International Journal of Advance and Applied Research
www.ijaar.co.in
ISSN – 2347-7075 Impact Factor – 8.141
Peer Reviewed Bi-Monthly
Vol. 7 No. 28 March - April - 2026

Explainable Hybrid Machine Learning Models for Soil Fertility Prediction with Generative AI-Based Crop Recommendation

Kewal Deshpande[1], Suprabha Kulkarni[2] & Mrinmayee Deshpande[3]

[1]*Department of Mechanical Engineering, All India Shri Shivaji Memorial Society's College of Engineering, Pune, India*

[2]*Department of Mathematics, Pemraj Sarda College, Ahmednagar, India*

[3]*Department of CSE - Artificial Intelligence, Vishwakarma Institute of Information Technology Pune, India*

Corresponding Author – Kewal Deshpande

DOI - 10.5281/zenodo.19396490

Abstract:

Soil fertility evaluation is one of the key aspects for improving agricultural productivity and developing sustainable agriculture practices. Soil fertility evaluation is essential for farmers to take appropriate decisions regarding crop cultivation. In this research work, we propose an intelligent decision-support framework for farmers based on machine learning, explainable AI, and generative AI for soil fertility evaluation and crop cultivation. In the proposed framework, key soil parameters such as Nitrogen (N), Phosphorus (P), Potassium (K), pH level, moisture, temperature, organic carbon, and rainfall are considered as input features for predictive modeling. Three different machine learning algorithms, namely Random Forest, XGBoost, and Support Vector Machine, are implemented for soil fertility evaluation. Experimental evaluation of these algorithms is carried out for soil fertility evaluation, and it is observed that XGBoost algorithm produces 97% accuracy for soil fertility evaluation, whereas Random Forest and Support Vector Machine produce 92% and 72% accuracy, respectively. To improve the explainability of the proposed framework, explainable AI is implemented based on SHAP (SHapley Additive exPlanations), where the contribution of each soil parameter is analyzed for soil fertility evaluation. This would facilitate a better understanding of the impact of different soil characteristics on fertility classification. Additionally, a generative AI-based advisory module is proposed for incorporating context-based crop and farming recommendations based on the soil fertility and other conditions. This would enable the generation of customized recommendations for farmers regarding crops, fertilizers, and other soil management practices. Thus, with the integration of predictive ML-based approaches, explainable AI-based approaches, and generative AI-based approaches, the proposed framework would facilitate an intelligent decision-support framework for farmers with the aim of improving crop productivity and promoting sustainable farming practices.

Keywords: Soil Fertility Prediction, Machine Learning, Explainable AI, SHAP, Generative AI, Crop Recommendation, Smart Agriculture.

Introduction:

Agriculture is considered to be one of the key sectors that can contribute to the sustenance of global food security and economic development in developing countries, as a major portion of the population depends on this sector for their livelihood [1][2]. Soil fertility is considered to be a key factor in the determination of agricultural productivity, as the availability of necessary nutrients in the soil will have a direct impact on the development of crops. In this regard, the conventional methods of evaluating soil fertility have been based on manual laboratory analysis. Such processes have been considered to be time-consuming and not cost-effective for small-scale farmers. In this context, there is a need for intelligent systems that can assist in the analysis of soil.

The recent advancements in artificial intelligence and machine learning have shown promising possibilities in terms of data-driven decision support in agriculture. Machine learning models can process large amounts of soil and environmental data to predict the fertility level of the soil with high accuracy. This can help farmers make decisions regarding crop selection, fertilizer use, and other soil management activities. Despite the increased interest in machine learning in agriculture, most of the current machine learning models in this area are mainly focused on prediction accuracy and do not emphasize the importance of model interpretability and recommendation aspects. In real-life agricultural applications, it is important to understand the impact of different soil parameters on the machine learning model and provide useful recommendations that can be implemented in real-life scenarios.

In this paper, we propose an intelligent decision support framework for agriculture based on machine learning, explainable artificial intelligence, and generative AI for soil fertility prediction and crop recommendation. In our proposed framework, soil parameters such as Nitrogen (N), Phosphorus (P), Potassium (K), pH level, moisture content, organic carbon, rainfall, and temperature are analyzed and evaluated for determining soil fertility levels. In our proposed framework, multiple machine learning algorithms, namely Random Forest, XGBoost, and Support Vector Machine, are implemented and evaluated for determining the best approach for soil fertility prediction. Experimental evaluation of multiple machine learning algorithms for soil fertility prediction using soil parameters indicates that XGBoost is the best approach for soil fertility prediction with 97% accuracy, followed by Random Forest with 92% accuracy, and Support Vector Machine with 72%. To increase the transparency of the soil fertility prediction process, explainable artificial intelligence is proposed and implemented using SHAP (SHapley Additive exPlanations), which provides detailed information regarding soil parameters for improving the interpretability of machine learning algorithms. In addition, a generative AI-based crop recommendation and decision support is proposed and implemented for generating crop recommendations based on soil fertility and other environmental conditions. This advisory module will be helpful for farmers as it provides suggestions like the choice of crops to be grown, fertilizer to be used, and ways to improve the soil. The major contributions of this research work can be identified as the creation of a system for the prediction of soil fertility with the help of machine learning technology, the inclusion of explainable AI technology to make the model interpretable, and the addition of generative AI technology to make intelligent recommendations to the farmers.

Literature Survey:

Recent research has focused on the application of artificial intelligence techniques in analyzing the fertility degradation of the soil and its health status. In this context, paper [3] has shown that machine learning and deep learning techniques can be effectively used to assess the quality of the soil and predict the reduction in fertility. However, the authors have mentioned that there are challenges in terms of data availability and practical implementation in real-life scenarios. In another research work, paper [4], the authors proposed an explainable artificial intelligence model to predict the fertility of the soil using a random forest model with high accuracy and visual explanations to understand the effect of different parameters on fertility levels.

The research work presented in [5] suggests an architecture for an AI-based decision support system for farmers, where data collection, intelligent analysis, and decision-making modules are integrated for improving agricultural productivity. This framework illustrates the role of AI-based analytics for improving crop planning, soil management, and other agricultural decision-making activities. In paper [6], the authors proposed a machine learning-based approach for developing a real-time soil fertility analysis and crop prediction system by incorporating soil parameter collection using sensors. This research work illustrates the effectiveness of ensemble learning and deep learning algorithms for predicting soil fertility and improving decision-making for farmers.

In paper [7], a review of the use of artificial intelligence and machine learning algorithms in the analysis of soil and sustainable agriculture was done. In this paper, the use of Random Forest, SVM, and neural networks in the analysis of soil and its properties, as well as the analysis of soil water content using machine learning algorithms, is reviewed. In paper [8], a review of the use of IoT and AI-based algorithms in estimating the nutrient and fertility level of the soil in agriculture was done. In this paper, the use of intelligent algorithms in analyzing the soil and its properties is reviewed.

Paper [9] presented the integration of artificial intelligence and robotics in the analysis of the soil and fertility management in modern agricultural practices. This paper, in essence, demonstrates the potential of autonomous technologies and artificial intelligence in the real-time monitoring of soil parameters. Paper [10] focused on the evaluation of soil fertility using deep learning techniques based on the analysis of soil organic carbon and macronutrient levels, including NPK. The authors used convolutional neural networks, including LeNet, AlexNet, and VGG16, for the classification of the soil, where VGG16 showed the highest prediction accuracy.

Methodology:

This section explains the overall methodology followed in developing the proposed intelligent agricultural decision-support system. It explains the intelligent agricultural decision-support system based on machine learning, explainable AI, and generative AI. It explains the overall methodology followed in developing the proposed intelligent agricultural decision-support system. It explains the overall methodology followed in developing the proposed intelligent agricultural decision-support system.

System Architecture:

The proposed system architecture comprises different connected components that handle soil information and provide smart recommendations for agriculture. The system starts with the acquisition of soil parameters and proceeds with processing them using machine learning algorithms for predicting soil fertility levels. The results are then examined using explainable AI techniques to understand feature contributions. Finally, a generative AI component provides customized crop recommendations and farming suggestions based on predicted fertility levels and environmental factors.

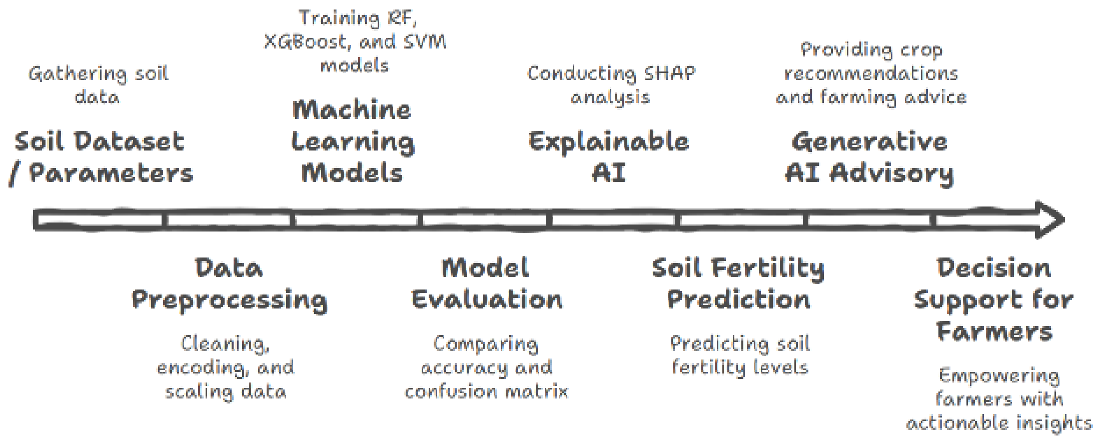

Fig 1. System Architecture of the Proposed Model

The major components of the architecture are as follows:

- Soil dataset and parameter input
- Data preprocessing and feature preparation
- Machine learning model training and evaluation
- Explanable AI using SHAP
- Generative AI-based crop recommendation and advisory

This architecture is designed in a way that it not only gives accurate predictions regarding soil fertility but also offers explanations and recommendations for farmers.

Data Collection and Dataset Description:

The system uses a dataset related to soil, which contains various agricultural and environmental parameters affecting soil fertility. It includes essential nutrients and parameters present in the soil, which have a significant impact on crop productivity. The parameters used in the research are agricultural input features, including Nitrogen (N), Phosphorus (P), Potassium (K), pH levels, soil moisture, organic carbon content, temperature, rainfall, etc.

All these parameters describe the chemical composition and environmental conditions present in the soil. Nitrogen, phosphorus, and potassium are essential macronutrients required for plant growth, while pH levels are essential for nutrient absorption. Environmental parameters like temperature and rainfall are essential for crop selection and soil productivity. Each record in the dataset represents the soil conditions for an agricultural scenario, along with the soil fertility.

Data Preprocessing:

Prior to the training of the machine learning models, there are several preprocessing activities that are undertaken on the data. This ensures that the data is in the best state before the training of the models. First, the data is loaded into the environment, after which the data is checked for inconsistencies and missing values. Once the data is loaded into the environment, the data is then converted into numerical values using label encoding. This ensures that the data can be analyzed using machine learning models. After this, the data is split into training data and test data, with an 80-20 split. This ensures proper evaluation of the data. Additionally, there are feature scaling activities undertaken on the data. This is especially true for algorithms such as Support Vector Machines, where there are concerns regarding the magnitude of the features.

Machine Learning Models for Soil Fertility Prediction:

To correctly predict the soil fertility levels, three different machine learning models are implemented and evaluated using the processed soil data. The models are implemented and evaluated based on their classification accuracy. The first model used in this study is the Random Forest Classifier. The Random Forest Classifier is an ensemble learning method that creates a forest of decision trees and combines their outputs to produce the final prediction result. The ensemble learning method reduces the chances of overfitting, thus increasing the accuracy of the model prediction. The second model implemented in this study is the XGBoost algorithm, which is an advanced version of the gradient boost algorithm and is considered the most efficient algorithm with the highest prediction accuracy. The XGBoost algorithm creates decision trees that learn from the mistakes of the previous trees, thus increasing the accuracy of the model prediction. The third model implemented in this study is the SVM algorithm, which is considered a powerful classification algorithm that can find the optimal hyperplane that separates the different classes in the feature space. The SVM algorithm is considered the most effective algorithm, especially in high-dimensional spaces, and is widely used in classification problems.

Model Evaluation:

To evaluate the performance of the implemented machine learning models, different performance metrics are applied. These performance metrics are accuracy, precision, recall, and F1 score. The dataset is divided into training and testing sets and compared with the predicted values of each model. The results of the experiments show that the performance of the XGBoost model is better than the other models in terms of prediction accuracy, with a maximum accuracy of 97%, followed by Random Forest with a maximum accuracy of 92%, and Support Vector Machine with a maximum accuracy of 72%. This shows that gradient boosting is best suited for complex relationships in soil parameter datasets.

Explainable AI Using SHAP:

Although machine learning models can give very accurate predictions, they are usually black-box models, i.e., the reasoning behind the prediction is not easily interpretable. Therefore, to overcome this disadvantage of machine learning models, the proposed system uses explainable artificial intelligence using SHAP (SHapley Additive exPlanations). SHAP analyzes the contribution of each feature to the prediction made by the model. It assigns importance values to all the input parameters and can be used to visualize how different characteristics of the soil affect the classification of soil fertility using interaction plots and SHAP summary plots to determine the most important parameters of the soil that affect the prediction of fertility.

Generative AI–Based Crop Recommendation System:

In addition to predicting the fertility of the soil, the proposed framework also includes the integration of a generative AI-based advisory module for intelligent recommendations for the agricultural field. Once the fertility of the soil is predicted by the machine learning model, the parameters of the soil and the prediction results will be fed into the generative AI model for intelligent recommendations and suggestions for the agricultural field.

The generative AI system will analyze the parameters of the soil and recommend the type of crops that can be planted based on the conditions of the soil. The system will also provide additional recommendations for the agricultural field, such as the usage of fertilizers and the conditions for irrigation.

Overall Workflow:

The overall workflow of the proposed system can be described as follows: first, the inputting of parameters, then the data is processed and passed to multiple machine learning models for prediction purposes, followed by the evaluation and interpretation of the prediction results using Explainable AI, and lastly, the generative AI model is used to generate crop recommendations and advisory information. Overall, the proposed framework is a comprehensive solution for intelligent decision-making in the field of agriculture.

Results:

This section deals with the experimental evaluation of the machine learning models developed for the prediction of soil fertility. Three different supervised learning models, namely Random Forest, XGBoost, and Support Vector Machine, were trained on the prepared dataset for the prediction of soil fertility. The models were evaluated based on classification metrics such as precision, recall, F1 score, and overall accuracy. Confusion matrices, feature importance, and explainability were used for the analysis of the models.

Model Performance Comparison:

The performance of the three models was evaluated on the testing dataset with 2000 samples. Table 1 presents a summary of the classification accuracy of each model.

Table 1: Model Accuracy Comparison

Model	Accuracy
XGBoost	97%
Random Forest	92%
SVM	72%

The results show that the XGBoost model performed significantly better than the other algorithms, as it recorded the highest classification accuracy of 97%. The Random Forest model also recorded impressive results with a classification accuracy of 92%, while the Support Vector Machine model recorded relatively lower performance compared to the other models with a classification accuracy of 72%.

The reason for the improved performance of the XGBoost model is its gradient boosting technique, which improves its performance by correcting its own prediction errors during each iteration of boosting. This enables it to learn complex relationships between soil parameters and fertility levels.

Confusion Matrix Analysis:

To further evaluate classification performance, confusion matrices were generated for each model.

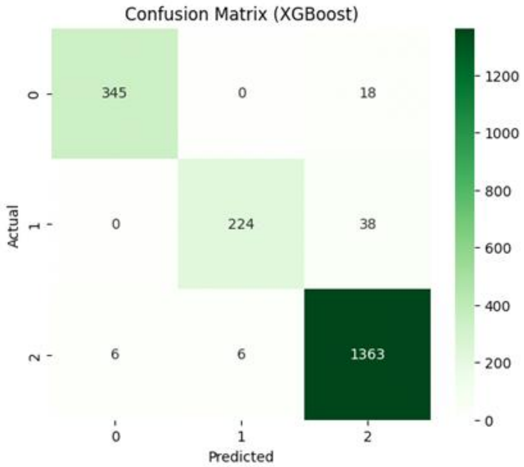

Fig 2. Confusion Matrix for XGBoost Model

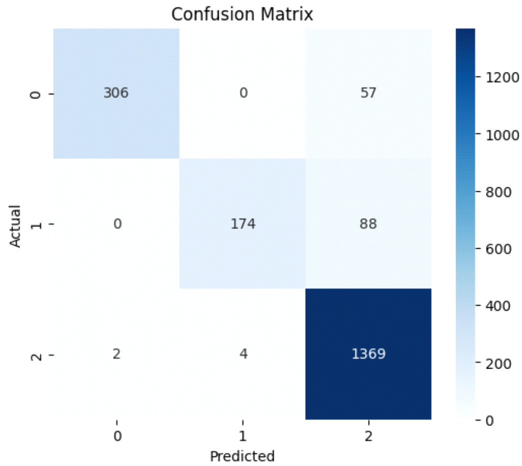

Fig 3. Confusion Matrix for Random Forest Model

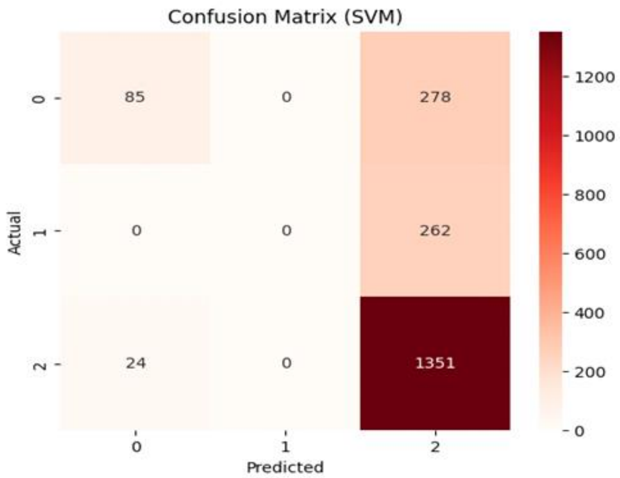

Fig 4. Confusion Matrix for SVM Model

The confusion matrix of the XGBoost model in figure 2 indicates that most samples are correctly classified in all fertility classes. For instance, out of 1375 samples in class 2 (high fertility), 1363 samples are correctly classified by the model, showing that the model is very reliable in classifying dominant fertility classes.

The confusion matrix of the Random Forest model in figure 3 indicates that the model performs very well in classifying the samples, especially in class 2, where 1369 samples are correctly classified by the model. However, the model performs poorly in class 1, showing that there is some confusion in classifying different fertility levels.

The confusion matrix of the Support Vector Machine model in figure 4 shows that there is a lot of confusion in classifying different fertility classes, especially class 1, where the model does not correctly classify most of the samples in this class, explaining why this model has a low level of accuracy compared to other models.

Feature Importance Analysis:

Feature importance analysis was carried out on the model to identify the parameters that have the most significant effect on the prediction of fertility. From the results obtained through the Random Forest and XGBoost models, several parameters are seen to have an effect on the prediction.

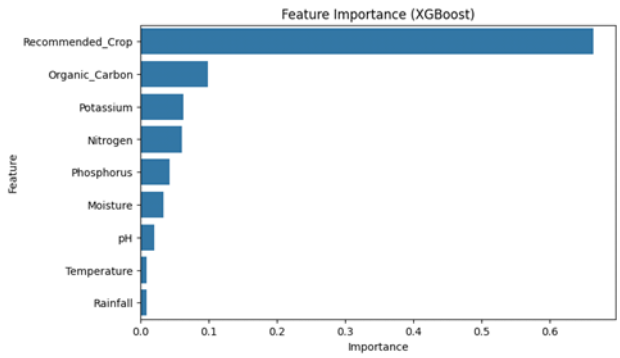

Fig 5. Feature Importance for XGBoost Model

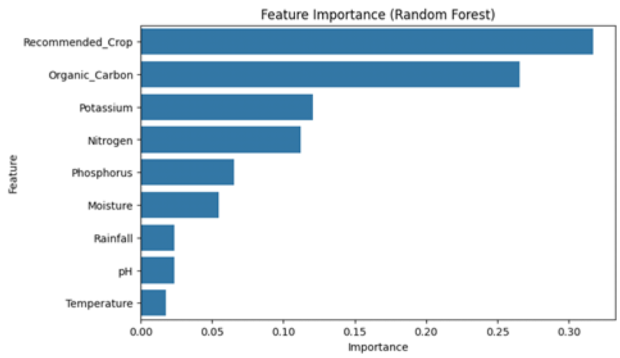

Fig 6. Feature Importance for Random Forest Model

As seen from figures 5 and 6, which are the feature importance plots, Organic Carbon, Potassium, and Nitrogen are seen to have the most significant effect on the prediction of soil fertility. These are essential nutrients in the growth of plants, hence affecting the productivity of the soil. Environmental factors such as rainfall, temperature, and pH are seen to have a less significant effect on the prediction compared to nutrient parameters.

Interestingly, the Recommended Crop parameter is seen as an influential parameter in the prediction. This shows that there is a high correlation between crop recommendations and soil fertility.

Explainability Analysis Using SHAP:

To increase the model's interpretability, Explainable AI was used, and the SHAP (SHapley Additive exPlanations) technique was employed. SHAP is used to increase understanding regarding the features' contribution to the model's predictions.

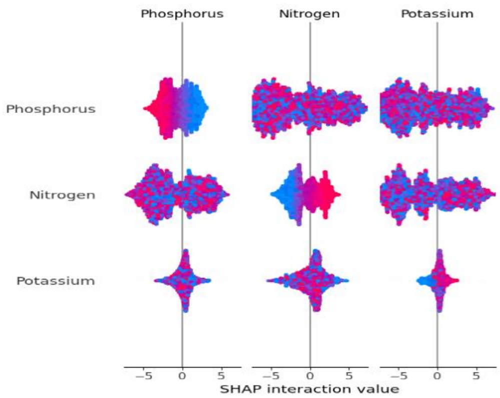

Fig 7. SHAP Summary Plot for XGBoost Model

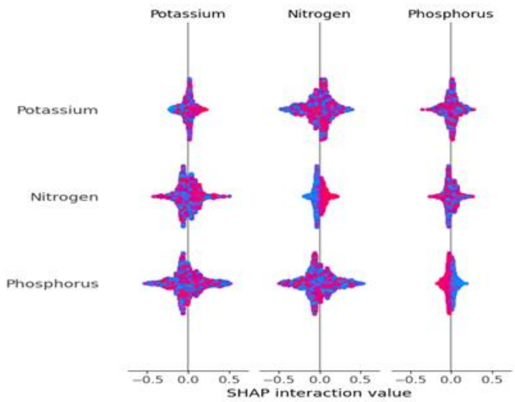

Fig 8. SHAP Summary Plot for Random Forest Model

The SHAP summary plot in figures 7 and 8 indicates that the features Nitrogen, Potassium, and Phosphorus have the most significant influence on the prediction of soil fertility. If the values of these nutrients are high, the prediction is more likely to fall in the higher fertility class, and vice versa.

Furthermore, the SHAP interaction plot in figure 9 is used to investigate the relationships between different parameters of the soil.

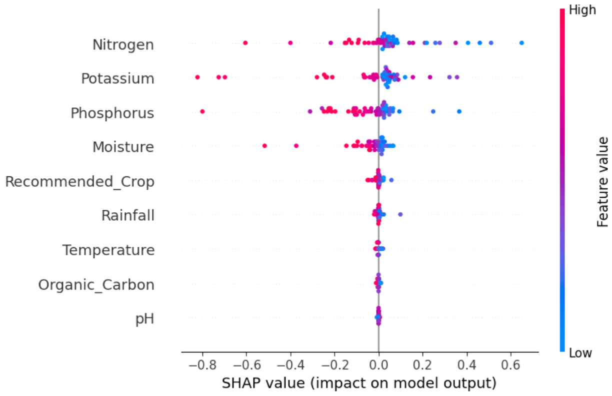

Fig 9. SHAP Summary Plot for SVM Model

As can be seen from the interaction plots, there are moderate interactions present, especially between the soil nutrients, such as Nitrogen and Phosphorus. These interactions show the effect of the combination of soil nutrients on fertility classification, further emphasizing the importance of nutrient balance.

Discussion And Conclusion:

This paper proposed an intelligent decision-support framework for soil fertility prediction and crop recommendation based on machine learning, explainable AI, and generative AI. In this proposed framework, soil fertility is classified based on key soil parameter analysis, i.e., Nitrogen, Phosphorus, Potassium, pH, moisture, organic carbon, rainfall, and temperature. For soil fertility prediction, three different machine learning algorithms were implemented and evaluated, namely Random Forest, XGBoost, and SVM. Among these algorithms, XGBoost showed maximum accuracy of 97%, followed by Random Forest with 92%, and SVM with 72%. To increase model interpretability, explainable AI is employed using SHAP for analyzing the contribution of each soil parameter towards soil fertility prediction. Further, a generative AI-based advisory module is incorporated to generate crop recommendations and farming suggestions based on soil conditions and fertility levels. The results show that it is possible to improve agricultural decision support systems by integrating predictive machine learning models with explainable AI and generative AI techniques. Future directions of this study are discussed below: First, we plan to improve the accuracy of the prediction by integrating real-time soil sensors and satellite-based environmental information systems. Further, we plan to develop this framework as a mobile application or a web application that provides farmers with real-time agricultural recommendations.

References:

1. A. D. Jones and G. Ejeta, "A new global agenda for nutrition and health: the importance of agriculture and food systems," Bulletin of the World Health Organization, vol. 94, no. 3, pp. 228–229, Mar. 2016, doi: 10.2471/BLT.15.164509.
2. W. D. Hopper, "The development of agriculture in developing countries," Scientific American, vol. 235, no. 3, pp. 196–205, Sep. 1976. Available: https://www.jstor.org/stable/24950445
3. Y. Kumar, S. C. Srivastava, D. Gupta, W. Singh, A. K. Pandey, and D. Arora, "Analysis of soil fertility reduction using AI techniques: Opportunities and challenges," AIP Conference Proceedings, vol. 3297, p. 030008, 2025, doi: 10.1063/5.0287014.

4. H. Chandra, P. M. Pawar, R. Elakkiya, P. S. Tamizharasan, R. Muthalagu, and A. Panthakkan, “Explainable AI for soil fertility prediction,” IEEE Access, vol. 11, pp. 97866–97878, Sep. 2023, doi: 10.1109/ACCESS.2023.3310115.
5. V. D. N. Varshitha and S. Choudhary, “An AI solution for soil fertility and crop friendliness detection and monitoring,” International Journal of Engineering and Advanced Technology (IJEAT), vol. 10, no. 3, pp. 172–176, Feb. 2021, doi: 10.35940/ijeat.C2270.0210321.
6. K. Gunasekaran, K. A., and P. Sreevardhan, “Real-time soil fertility analysis, crop prediction, and insights using machine learning and deep learning algorithms,” Frontiers in Soil Science, vol. 5, 2025, doi: 10.3389/fsoil.2025.1652058.
7. M. Awais, S. M. Z. A. Naqvi, H. Zhang, L. Li, W. Zhang, F. A. Awwad, E. A. A. Ismail, M. I. Khan, V. Raghavan, and J. Hu, “AI and machine learning for soil analysis: An assessment of sustainable agricultural practices,” Bioresources and Bioprocessing, vol. 10, no. 1, p. 90, Dec. 2023, doi: 10.1186/s40643-023-00658-5.
8. S. B. Charbhe, C. Konduri, R. Maurya, and R. Deshmukh, “Estimation of soil nutrients and fertility using IoT and AI-based algorithm: A review,” in Micro and Nanoelectronics Devices, Circuits and Systems (MNDCS 2024), Lecture Notes in Electrical Engineering, vol. 1382, pp. 469–478, 2025.
9. L. Singh, S. Kumari, and M. Rai, “Using AI robotics for soil analysis and fertility management in agriculture,” in Artificial Intelligence Applications in Agriculture, J. K. Pandey, M. Rai, and T. Sarkar, Eds. Hoboken, NJ, USA: Wiley, 2025, ch. 13, doi: 10.1002/9781394320905.ch13.
10. A. M. Pathumuthusabana and S. S. Priyadharsini, “An artificial intelligence-based evaluation of soil fertility,” in Handbook of Green Engineering Technologies for Sustainable Smart Cities, 1st ed., Boca Raton, FL, USA: CRC Press, 2021, pp. 21–41.