AI-Based Soil Microbiome Analytics for Optimized Crop and Fertilization Selection

[1]Tarun Bankar, [2]Prof. Anil Band

[1]Department Of Artificial Intelligence
[1,2]Priyadarshini JL College Of Engineering, Nagpur, India

---***---

Abstract - The In order to enhance precision agriculture, this study presents a web based system that leverages artificial intelligence to provide personalised fertiliser recommendations. It combines microbiome studies from QIIME2 and PICRUSt2 with soil physicochemical data (crop type, yield target, field size, pH, organic carbon, total nitrogen, phosphorus, potassium, and moisture) to produce insights relevant to a given field. A key component of the platform is the Mistral-7B AI model, which generates actionable recommendations by analysing soil microbiome profiles across 22 health metrics.Utilising a Python FastAPI backend for effective data processing and real-time inference and a ReactJS frontend for user interaction, the system eventually supports sustainable farming methods and enhanced soil health.

***Key Words*:** Precision Agriculture, Artificial Intelligence, Soil Microbiome, Fertilizer Recommendation, QIIME2, PICRUSt2, Soil Physicochemical Properties, Mistral-7B, FastAPI, ReactJS, Sustainable Farming, Soil Health.

1. INTRODUCTION

Maintaining healthy soils is becoming more and more important to sustainable agriculture in order to protect the environment and increase agricultural yields. The use of 16S rRNA sequencing, a molecular method that separates bacterial species by looking at both conserved and variable sections of the gene, has advanced modern soil evaluation. This method provides crucial information about the dynamics of nutrients, the decomposition of organic matter, and the general functioning of ecosystems in soil environments. Since these microbial interactions are essential to soil fertility and nutrient cycle, it is imperative to comprehend them.Conventional fertiliser recommendations frequently overlook the important role of soil bacteria and are mostly based on chemical analyses. This omission may result in imbalanced fertilisation techniques and less-than-ideal nutrition management. This study responds by presenting an integrated strategy that blends in-depth microbiome profiling with thorough chemical analysis. Together with powerful bioinformatics tools like QIIME2, high-throughput 16S rRNA sequencing converts raw microbial data into diversity indices and operational taxonomic units that precisely reflect the composition of the soil's bacterial community.

A comprehensive picture of soil health is obtained by combining this microbial data with important soil characteristics, such as crop type, yield goals, field size, pH, organic carbon, total nitrogen, phosphorus, potassium, and moisture levels. To forecast fertiliser needs that are most appropriate for the particulars of each field, sophisticated machine learning algorithms—in particular, Random Forest and Support Vector Machines—are used. This method is further improved by the addition of the Mistral-7B AI model, which makes it possible to synthesise intricate chemical and microbiological data into clear, practical suggestions for 22 different soil health categories.

The suggested system exhibits resilience and usefulness by tackling issues like data heterogeneity and computing needs with scalable cloud-based solutions and optimised preparation pipelines. In addition to improving fertiliser accuracy, the integration of multidisciplinary techniques promotes long-term soil conservation and sustainable agricultural methods. By guaranteeing that nutrient administration techniques are effective and environmentally sound, this thorough methodology opens the door for better agricultural management and eventually supports sustainable food production systems.

2. LITERATURE REVIEW

[1] Matias Fernández-Huarte, Ivan Mijangos, Blanca Etayo, Diego Sánchez-Pinillos, Carlos Garbisu, "Metabarcoding reveals that soil bacterial and fungal diversity is strongly affected by location and only variably by agricultural management," ***Soil Biology and Biochemistry*****, Vol. 184, 2023, Article 109104.**
This study introduced DNA-BSAM, a standardised biodiversity analysis tool, addressing a significant difficulty in soil microbiome research. The findings indicated that microbial diversity was more or less influenced by geographic location than by agricultural management, with site-dependent management effects.

[2] Don A. Cowan, Anusha Mothilal, Bryan Treasure, et al., "Biogeographical survey of soil microbiomes across nine sub-Saharan African countries," ***Microbiome*****, Vol. 10, 2022, Article 131.**
A comprehensive investigation of topsoil microbiomes in nine African nations. The study found that environmental factors like temperature, precipitation, and pH have a big impact on the organisation of microbial communities and are crucial for forecasting the implications of climate change.

[3] A.P. Keinath, J.M. DuBose, R.B. Leidy, "Effect of Velvetbean Cover Crop Mulch on Wirestem of Collard Caused by Rhizoctonia solani," ***Plant Disease*****, Vol. 87, No. 12, 2003, pp. 1490–1494.**
This study highlighted the complicated effects of some

agricultural methods by demonstrating that, whereas velvet bean mulch is frequently used to improve soil quality, it can also unintentionally encourage soil diseases like Rhizoctonia solani.

[4] J.L. Kirk, L.A. Beaudette, M. Hart, P. Moutoglis, J.N. Klironomos, H. Lee, J.T. Trevors, "Methods of studying soil microbial diversity," ***Journal of Microbiological Methods*****, Vol. 58, No. 2, 2004, pp. 169–188.**
This study examined developments in the investigation of microbial diversity, with a focus on the move to molecular techniques like 16S rRNA sequencing. It underlined the necessity of analytical standardisation and the use of DNA-based techniques in identifying species that cannot be cultured.

[5] C.J. Feeney, M.A. Bruns, T. Adams, et al., "Transparency in soil microbiome research: Introducing CRediT authorship," ***Science of the Total Environment*****, Vol. 951, 2024, Article 175642.**
By endorsing the CRediT taxonomy for author contributions, the report highlighted ethical research procedures. Despite not having a direct focus on microbiomes, it encourages transparency and responsibility in scientific partnerships.

[6] T. Morvan, L. Beff, Y. Lambert, et al., "An original experimental design to quantify and model net mineralization of organic nitrogen in the field," ***Nitrogen*****, Vol. 2, 2022, pp. 197–212.**
T. Morvan, Y. Lambert, P. Germain, et al., "A dataset from a 3-year network of field measurements of soil organic nitrogen mineralization under a mild oceanic temperate climate," ***Data in Brief*****, Vol. 35, 2021, Article 106841.**
These research examined the effects of soil types and parent materials on microbial habitats and organic nitrogen mineralisation. The results highlight how microbial community architectures are shaped by the physical and chemical characteristics of the soil.

3. PROPOSED SYSTEM

1. Access to the dashboard and input upload
Users can add soil physicochemical data (pH, nitrogen, phosphorus, potassium, organic carbon, moisture, etc.) after logging into the application.
• Provide filtered microbiome data files, particularly the PICRUSt2 filtered_abundances.tsv output.
• See the most recent evaluations of soil health or earlier submissions.

2. Integration of Microbiome Analysis Pipelines
The system parses the microbial abundance data (from the QIIME2 and PICRUSt2 pipelines) after it has been uploaded. Maps predicted functions using pathways from KEGG Orthology.
• Combines important functional parameters, microbial diversity, and richness into 20 categories of soil health (e.g., nitrogen metabolism, stress tolerance, organic matter breakdown).
This information is connected to the user's session and kept in an organised fashion.

3.The Fertiliser Suggestion Engine
Making use of the KEGG functional outputs and the processed soil chemical inputs:
• To forecast the ideal fertiliser doses (such as urea, DAP, and potash), the backend activates a Random Forest Regressor model that has already been trained.
• The model recommends biofertilizer strains (e.g., Bacillus for phosphate solubilisation, Rhizobium for nitrogen fixation) based on biological indicators and nutrient imbalances.
Based on their individual soil profile, users are provided with tabular results for fertiliser recommendations.

4.Clustering & Crop Suitability Simultaneously, the system recommends crops using classification and clustering models: Clustering algorithms use chemical and microbiological characteristics to group related soil profiles. • For each cluster, a crop mapping dictionary recommends crops that are suitable for the area (legumes for nitrogen-deficient soils, for example).

5. Classification for Soil Health
The model assesses the soil in 20 standardised categories, such as enzyme activity, microbial diversity, and nutrient availability. It creates a Soil Health Card that shows:
• A combined score between 0 and 100.
• Visual cues (traffic light: green indicates health, yellow indicates intermediate, and red indicates poor).
• The degree of risk and comments on important indications, such as "Low P availability, consider PSB biofertilizer."

6. Natural language summarisation powered by AI
The 20-category soil health outputs are transformed into easily comprehensible language using the Mistral-7B GenAI model.
• Produce practical suggestions like: "There is not enough organic carbon in your soil. Nitrogen retention can be enhanced by adding compost and Azotobacter biofertilizer. Particularly for non-technical farmers, the system makes sure these outputs are user-friendly and contextual.

7. Frontend: Interface for ReactJS
The frontend offers: • Interactive soil input dashboards.
• Data visualisations, such as KEGG bar charts, diversity plots, and PCoA.
• The ability to download soil health cards in PDF format.
• Integrated microbial index and nutrient balance charts.
Because of its complete responsiveness, it is accessible on both desktop and mobile platforms.

8. Backend: Microservices using FastAPI
FastAPI manages: • Model execution (Crop Classifier, Fertiliser Regressor).
• API endpoints for GenAI output generation, prediction, and file parsing.
• Safe database operations and routing (MongoDB).

9. Environment for Data Processing
Conda environments are used in Google Colab to run all QIIME2 and PICRUSt2 procedures. This makes it possible to process huge FASTQ and TSV files quickly.
- If necessary, GPU acceleration.
- Notebook sharing and replication are simple.

4. MODULES DESCRIPTION

4.1 Module for Data Acquisition and Preprocessing

This module gathers soil physicochemical data (pH, N, P, K, OC, and moisture) and 16S rRNA microbial sequencing data from NCBI. Trimmomatic and QIIME2 (with DADA2) are used for preprocessing in order to eliminate noise and produce high-quality ASVs. The SILVA database is used to assign taxonomy, and then PICRUSt2 integration is used to estimate microbial functional pathways using KEGG, resulting in the crucial filtered_abundances.tsv.

4.2 Predictive Modelling & Machine Learning Module

This module builds models that forecast fertiliser needs and suggest crops using scikit-learn. Crop selection is supported by classification and clustering techniques, while a Random Forest Regressor links soil and microbial characteristics to fertiliser requirements. Based on nutrient levels and microbial activity, soil health is assessed under 20 functional categories.

4.3 An Interpretation Module powered by AI

The Mistral-7B The ML results are processed by the generative AI model, which then transforms them into suggestions in natural language. It analyses chemical profiles, microbial activity, and soil health indicators to give farmers practical advice on fertiliser and biofertilizer usage.

4.4 Backend Module & Web Interface

This module facilitates backend processing and user interaction through the usage of ReactJS and FastAPI. The frontend provides textual and graphic reports and accepts soil data input. Input management, model execution, and API responses are all handled by the FastAPI backend, which guarantees smooth operation and instantaneous feedback.

Fig-1: Flowchart

5. TECHNOLOGY USED

1. Frontend (Interface for ReactJS)
ReactJS, a contemporary JavaScript package well-known for its effectiveness in creating dynamic user interfaces, is used to build the application's frontend. Users can upload microbiome datasets, enter soil test parameters, and evaluate predictions produced by the backend models on this highly interactive site. Reusable components guarantee clear and maintainable code, and the platform's device responsiveness makes it useful by a wider range of users, including farmers using mobile devices. Additionally, the interface offers visualisations like KEGG pathway summaries and diversity bar graphs, which simplify the interpretation of complex microbiological data for consumers. These design components convert research results into useful agricultural decision-support instruments.

2. Backend (Mistral-7B AI + FastAPI + Machine Learning)
Python powers the backend, which is organised using FastAPI, a high-performance framework built for creating API-driven apps quickly. It manages the ingestion of data, calls upon machine learning models that have been trained, and instantly returns the results to the frontend. In order to forecast fertiliser dosages, Random Forest regressors are trained on integrated soil chemical and microbial function data using scikit-learn to create core models. On the basis of enriched soil profiles, classification models recommend crop compatibility. Furthermore, automated creation of human-readable soil health advice is made possible by a GenAI integration (Mistral-7B). This extensive language model improves accessibility for stakeholders who are not technical by translating scientific findings—like KEGG pathway impacts or microbial diversity

scores—into easily understood text summaries.

3. Microbiome Analysis (KEGG Mapping, PICRUSt2, and QIIME2)
The application's scientific foundation is microbial analysis, which is carried out with QIIME2, a complete bioinformatics toolset for microbiome profiling. High-resolution Amplicon Sequence Variants (ASVs) are produced by rigorous preparation of sequencing data, which includes quality filtering, chimaera removal, and denoising using the DADA2 plugin. The SILVA database is used to taxonomically classify these ASVs. PICRUSt2 then uses the KEGG (Kyoto Encyclopaedia of Genes and Genomes) database to correlate ASVs to known gene functions in order to predict the functional capability of microbial communities. The filtered_abundances.tsv file, which includes normalised values for gene families implicated in nutrient cycling and soil health, is produced as a result of this operation. These results are essential for machine learning analysis that comes later.

4. The Processing Platform, Google Colab

All analyses are conducted in a Google Colab environment to manage the computational burden of bioinformatics procedures and promote reproducibility. For executing QIIME2 and PICRUSt2 pipelines, Google Colab offers a scalable and adaptable platform that supports Python and Conda-based systems. Large sequencing datasets can be processed effectively without the requirement for local high-performance computers because to its free GPU and high-memory setups. Cloud-based notebooks also improve teamwork by enabling several researchers to replicate and expand analysis workflows with less setup. The project stays scalable and accessible by integrating the pipeline into Colab, which makes it a perfect fit for academic and field research environments.

6. ADVANTAGES AND APPLICATIONS

1. Precision Recommendations: By integrating microbial functional profiles with soil physico-chemical data, this tool offers tailored crop and fertilizer recommendations for specific fields. This makes it possible to implement customized farming methods that raise production and cut down on input waste.

2. AI-Powered Interpretation: GenAI (Mistral-7B) generates comprehensible, text-based soil health assessments, simplifying complex data for farmers. It connects sophisticated data analytics to practical farming choices.

3. Dual Recommendation Engine: By providing guidance on chemical and biofertilizers on a same platform, this tool promotes sustainable input usage. This preserves output while promoting ecologically friendly farming.

4. Functional Microbial Insights: identifies the roles of microorganisms in nitrogen fixation, phosphorus solubilization, and other activities using KEGG mapping and PICRUSt2. These revelations aid in locating organic mechanisms that can lessen reliance on artificial inputs.

5. Modular Architecture: The seamless integration of future ML models into the FastAPI backend makes the system scalable and adaptable. The platform is thus future-proofed for developing research and regional integration.

6. Educational Value: Offers students, agritech companies, and institutes researching microbiome-based agriculture a resource. It raises awareness of AI-driven agriculture and encourages practical learning.

7. Regional Adaptability: Able can be modified using local datasets to account for regionally specific farming practices and soil properties. This guarantees efficacy and relevance in a variety of agricultural environments.

7. CONCLUSION

In order to promote sustainable and precision agriculture, this project introduces a novel AI-powered web application that combines physicochemical soil profiles with data from soil microbiomes. The system provides customised recommendations for crop selection, fertilisers, and biofertilizers by fusing traditional soil test results with microbial diversity analysis (using QIIME2 and PICRUSt2). The platform fills important gaps in standardised microbial assessment and location-specific variability by using machine learning and big language models to understand soil health across functional categories. It provides a user-friendly interface for data-driven, real-time agricultural decision-making with a ReactJS frontend and FastAPI backend. In addition to supporting microbiological balance and encouraging resource efficiency, this approach lays the foundation for future integration with IoT sensors and climate-smart technology.

REFERENCES

[1] Bhogal, A., Nicholson, F.A., Young, I., Sturrock, C., Whitmore, A.P., Chambers, B.J. Effects of Recent and Accumulated Livestock Manure Carbon Additions on Soil Fertility and Quality. *European Journal of Soil Science*, Vol. 62, pp. 174–181, 2011.

[2] Maskova, L., Simmons, R.W., Deeks, L.K., De Baets, S., Drost, D.T. Impacts of Long-Term Application of Best Management Practices on Yields and Root Carbohydrate Content in Asparagus (*Asparagus officinalis*) (UK). *Available at SSRN*, Paper No. 4349478, 2022.

[3] Morrison-Whittle, P., Goddard, M.R. From Vineyard to Winery: A Source Map of Microbial Diversity Driving Wine Fermentation. *Environmental Microbiology*, Vol. 20, pp. 75–84, 2018.

[4] Nicholson, F.A., Bhogal, A., Chadwick, D., Gill, E., Gooday, R., Lord, E., Misselbrook, T., Rollett, A., Sagoo, E.,

Smith, K. An Enhanced Software Tool to Support Better Use of Manure Nutrients: MANNER-NPK. *Soil Use & Management*, Vol. 29, pp. 473–484, 2013.

[5] Odelade, K.A., Babalola, O.O. Bacteria, Fungi and Archaea Domains in Rhizospheric Soil and Their Effects in Enhancing Agricultural Productivity. *International Journal of Environmental Research and Public Health*, Vol. 16, Article 3873, 2019.

[6] Zhou, T., et al. Age-Dependent Forest Carbon Sink: Estimation via Inverse Modeling. *Journal of Geophysical Research: Biogeosciences*, Vol. 120, pp. 2473–2492, 2015.

[7] Mattila, T.J. Redox Potential as a Soil Health Indicator – How Does It Compare to Microbial Activity and Soil Structure? *Plant and Soil*, 2023.

[8] Moebius-Clune, B.N., Moebius-Clune, D.J., Gugino, B.K., Idowu, O.J., Schindelbeck, R.R., Ristow, A.J., van Es, H.M., et al. Comprehensive Assessment of Soil Health - The Cornell Framework (3.2). *Cornell University*, 2017.

[9] Hosseini, M., Agereh, S.R., Khaledian, Y., Zoghalchalid, H.J., Brevik, E.C., Naeini, S.A.R.M. Comparison of Multiple Statistical Techniques to Predict Soil Phosphorus. *Applied Soil Ecology*, Vol. 114, pp. 123–131, 2017.

[10] Rachwał, K., Gustaw, K., Kazimierczak, W., Waśko, A. Is Soil Management System Really Important? Comparison of Microbial Community Diversity and Structure in Soils Managed Under Organic and Conventional Regimes. *PLOS ONE*, Vol. 16, e0256969, 2021.

[11] Veenadhari, S., Misra, B., Singh, C. Machine Learning Approach for Forecasting Crop Yield Based on Climatic Parameters. *2014 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, IEEE, pp. 1–5, 2014.

[12] Dasgupta, D., Brahmaprakash, G. Soil Microbes Are Shaped by Soil Physico-Chemical Properties: A Brief Review of Existing Literature. *International Journal of Plant & Soil Science*, Vol. 33, pp. 59–71, 2021.