

Predictive Modeling of Soil Microbiome Health Using Agentic AI Frameworks

¹Saudamini Samantaray, ²Priyadarsini Sahoo

^{1,2} Assistant Professor, Department of Computer Science Engineering, GIFT Autonomous, Bhubaneswar

Abstract

Soil microbiomes play a crucial role in maintaining agricultural productivity, supporting nutrient cycling, and promoting environmental sustainability. However, conventional methods used to assess soil health often provide limited predictive insights and struggle to adapt to changing environmental conditions. This study presents an agentic Artificial Intelligence (AI) framework that combines machine learning, reinforcement learning, and multi-agent systems to predict soil micro biome health using environmental and genomic data. The experimental results indicate that the proposed approach achieves better predictive accuracy, adaptability, and decision-making performance than traditional models. These findings demonstrate the potential of agentic AI to support precision agriculture and encourage more sustainable soil management practices.

Keywords: Soil Microbiome, Agentic AI, Precision Agriculture, Predictive Modelling.

1. Introduction

Soil health is a fundamental factor influencing both agricultural productivity and ecological sustainability. Among the various indicators of soil quality, the soil microbiome—the diverse community of bacteria, fungi, archaea, and viruses found within soil ecosystems—plays a particularly important role. These microorganisms contribute to nutrient cycling, decomposition of organic matter, carbon sequestration, and protection against plant diseases. Variations in microbial diversity and population structure can therefore have a direct impact on crop yield and ecosystem resilience.

Conventional methods for assessing soil health primarily rely on laboratory procedures such as pH analysis, nutrient measurement, and microbial culturing. Although these approaches are effective for localized studies, they are often time-consuming, costly, and limited in their ability to capture the dynamic nature of microbial ecosystems. In addition, environmental pressures such as climate change, excessive fertilizer use, and industrial pollution create highly complex and nonlinear interactions that traditional statistical techniques often struggle to model accurately.

Recent advances in Artificial Intelligence (AI) and machine learning have introduced new opportunities for agricultural analytics. Supervised learning models have shown promise in applications such as crop disease prediction, soil fertility estimation, and microbial imbalance detection. However, most existing systems remain reactive rather than adaptive. They typically require manual retraining, have limited reasoning capabilities, and cannot efficiently coordinate decisions across multiple heterogeneous datasets.

Agentic AI frameworks offer a more advanced and flexible alternative. Unlike traditional AI models, agentic systems are autonomous, goal-oriented, and capable of adapting to changing environments through continuous learning. They can also collaborate through multi-agent coordination, making them well suited for soil microbiome analysis, where environmental conditions evolve constantly and decisions often need to be made in real time.

Despite ongoing progress in predictive soil analytics, current models for soil microbiome health still face several important limitations:

- Difficulty in accurately modeling dynamic ecological interactions.
- Limited ability to integrate heterogeneous environmental and genomic datasets.

- Weak adaptability to changing climatic and agricultural conditions.
- Absence of autonomous decision-making and adaptive learning mechanisms.

These limitations reduce the effectiveness of predictive systems in precision agriculture and sustainable environmental management.

2. Research Objectives

This research aims to:

1. To develop a predictive framework for soil microbiome health using agentic AI architectures.
2. To compare the performance of conventional machine learning models with agentic AI systems.
3. To assess the adaptability and interpretability of autonomous AI agents in environmental prediction tasks.
4. To identify the practical applications and current limitations of agentic AI in soil health monitoring.

This study contributes to the growing integration of AI and environmental science by introducing an adaptive and intelligent framework for soil health prediction. By combining autonomous agents with predictive analytics, the proposed approach has the potential to support sustainable agriculture, improve climate resilience, and enhance precision farming practices.

3. Methodology and Research Planning

3.1 Soil Microbiome and Agricultural Sustainability

The ecological importance of soil microbiomes has been widely recognized in environmental research. Fierer (2017) demonstrated that microbial diversity strongly influences nutrient availability and crop resilience, while Delgado-Baquerizo et al. (2018) emphasized the relationship between microbial biodiversity and ecosystem multifunctionality.

Microbial communities are highly sensitive to environmental conditions such as temperature, moisture, soil pH, and fertilizer application. Advances in metagenomic sequencing technologies have enabled researchers to study microbial populations at unprecedented depth and scale. However, interpreting these large and complex datasets remains a major computational challenge.

3.2 Machine Learning in Soil Health Prediction

A variety of machine learning techniques have been applied to soil analysis and agricultural prediction tasks.

Table 1 Soil analysis and Agricultural prediction Model

Model	Application	Strengths	Limitations
Random Forest	Soil fertility prediction	High predictive accuracy	Limited ability to model temporal patterns
Support Vector Machine	Microbial classification	Effective for smaller datasets	Limited scalability
Gradient Boosting	Nutrient prediction	Strong feature handling capabilities	Computationally intensive
Neural Networks	Temporal forecasting	Effective for nonlinear	Requires large datasets

Model	Application	Strengths	Limitations
		relationships	

Li et al. (2020) reported that ensemble learning approaches often outperform traditional regression methods in predicting soil nutrient composition. Deep learning techniques, including Convolutional Neural Networks (CNNs) and LSTM networks, have also demonstrated strong performance in analysing temporal environmental data.

Despite these developments, most existing approaches remain static and dependent on centralized training processes, limiting their adaptability in dynamic environments.

3.3 Agentic AI Frameworks

Agentic AI systems differ from conventional AI models because of their ability to reason autonomously and adapt continuously. These systems are designed around intelligent agents capable of:

- Goal-oriented decision-making
- Environmental adaptation
- Self-directed learning
- Multi-agent collaboration
- Reinforcement-driven optimization

Multi-Agent Systems (MAS) have already been applied successfully in areas such as robotics, autonomous transportation, and smart energy management. Reinforcement learning techniques, including Deep Q-Networks (DQN) and Proximal Policy Optimization (PPO), allow agents to learn optimal strategies through interaction with changing environments.

Emerging studies suggest that agentic AI can significantly improve environmental monitoring systems by supporting decentralized decision-making and adaptive responses.

3.4 Existing Research Gaps

Although AI-based agricultural research has progressed considerably, several gaps remain:

1. Limited integration of agentic AI with soil microbiome analytics.
2. Insufficient development of autonomous adaptive systems for environmental prediction.
3. Limited emphasis on explainable AI in ecological applications.
4. Minimal research on real-time microbial forecasting using multi-agent coordination.

This study addresses these gaps by proposing a hybrid agentic AI framework specifically designed for predictive soil microbiome modelling.

3.5 Dataset Description

The study used publicly available soil microbiome and environmental datasets collected from agricultural research repositories and metagenomic databases. The dataset included:

- Soil pH measurements
- Nitrogen, phosphorus, and potassium concentrations
- Temperature and moisture records
- Microbial diversity indices

- Metagenomic sequencing information
- Crop management data

Table 2 Dataset Summary

Variable Category	Number of Features
Environmental Variables	18
Chemical Properties	12
Microbial Genomic Features	145
Temporal Variables	8

The final dataset consisted of approximately 50,000 soil samples collected from multiple climatic regions.

3.2 Data Preprocessing

The preprocessing workflow involved several stages to improve data quality and model performance.

3.2.1 Missing Value Handling

Missing environmental sensor values and genomic records were handled using K-Nearest Neighbour (KNN) imputation techniques.

3.2.2 Normalization

Continuous variables were normalized using Min-Max scaling to improve training efficiency and model convergence.

3.2.3 Feature Selection

Principal Component Analysis (PCA) and mutual information scoring were applied to reduce dimensionality and eliminate redundant features.

3.2.4 Data Encoding

Categorical agricultural management variables were transformed using one-hot encoding methods.

3.3 Proposed Agentic AI Framework

The proposed framework consists of four interconnected intelligent agents

Table 3 Dataset flow

Agent Type	Primary Function
Monitoring Agent	Collects environmental and microbial data
Prediction Agent	Forecasts microbiome health indicators
Optimization Agent	Recommends soil treatment strategies
Coordination Agent	Synchronizes decisions across agents

The framework combines several advanced AI techniques, including:

- Reinforcement learning for adaptive decision-making
- LSTM networks for temporal prediction
- Graph Neural Networks (GNNs) for modeling microbial interactions
- Explainable AI (XAI) methods for interpretability

3.4 Algorithms Used

Traditional Machine Learning Models

- Random Forest (RF)
- Support Vector Machine (SVM)
- Gradient Boosting Machine (GBM)

Deep Learning Models

- LSTM Networks
- Convolutional Neural Networks (CNN)

Agentic AI Approaches

- Multi-Agent Reinforcement Learning (MARL)
- Deep Q-Network (DQN)
- Proximal Policy Optimization (PPO)

3.5 Evaluation Metrics

The study used multiple evaluation metrics to assess predictive performance.

Classification Metrics

- Accuracy
- Precision
- Recall
- F1-score

Regression Metrics

- Root Mean Square Error (RMSE)
- Mean Absolute Error (MAE)
- R^2 Score

Explainability Metrics

- SHAP value consistency
- Feature importance stability

3.6 Experimental Setup

Experiments were conducted using Python-based frameworks such as TensorFlow, PyTorch, and Ray RLlib. Model training was performed on GPU-enabled cloud infrastructure.

Table 4 Hardware Configuration

Component	Specification
CPU	Intel Xeon 32-core
GPU	NVIDIA A100
RAM	128 GB
Storage	4 TB SSD

The dataset was divided into:

- 70% training data
- 15% validation data
- 15% testing data

Cross-validation techniques were also applied to improve robustness and reduce over fitting.

4. Results and Discussion

4.1 Predictive Performance

The proposed agentic AI framework achieved the highest predictive accuracy and F1-score, demonstrating its effectiveness in modelling complex ecological interactions and adapting to changing environmental conditions.

Table 5 Classification Results

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	84.2%	83.7%	82.9%	83.3%
SVM	81.4%	80.8%	79.6%	80.2%
GBM	86.7%	85.9%	85.1%	85.5%
LSTM	89.3%	88.8%	88.1%	88.4%
Agentic AI Framework	93.8%	93.2%	92.9%	93.0%

Table 6 Regression Performance

Model	RMSE	MAE	R ² Score
Random Forest	0.42	0.31	0.79
GBM	0.37	0.28	0.84
LSTM	0.29	0.22	0.89

Model	RMSE	MAE	R ² Score
Agentic AI Framework	0.18	0.14	0.95

The significantly lower RMSE values achieved by the agentic AI framework indicate improved precision in predicting microbial diversity and soil nutrient dynamics.

The results demonstrate that agentic AI systems provide clear advantages over traditional predictive approaches.

Key Findings

Adaptive Learning

The agentic framework continuously updated its prediction strategies in response to environmental changes, enabling more flexible and accurate forecasting.

4.2 Enhanced Temporal Modelling.

The integration of LSTM networks with reinforcement learning improved the system's ability to capture long-term temporal patterns within environmental data.

Autonomous Decision-Making

The optimization agent successfully generated recommendations for corrective actions, including irrigation adjustments and nutrient balancing strategies.

Improved Interpretability

SHAP analysis identified soil moisture, pH levels, and microbial richness as the most influential variables affecting predictive outcomes.

Implications for Precision Agriculture

The proposed framework has several important applications in precision agriculture, including:

- Early identification of soil degradation
- Sustainable fertilizer and nutrient management
- Climate-resilient farming strategies
- Automated agricultural advisory systems

4.3 Limitations

Despite the promising results, several limitations remain:

1. Multi-agent training requires substantial computational resources.
2. Standardized soil microbiome datasets are still limited.
3. Real-time deployment may be difficult in resource-constrained environments.
4. Geographical imbalance in datasets may introduce predictive bias.

5. Conclusion and Future Work

This study presented a comprehensive framework for predictive modelling of soil microbiome health using agentic AI architectures. The findings demonstrate that autonomous multi-agent systems outperform conventional machine learning models in predictive accuracy, adaptability, and temporal forecasting.

By integrating reinforcement learning, deep neural networks, and explainable AI methods, the proposed framework effectively captured complex interactions between microbial communities and environmental conditions. The ability of intelligent agents to autonomously monitor, predict, and optimize soil health represents a significant advancement in precision agriculture and environmental sustainability.

At the same time, challenges related to computational efficiency, data standardization, and large-scale deployment must be addressed before widespread adoption becomes practical.

Future Research Directions

Future research should focus on the following areas:

1. Federated learning for distributed soil monitoring systems.
2. Integration of IoT sensor networks with agentic AI frameworks.
3. Development of lightweight edge AI models for real-time applications.
4. Incorporation of climate simulation models for long-term ecological forecasting.
5. Establishment of ethical and governance frameworks for autonomous agricultural technologies.

Overall, agentic AI presents a promising pathway toward intelligent, adaptive, and sustainable management of soil ecosystems.

References

1. Delgado-Baquerizo, M., Oliverio, A. M., Brewer, T. E., et al. (2018). A global atlas of the dominant bacteria found in soil. *Science*, 359(6373), 320–325.
2. Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology*, 15(10), 579–590.
3. Li, Y., Wu, J., Liu, S., & Shen, Q. (2020). Machine learning approaches for predicting soil properties and crop productivity. *Computers and Electronics in Agriculture*, 170, 105–117.
4. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
5. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
6. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press.
7. Silver, D., Huang, A., Maddison, C. J., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489.
8. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
9. Zhou, Z.-H. (2021). *Machine Learning*. Springer.
10. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260.