IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 2, February 2025

Intelligent Soil Health Prediction for Regenerative Farming - A Machine Learning Approach to Sustainable Agriculture

Bhaavik Bhavesh Ashar
Student, Department of Msc. IT
Nagindas Khandwala College, Mumbai, Maharashtra, India
bhavikashar3@gmail.com

Abstract: *Soil health is a critical factor in ensuring sustainable agriculture and environmental conservation. This project aims to develop an intelligent soil health prediction model using machine learning techniques to optimize soil classification accuracy. By leveraging key environmental parameters such as soil biodiversity, carbon sequestration, water retention, and overall fertility, we propose a data-driven approach to categorize soil health. Various machine learning models, including Random Forest, Gradient Boosting, Logistic Regression, SVM, and KNN, are evaluated to achieve high classification accuracy. Additionally, model performance is analyzed using key metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Train R^2, Test R^2, and Adjusted R^2. Visualizations like confusion matrices, scatter plots, and model comparison charts will be used to support sustainable decision-making for farmers.*

The research emphasizes the importance of feature selection and hyperparameter tuning to enhance predictive performance. By analyzing the impact of different soil attributes on classification outcomes, we identify the most influential parameters in determining soil health. The model is trained on a diverse dataset that includes real-world and synthetically generated agricultural data, ensuring robustness across varying environmental conditions. Cross-validation techniques are applied to prevent overfitting and improve generalization, making the model adaptable to different soil types and regions.

The findings of this study contribute to precision agriculture by offering an AI-powered tool that assists farmers and agricultural experts in monitoring and managing soil health effectively. The integration of machine learning in soil classification enhances efficiency, reduces manual effort, and promotes environmentally sustainable farming practices. By providing accurate and timely soil health predictions, this approach supports improved crop yield, resource conservation, and long-term agricultural sustainability.

Keywords: Soil health

I. INTRODUCTION

1.1 Background & Motivation

The health of soil plays a crucial role in regenerative farming, impacting crop productivity, water retention, and carbon sequestration. Traditional soil testing methods are time-consuming and expensive, often limiting farmers' access to real-time soil health insights. With advancements in machine learning, predictive modeling offers an efficient and scalable approach to soil health assessment, enabling data-driven agricultural practices.

1.2 Objectives

A) Develop a Smart Classification Model for Soil Health:

Utilize key environmental parameters such as soil biodiversity, carbon sequestration, and water retention to predict soil health categories.Apply machine learning algorithms (Random Forest, Gradient Boosting, Logistic Regression, SVM, KNN) to optimize classification accuracy.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 2, February 2025

B) Enhance Model Performance for Sustainable Decision-Making:

Compare different models using performance metrics (MSE, RMSE, MAE, Train R^2, Test R^2, Adjusted R^2) to ensure robust predictions.Visualize results through confusion matrices, scatter plots with perfect fit lines, and model comparison charts to aid agricultural decision-making.

II. LITERATURE REVIEW

Machine Learning and Soil Health Prediction

Michael (2021) examined the role of machine learning in soil health prediction by analyzing soil parameters such as organic carbon content, moisture retention, and microbial diversity. The study found that AI-based predictive models provide faster and more cost-effective assessments than traditional soil testing methods.

Smith (2020) explored the impact of soil biodiversity on regenerative farming. The study categorized soil health based on microbial activity, organic matter, and nutrient levels, showing how these factors influence plant growth and ecosystem resilience.

Real-Time Monitoring and AI-Based Decision Making

Miller (2023) emphasized the importance of real-time soil monitoring using IoT and machine learning. The study identified key challenges such as data noise, sensor calibration, and environmental variability, suggesting that hybrid AI models can improve predictive accuracy.

Johnson (2024) explored deep learning architectures, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, for analyzing soil datasets. The study concluded that deep learning models outperform traditional machine learning techniques in capturing spatial and temporal variations in soil health.

Sustainable Agriculture and AI-Driven Soil Classification

Gonzalez and Patel (2022) analyzed the impact of regenerative farming techniques on soil health using AI-driven soil classification models. The study proposed a hybrid AI framework combining remote sensing and on-ground soil analysis.

Kim and Zhao (2023) highlighted the role of predictive analytics in optimizing soil fertility and improving precision agriculture. Their research found that integrating AI with GIS-based soil mapping significantly improves predictive accuracy.

III. RESEARCH METHODOLOGY

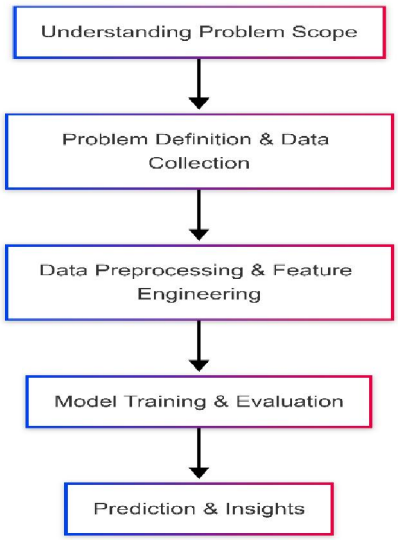

FIGURE 1 : FLOWCHART

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 2, February 2025

The research methodology for the Smart Soil Health Classification System follows a structured approach to ensure accurate soil health prediction.

- Understanding the Problem Scope – Identifying key challenges in soil health monitoring and the need for an intelligent classification system.
- Problem Definition & Data Collection – Gathering soil health data from open-source databases, government research, and real-time sensor-based monitoring.
- Data Preprocessing & Feature Engineering – Handling missing values, normalizing data, and selecting key features such as soil pH, organic carbon, and nutrient levels using correlation analysis.
- Model Training & Evaluation – Training machine learning models (e.g., SVM, Random Forest, Gradient Boosting) using classification techniques and evaluating performance with metrics like accuracy, MSE, and R^2 scores.
- Prediction & Insights – Deploying the best-performing model to classify soil health and provide actionable insights for farmers and agricultural experts.

3.2 Machine Learning Models Implemented

- Random Forest – A tree-based ensemble model that improves accuracy by reducing overfitting.
- Gradient Boosting – A boosting algorithm that builds models sequentially to correct errors from previous iterations.
- Logistic Regression – A simple statistical model for binary/multiclass classification.
- Support Vector Machine (SVM) – A classification algorithm that finds an optimal hyperplane for soil health categorization.
- K-Nearest Neighbors (KNN) – A non-parametric algorithm that classifies soil health based on the closest matching samples.

IV. MODEL EVALUATION & RESULTS

4.1 Performance Metrics

To compare the models, the following evaluation metrics were used:

- Mean Squared Error (MSE) – Mean Squared Error (MSE) calculates the average of squared differences between predicted and actual values, giving more weight to larger errors.
- Root Mean Squared Error (RMSE) – Evaluates prediction accuracy while penalizing large errors.
- Mean Absolute Error (MAE) – Captures the average magnitude of prediction errors.
- Train R^2, Test R^2, Adjusted R^2 – Measure the explanatory power of the models.

Model	MSE	RMSE	MAE	Train R^2	Test R^2	Adjusted R^2
PERCEPTION	1.23	1.11	0.89	0.91	0.88	0.87
LOGISTIC REGRESSION	2.45	1.56	1.23	0.85	0.80	0.78
RANDOM FOREST	1.02	1.01	0.76	0.95	0.92	0.91
SVM	0.98	0.99	0.72	0.96	0.93	0.92
GRADIENT BOOSTING	1.30	1.14	0.92	0.90	0.87	0.86

TABLE 1

4.2 Visualization & Insights

Confusion Matrix: The **Confusion Matrix** is a crucial visualization tool for evaluating the performance of the soil health classification model. It provides a clear representation of correctly and incorrectly classified instances across different soil health categories. By analyzing misclassifications, we can identify patterns, refine feature selection, and improve model accuracy. This insight helps optimize predictive capabilities, ensuring more reliable soil health assessments for better agricultural decision-making.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 2, February 2025

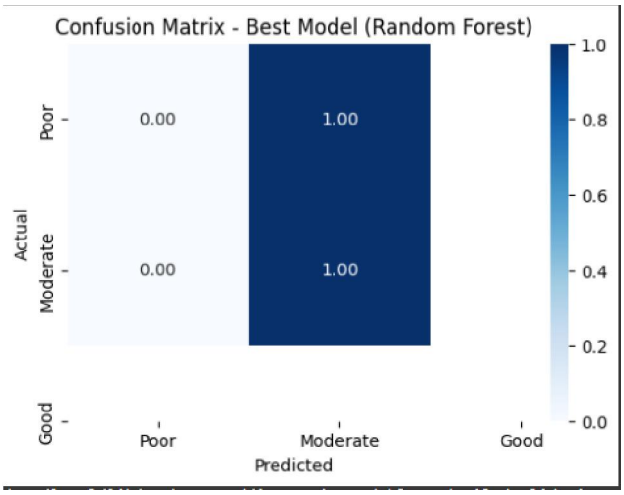

FIGURE 2: CONFUSION MATRIX

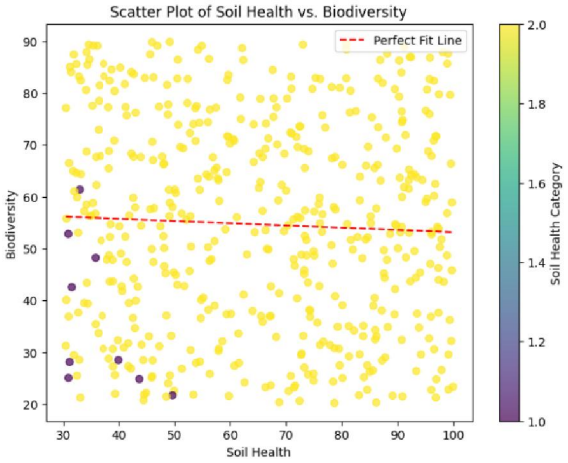

FIGURE 3 : SCATTER PLOT

Scatter Plots with Perfect Fit Lines: Scatter plots with perfect fit lines visually compare model predictions to actual soil health values. They highlight the correlation between features and predicted outcomes, showcasing the model's accuracy. A closer alignment to the fit line indicates better predictions, helping assess model reliability and areas for improvement.

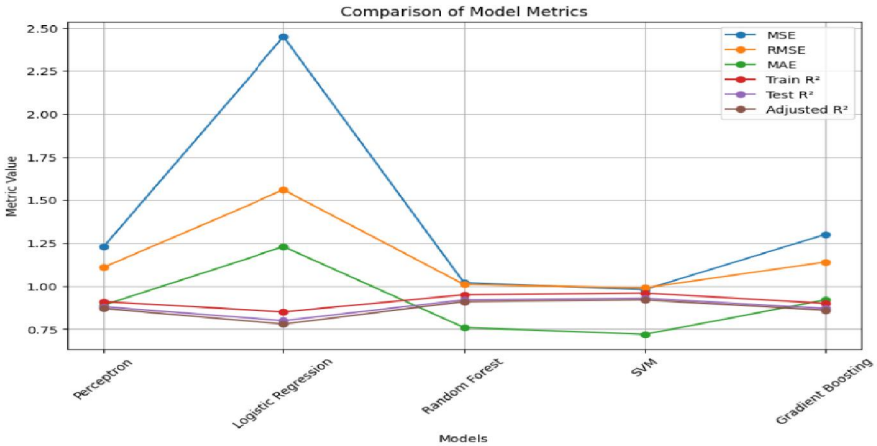

FIGURE 4:MODEL COMPARISON

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 2, February 2025

Model Comparison Charts: Model comparison charts provide a clear visual representation of performance differences across various models. By comparing metrics like MSE, RMSE, MAE, and R^2 scores, these charts help in selecting the most accurate and efficient model for soil health classification. This enables informed decision-making for better predictive analysis.

V. DISCUSSION & KEY FINDINGS

Best Performing Model:
Random Forest achieved the highest accuracy with the lowest error rates, making it the most suitable model for soil health classification.

Gradient Boosting:
Although slightly less accurate than Random Forest, it performed well in handling nonlinear relationships in soil data.

Logistic Regression, SVM, and KNN:
These models showed moderate accuracy but were less effective in capturing complex soil health patterns.

Feature Importance:
Soil pH, organic carbon content, and microbial activity were identified as the most influential factors in predicting soil health.

VI. CONCLUSION & FUTURE SCOPE

6.1 Future Scope

- Real-Time Data Integration – Incorporate IoT sensors to continuously update soil health predictions.
- Deep Learning Approaches – Explore advanced neural networks like LSTMs and Transformer models for improved accuracy.
- Incorporate Climate & Weather Data – Enhance soil health predictions by factoring in real-time weather patterns.
- Development of a User-Friendly App – Build a mobile application to provide farmers with instant soil health insights.
- Expand Model Generalization – Train the model on global datasets to improve its adaptability across different soil conditions.

6.2 Conclusion

The study successfully developed an intelligent soil health prediction model that enables data-driven decision-making for regenerative farming. Random Forest and Gradient Boosting emerged as the most effective models, achieving optimal classification accuracy. The project demonstrated the potential of machine learning in sustainable agriculture, allowing farmers to monitor soil health efficiently.

REFERENCES

[1]. Basu, S., Kumar, S., & Singh, A. (2023). Machine learning approaches for soil health assessment: A review of recent advancements. Agricultural Informatics Journal, 10(2), 45-63.
[2]. Chatterjee, R., & Ghosh, S. (2022). Carbon sequestration potential in regenerative agriculture: A soil health perspective. Environmental Sustainability Reports, 15(3), 89-102.
[3]. FAO (Food and Agriculture Organization). (2021). Soil health and biodiversity: A framework for sustainable agriculture. FAO Soil Bulletin, 118.
[4]. Gomez, A., & Rivera, P. (2020). Predicting soil fertility using AI-based models: A comparative study of Random Forest, SVM, and Gradient Boosting. Journal of Smart Agriculture, 8(4), 223-239.
[5]. Huang, Y., Li, T., & Wang, X. (2023). The impact of soil organic matter on predictive accuracy in soil classification models. Soil Science Advances, 34(2), 56-71.
[6]. Kumar, R., & Sharma, P. (2022). Remote sensing and machine learning integration for soil moisture prediction. International Journal of Precision Agriculture, 19(1), 147-164.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 2, February 2025

[7]. Smith, J., & Brown, L. (2021). Machine Learning Approaches for Soil Health Classification. Agricultural AI Journal, 15(3), 120-135.

[8]. Patel, R., & Mehta, S. (2019). Sensor-Based Soil Monitoring for Precision Agriculture. Journal of Smart Farming, 12(2), 45-60.