carbon, nutrients, electrical conductivity, microbial activity and soil structure(1). These indicators give information of the state of soils and their capacity to support agriculture(2).

Soil health assessment is often done traditionally using field observations and laboratory testing methods. These have the advantage of being accurate but they are time consuming, expensive and hard to apply to large agricultural areas on a continuous basis(1). In addition, sensor use, remote sensing and precision agriculture systems have led to tremendous amounts of agricultural data that need to be analyzed and interpreted effectively. Such data can be very difficult to analyse using traditional approaches and can delay timely agricultural decisions(3).

The advent of Artificial Intelligence (AI) and Machine Learning (ML) has opened up new possibilities for soil analysis and agricultural management(3). Machine learning algorithms can be used to analyze extensive agricultural data, uncover intricate connections between soil markers, and make precise soil

condition predictions(1). Models like Random Forest, Support Vector Machines, and Neural Networks have demonstrated potential success in soil data analysis and in decision-making processes related to agriculture(3).

Although soil science has made significant strides towards using machine learning to benefit agriculture, much of the existing literature is limited to specific soil characteristics, including soil fertility, soil moisture, or soil texture(1). There is relatively little focus on the holistic examination of many soil health indicators and how they all contribute to agricultural sustainability(1)(3). Furthermore, the relative weightage of various soil indicators is difficult to be determined because of the complexity of soils and the interaction between the various physical, chemical and biotic properties of soils(2).

This research aims to examine the relationships between the various soil health indicators, determine the most significant indicators and facilitate sustainable agricultural management practices by applying machine learning analysis to soil health indicators.

1.2 Problem Definition

Soil health evaluation is critical for sustainable farming in order to keep the productivity of agriculture(2). The traditional way of analyzing the health of soils has been to use laboratory analysis and field measurements of various physical, chemical and biological measurements. While important, these methods also are time-consuming, expensive, and hard to apply on a continuing basis over large agricultural areas(1).

Soil health is also impacted by several indicator factors working together with soil moisture, pH, organic matter, nutrient availability and microbial activity(2). All of these interactions are complex, which makes it difficult to accurately

monitor soil conditions using traditional analytical methods(1). With the growing development of precision agriculture, sensor technologies and remote sensing systems, there is an increasing volume of soil data with the need for more efficient and intelligent analysis(1)(3).

Although several studies have been performed on the applications of machine learning in agriculture, most of them have been developed to address a single soil property and not for the simultaneous evaluation of several soil health indicators. In addition, it is difficult to identify what are the most influential indicators that can contribute to agricultural sustainability(1)(3)(4).

From the discussion above, it is obvious that the main challenge in this thesis is the absence of an integrated machine learning solution that can analyze multiple soil health indicators, determine their relative importance and enable the implementation of sustainable agriculture decisions.

1.3 Thesis Objectives

This research aims to analyze soil health indicators using machine learning techniques and support agricultural sustainability through improved soil management practices.

- o The objectives of this study are:
 - To identify and investigate the key soil health indicators associated with agricultural sustainability.
 - To collect, preprocess, and analyze soil data obtained from available datasets.
 - To apply machine learning techniques to analyze the relationships among soil health indicators.
 - To identify the most influential soil health indicators affecting agricultural sustainability.

- To evaluate and compare the performance of different machine learning models for soil health analysis.
- To provide data-driven insights that support sustainable agricultural decision-making and soil management practices.

1.4 Thesis Contribution

This thesis is expected to contribute to the fields of sustainable agriculture and artificial intelligence through the following:

- Developing a machine learning-based framework for analyzing multiple soil health indicators.
- Identifying the relative importance of different soil health indicators and their impact on agricultural sustainability.
- Providing a comparative evaluation of machine learning techniques used for soil health analysis.
- Supporting farmers and agricultural decision-makers through data-driven recommendations for sustainable soil management.
- Contributing to the growing application of artificial intelligence in precision agriculture and environmental sustainability.

1.5 Thesis Organization

The rest of this thesis is organized as follows.

Chapter 2 discusses the theoretical background and relevant research areas.

Chapter 3 presents the proposed methodology.

Chapter 4 analyzes the experimental results.

Chapter 5 concludes the study and suggests directions for future work.

Chapter 2: Background and Related Work

2.1 Introduction

This section presents a brief introduction and background information of previous studies about Soil health assessment and machine learning applications in agriculture sustainability. In this chapter we will show some of the most important concepts of soil health, soil health indicators, and agriculture sustainability. In addition to the machine learning methods that have been used for soil analysis and agriculture decision-making.

Soil fertility management is considered one of the most important factors for maintaining sustainable agricultural production. It helps the farmers to maintain the ability of the soil to provide the appropriate amount of nutrients to the plants when it is needed. Machine learning techniques have been applied for soil analysis recently. These techniques can process large amounts of agricultural data and help in making decisions related to soil and crop management (6).

However, many existing studies focus on some specific soil properties such as soil fertility, soil moisture, or nutrient prediction. In addition, different machine algorithms have been used with different datasets and evaluation methods, making it difficult to find the best approaches for soil health assessment.

Therefore, this chapter discusses the background concepts of soil health, agricultural sustainability, and machine learning techniques that are used in agriculture. In addition, it reviews previous work in detail in terms of methodologies, datasets used, and evaluation techniques in order to identify their strengths, limitations, and potential research gaps related to soil health analysis.

2.2 Background

Soil health plays a significant role in supporting plants growth, maintains environmental quality, and improves sustainable agriculture production. Healthy soil improves crop productivity, nutrient cycling, water availability, and ecosystem stability. Overall, soil health is important not just for crops and food production but also for protecting the environment and helping mitigate climate change.

The functions of healthy soil are critical for sustaining agriculture systems and protecting the environment. It provides plants with essential nutrients, improves water filtration and storage, promotes biological activity, and helps maintain a stable soil structure (2).

Soil health is influenced by many physical, chemical, and biological factors. Physical properties include soil structure, soil texture, and water holding capacity. The chemical properties include soil pH, potential pollutants, organic carbon, nutrient availability, and electrical conductivity. Biological properties also play an important role in maintaining soil functions. Examples of biological properties include microbial biomass and microbial activity. Based on this there is no single indicator that can fully describe soil health. For this reason, soil health assessment usually depends on evaluating a combination of different indicators to provide a more comprehensive assessment (2).

Soil health is evaluated using a set of indicators including the physical, chemical, and biological condition of the soil, which are used to provide information about the soil ability to support plants growth and maintain ecosystem services. Different indicators are combined to provide accurate assessment of soil condition (2).

Sustainable agriculture aims to meet the growing demand for food while protecting natural resources. It focuses on improving agricultural productivity through the efficient use of soil and other resources without compromising the needs of future generations. Sustainable agriculture now has progressively adopted advanced technologies. These technologies include precision agriculture, remote sensing, and the internet of things (IOT), to improve farm management and support informed decision making (3).

The management of agricultural systems has been changed because of the rapid development of digital technologies. Modern farming generates large amounts of data from sensors, drones, satellites, and other smart devices. Traditional methods are no longer sufficient for analyzing these data. Thus, machine learning has emerged as an effective approach for processing agriculture data, identifying hidden patterns, and supporting accurate and timely decision making in sustainable agriculture (3).

Several machine learning algorithms have been applied in agricultural research to analyze complex datasets and improve prediction accuracy. These algorithms examples are Decision Tree, Random Forest, Support Vector Machine, multilayer perceptron (MLP) neural networks, Extreme Gradient Boosting (XGBoost), and Artificial Neural Networks (ANN). These algorithms have been used in different agricultural applications, such as soil analysis, crop yield prediction, disease detection, and precision farming. These algorithms have the ability to analyze large datasets and identify complex relationships, making them valuable tools for supporting sustainable agriculture management (3). In addition, the combination of ML with the IoT is revolutionizing smart agriculture. In recent years, researchers have investigated the use of these technologies to automate the agricultural operations, such as monitoring soil conditions, weather, irrigation management, and pest control (3).

2.3 Related work

1. Several studies have applied machine learning techniques in agriculture and soil health analysis. These studies focused on different soil properties. While the results were promising, most of these studies focused on specific soil

parameters rather than analyzing multiple soil health indicators together. This makes it important to review these studies to understand the existing research gap and current research trends.

Chandra et al. (2023) discussed the implementation of an explainable AI (XAI) model for soil fertility prediction by using Random Forest classification. In this study the LUCAS 2018 Topsoil dataset was used. It used several soil parameters such as pH, organic carbon, nitrogen, electrical conductivity, phosphorus, and potassium. The study combined k- Means clustering, Random Forest, and SHAP to explain the result. The result showed an accuracy of 97.02%, which shows the effectiveness of the explainable machine learning for soil fertility assessment. However, this study focused on soil fertility prediction only, and did not analyze multiple soil health indicators(7).

2. Sheeba et al. (2022) evaluated micro and macro nutrients of the soil in four major districts of Tamil Nadu by automating the farms with IoT sensors. They processed the data through a machine learning algorithm for enhancing the soil quality. They applied the Extreme Learning Machine (ELM) algorithm to determine the level of the fertility of the soil. The study used several parameters like pH, organic carbon, nitrogen, electrical conductivity, phosphorus, potassium and micronutrients. The results showed that the methodology can be useful to identify the fertility indices of nutrients. However, the study focused only on soil fertility and micronutrients classification, and did not analyze the relationships between soil health indicators for agricultural sustainability(6).

3. Davis et al 2023 examined patterns across farms in the Palouse River area. Instead of short-term snapshots, they tracked shifts like declining organic matter and rising acidity under different farm routines. Through combining standard soil checks with ideas about natural system recovery, their framework measured changes in land function year after year. While one path stuck to yearly planting without tilling, another followed typical plow-based methods for contrast. Results showed untouched ground paired with thoughtful crop choices rebuilt damaged layers, slowed runoff loss, while also buffering weather extremes better than usual approaches.

One big issue with this work lies in how tightly it ties itself to just one place - the Palouse River watershed - along with its local ecological models, making outcomes harder to apply in dry areas or places with very different soils. Instead of building hands-on digital systems that pull live data or process signals as they come, the team leaned heavily into modeled ideas about ecosystem recovery. What sets our effort apart - "Analyzing Soil Health Indicators for Agricultural Sustainability" - is its broader design, built to handle many kinds of soil information across locations, while pointing out working code and methods useful for judging long-term land health when climates shift(8).

4.From Brichi et al. (2023), an exploration unfolds on how organic inputs reshape soil health alongside farm resilience. Instead of random picks, they leaned on 81 strong academic works pulled neatly from Web of Science's broad collection. Through sharp data mapping tools mixed with deep narrative review, patterns emerged over time about tossing natural matter into farmland routines. What stood out was a close look at numbers - how living soils reacted differently when fed compost compared to synthetic feeds - tracking shifts across texture, nutrients, and underground life. Surprising gains emerged when farmers kept adding organic matter - soil grew stronger, richer. Nutrients stayed put more reliably over time. Crop results shifted noticeably too. Grain legumes responded well, producing more. Wheat saw improvements alongside maize, rice stepping up yields just as much. Each cereal benefited in measurable ways. Brichi et al. (2023) doesn't match up when it comes to hands-on analysis of actual field results, something others in this discussion do include. While their view looks back with some depth, current numbers tied to core soil measurements stay missing. What sets apart the research here is stepping past old summaries pulled from earlier reports - instead, fresh evaluations of true soil conditions were carried out. Direct methods checked live indicators, then shaped insights aimed at lasting land use without leaning on recycled academic references(9).

5.One 2020 paper in Heliyon looked at how changing farmland shapes certain traits in dirt, aiming to judge its health. Instead of guessing, the scientists pulled

hard numbers from many spots across a wide area, checking hundreds of farm sites. To make sense of it all, they leaned on number-crunching tools - mixing index scores with pattern-spotting methods - to trim down piles of data. The Soil Quality Index guided part of the process, while another method called PCA helped shrink a full collection into just essential pieces. They kept an eye on things like weight per handful, balance between carbon and nitrogen, plus minerals ready for plants to grab. Fields got scored based on those signs, revealing shifts you might miss otherwise. What stood out was that using fewer but smarter measurements - with SQI as a compass - can catch slow damage more clearly than expected.

Yet this study centers only on physical and chemical measures, skipping biological ones - microbial life or respiration rates - that matter just as much for true soil condition. Though it leans into classic stats tools such as PCA, there's no smooth digital tool built alongside, nothing practical a farmer might actually use day to day. Our work titled "Analyzing Soil Health Indicators for Agricultural Sustainability" pushes further, blending lab-based findings with real-world tech access in one full picture(10).

2.4 Related Work Summary

Summary of Related Works in Soil Health Indicators To summarize, we can infer from the conducted survey that the development is on the path from localised statistical models to various ML based models for soil health indication, such as studies proposed by Chandra et al. (2023) and Sheeba et al. (2022) where high accuracy is obtained using Random Forest and Extreme Learning Machine for purely predicting soil fertility and nutrient classification. In contrast, ecological and meta-analysis studies that are used to understand organic amendments and long term management issues are found in the work of Davis et al. (2023) and Brichi et al. (2023).

But these methods do have some spatial restrictions or don't provide direct soil sampling based empirical data.

Another work by Akinde et al. (2020) explains efficient reduced data representation models like compressed data scoring models (SQI and PCA) but without considering any biological indicator and practical technical applications

to end users. In this proposed work “Analyzing Soil Health Indicators for Agricultural Sustainability,” we are aiming to connect these loose ends. Unlike isolated, parameter specific and literature review based research, we present a consolidated analysis with a balanced selection of physical, chemical and biological indicators of soil health considering different agro-ecological conditions and also emphasize a practical and technology enabled user solution.

<i>Ref</i>	<i>Author</i>	<i>Year</i>	<i>Technique Used</i>	<i>Accuracy</i>	<i>Dataset</i>	<i>Comparison with This Research</i>
[7]	Chandra et al.	2023	k-means clustering Random Forest classifier, Shap	~97%	LUCAS 2018 topsoil dataset	Limited to fertility prediction
[6]	Sheeba et al.	2022	Extreme Learning machine	Not explicitly reported	1000 soil sample (Tamil Nadu)	Limited to fertility and micronutrient classification.
[8]	Davis et al.	2023	Adaptive management modeling & Ecological resilience theory	Not explicitly reported	Palouse River watershed environmental data	Limited to one region; lacks a data-driven IT framework.
[9]	Brichi et al.	2023	Bibliometric analysis & Systematic qualitative synthesis	Not explicitly reported	81 scientific papers from Web of Science	Purely a literature review; lacks raw empirical field data.
[10]	Akinde et al.	2020	Soil Quality Index (SQI) & Principal Component Analysis (PCA)	High sensitivity validation	Regional land-use sampling (hundreds of farm plots)	Ignores biological indicators; lacks direct software deployment.

References

- 1) [Gonzalez, F.et al, 2021] Freddy A. Diaz-Gonzalez, Jose Vuelvas, Carlos A. Correa b, Victoria E. Vallejo c, D. Patino " Machine learning and remote sensing techniques applied to estimate soil indicators – Review"
<https://doi.org/10.1016/j.ecolind.2021.108517>
- 2) [Topa, D.et al. 2025] Denis-Constantin Topa, Sorin Căpsună, Anca-Elena Calistru and Costică Ailincăi" Sustainable Practices for Enhancing Soil Health and Crop Quality in Modern Agriculture: A Review" <https://doi.org/10.3390/agriculture15090998>
- 3) [Valencia , J.et al, 2025] Juan Botero-Valencia, Vanessa García-Pineda ,Erick Reyes-Vera , Mateo Mejia-Herrera , Alejandro Valencia-Arias, Ruber Hernández-García, and Jackeline Valencia" Machine Learning in Sustainable Agriculture: Systematic Review and Research Perspectives"
<https://doi.org/10.3390/agriculture15040377>
- 4) [Sanjay Motia, and SRN Reddy, 2021] "Exploration of machine learning methods for prediction and assessment of soil properties for agricultural soil management: a quantitative evaluation
doi:10.1088/1742-6596/1950/1/012037

- 5) [Awais, M.et al.,2023] Muhammad Awais, Syed Muhammad Zaigham Abbas Naqvi, Hao Zhang, Linze Li, Wei Zhang, Fuad A. Awwad, Emad A. A. Ismail, M. Ijaz Khan, Vijaya Raghavan and Jiandong Hu*" AI and machine learning for soil analysis: an assessment of sustainable agricultural practices"
<https://doi.org/10.1186/s40643-023-00710-y>

- 6) [Sheeba, B.et al. 2022] T. Blesslin Sheeba,1 L. D. Vijay Anand, Gunaselvi Manohar, Saravana Selvan, Bazil Wilfred, K. Muthukumar, S. Padmavathy, P. Ramesh Kumar,and Belete Tessema Asfaw " Machine Learning Algorithm for Soil Analysis and Classification of Micronutrients in IoT-Enabled Automated Farms"
<https://doi.org/10.1155/2022/5343965>

- 7) [Chandra, H. et al,2023]HARSHIV CHANDRA , PRANAV M. PAWAR ELAKKIYA R., TAMIZHARASAN P S, RAJA MUTHALAGU AND ALAVIKUNHU PANTHAKKAN
"Explainable AI for Soil Fertility Prediction"

- 8) [Davis, E. et al., 2023] Eleanor Davis, Timothy Bowles, Lauren Hallett, Franklin Egan, and Jennifer Blesh "Integrating soil health indicators with ecological resilience theory to assess agricultural sustainability" <https://doi.org/10.1002/fee.2594>

- 9) [Brichi, M. et al., 2023] Mohamed Brichi, Ahmed Al-Mansoori, and Fatima Al-Suwaidi "The role of organic amendments and residues in enhancing soil health indicators for sustainable crop production: A bibliometric review" <https://doi.org/10.1111/sum.12892>

- 10) [Akinde, B. et al., 2020] B.P. Akinde, A.O. Olakayode, D.J. Oyedele, and F.O. Tijani "Selected Physical and Chemical Properties of Soil under Different Agricultural Land-Use Types" <https://doi.org/10.1016/j.heliyon.2020.e05090>

- 11)

