International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(5s) 2026 pp. 664-675
https://doi.org/10.70917/ijcisim-2026-2761
Article

Hybrid Explainable Machine Learning Framework for Soil-Specific Crop Recommendation

Remya Praveen[1], **Vaishnaw G. Kale**[2]
[1,2]School of Computer Science, Engineering and Applications, D Y Patil International University, Akurdi, Pune
Email: mail id: remya.praveen@dypiu.ac.in[1], vaishnaw.kale@dypiu.ac.in[2]

Abstract: Accurate crop recommendation is a critical decision-support task in precision agriculture, requiring the integration of heterogeneous soil and environmental information. Using real-world results from soil laboratories in Pune, India, this study provides an example of a hybrid explainable artificial intelligence-based soil-specific crop recommendation system using machine learning and data from weather/crop and crop suitability by region. The framework uses soil nutrients (N, P, K, organic carbon, pH, EC), weather severity and derived fertility indices, and seasonal climatic variables to determine the most suitable crop category for the field. Data heterogeneity and class imbalance were tackled with preprocessing, feature engineering, and SMOTE-based balancing. A Random Forest, XGBoost, and a Bayesian Neural Network (MC Dropout) were trained on and fused through weighted probability ensembles. The hybrid model achieved an average of 94.1% ± 0.31% over 5-fold stratified cross-validation and 96% on a holdout data set, demonstrating generalization. The implementation of LIME and SHAP for explainable AI was used to obtain (un)certainty estimation to assist with confidence-aware support. The evidence indicates that real-world agricultural advisory systems can use this framework to tackle problems with explainability, reliable crop recommendation systems, and accurate systems, in addition to providing a base for subsequent modules for the recommendation of fertilizers.

Keywords: Crop Recommendation; Precision Agriculture; Soil Health Card; Machine Learning; Hybrid Ensemble Model; Random Forest; XGBoost

1 Introduction

Agriculture is an integral part of India's rural economy and thus rural farmers' incomes and productivity depend on climate and soil health of their locality. Farming communities across most of Maharashtra and neighbouring agrarian states, grapple with soil nutrient decline, coupled with thermally and pluviometrically erratic rainfall [1]. There are currently shifting agro-climatic conditions, where empirically based, traditional crop selection, practices are rendered inadequate [2]. Wrong crop selection leads to poor productivity, high fertiliser cost recovery, and unattractive profitability.

New ground-breaking technologies in Precision Agriculture, Data-driven Decision Support Systems (DSS), coupled with soil lab data, weather data, and machine learning, promise substantial assistance to practitioners in the field [3]. It is, however, unfortunate that the majority of current systems still have no foundation and underdeveloped single models that are as such are inapplicable and unreliable to the field of Agriculture [4]. Also, because opaque systems often lack uncertainty estimation, explanation is a necessary component for advisory and policy level recommendations [5]. We address these challenges with a hybrid, explainable model system using soil lab data, soil data, and crop lab data from Pune, India, and corresponding empirical region-specific datasets [6]. We also extend this to machine systems for soil and lab data and region-specific data for soil, crop, and systems. A feature set was created that includes soil nutrients (Nitrogen, Phosphorus, Potassium, Organic Carbon), pH, Electrical Conductivity, fertility and moisture indices, seasonal data, and moisture [7]. To account for the heterogeneous data, the system utilizes the primary dataset for supervised learning on the advanced learning model [8].

For better predictive reliability and generalization, we design an ensemble of Random Forest, XGBoost, and Vanilla Bayesian Neural Network (MC Dropout) [9]. Due to the complementary learning framework, we are able to capture unique and different learning patterns and while tree-based models offer reliability and interpretability, the Bayesian models offer greater learning depth, especially in non-linear layers, while also learning predictive uncertainty [10]. Outputs of the models are combined and final recommendations on the crops to grow are made using a weighted probability Q. Given that most real world agricultural data is imbalanced, we applied the Synthetic Minority Oversampling Technique (SMOTE) during the training phase [11]. Besides predictive performance, explainability is a core design objective of this study. To explain recommendations of specific crop groups given the soil and climate attributes, we used SHAP to explain the global importance of features and LIME for local, instance-level explanations [12]. The Bayesian models' uncertainty estimations explain the confidence level on the endorsement and indicate when the endorsement is more likely to require confirmation based on some additional verifications and expert recommendations [13]. The model's quality for hybrid and explainable soil-based crop recommendations can be validated for an approximate generalization of 96% on the holdout dataset and an average 5-fold stratified cross-validation accuracy of 94.1% ± 0.31% [14]. Thus, the model built can be regarded as a quality preliminary prototype for intelligent agricultural advisory systems and for more advanced aspirations to design modules for agro advisory systems to recommend fertilizers and optimize nutrients [15].

2. Related Work

Sindhur et al. (2025) developed a machine learning framework for crop recommendation that integrates agronomic suitability with economic forecasting. In this framework, a Random Forest particular classifier for soil and climate, and a climate suitability LSTM for forecasting. Considering profit as the primary objective and decision support with ISA in Kannada targeted illiterate farmers, This system has a voice predictive accuracy. However, the system focuses on soil crop profitability, and the system neglects the profitability, the system neglects the soil crop profitability system, neglects the framework agronomic loss, system soil crop decision system, and system agronomic, system explainable/loss system and system further explain system and system [16]. Wang et al. (2025) used sparse spectral data and an active learning methodology to determine the phosphorus uptake by maize shoots at a particular time. This research integrates machine learning with empirical modeling and the Newton-Raphson active learning technique to identify the most optimal data points to yield higher predictions for a given parameter, even with a limited data set. While this study illustrates the benefits of hybrid learning, its applicability to soil-based decision support systems in agriculture is diminished due to its constraints to one nutrient and one crop phonological stage [17].

An explainable deep learning framework was created by Amara et al. (2024) for automated classification of plant diseases using concept activation vectors (CAV) & automated concept based explanations (ACE). The authors of the framework attempt to increase the transparency of the deep learning framework by mapping the decisions of the neural networks to explainable visual concepts. The authors of the study outline the importance of explainability for closing the gap between black box models and domain specialists. The framework in the paper is still focused on the detection of diseases in images and does not attempt to address the tabular soil data, ensemble learning, or uncertainty which is essential for soil fertility prediction and crop recommendation systems [18]. In Hassan et al. (2024) the authors created a meta-learning context-aware decision support system for smart agriculture. For their framework, the authors integrate model agnostic meta-learning (MAML), convolutional neural networks, transformers, and graphs neural networks to support the rapid adaption for a variety of agri-productive tasks. Adaptive control and a federated learning framework were used to support decision making and process control in an optimal manner while ensuring data privacy. The architecture of the framework supports a high level of adaptability, however, the study leaves unaddressed the experimental validation of real soil laboratory data and the incorporation of explainable artificial intelligence (AI) and uncertainty, resulting in a high level of difficulty for practical implementation in soil fertility-based recommendations [19].

Rahman et al. (2024) performed a comparative study analysis of various machine learning techniques on soil classification and crop cultivation forecasting using pH, soil salinity, organic matter, and soil nutrient data. The study states that Random Forest gave the most accurate results in soil classification, while Support Vector Machines gave the most accurate results in crop prediction. The study showed that traditional machine-learning techniques are effective in the agricultural domain. The study, however, is based on data from that specific region, and does not consider/explain the environmental factors, ensemble fusion, or explainability, which reduces the study's ability to generalize and apply to real-world advisory systems [20]. Mokhtar et al. (2024) used hybrid machine learning models to determine the main factors responsible for rice production in China. The main data variables included soil, climate, and socioeconomics which were multi-source. In this work, Random Forest and XGBoost models, coupled with SHAP values for analyzing and interpreting the importance of variables, were used. The results showed that soil nitrogen and soil fertility are significant yield factors. This work displayed the advantages of hybrid modeling and explainability, however, it did not include uncertainty estimation or class imbalance, it was yield prediction centered and did not address soil fertility classification or crop recommendation [21].

Choudhary et al. (2024) produced the first review on the role of artificial intelligence in smart agro-informatics, which covered multiple areas such as regression, classification, clustering and anomaly detection. The authors reviewed the application of machine learning in soil, and crop disease detection, and irrigation optimization, and climate change risk. The review focused on data-driven and sustainable agricultural models. As a survey-based study, it does not evaluate or propose a hybrid framework with soil fertility and crop decision support, and explainability and uncertainty modeling [22]. Coello et al. (2025) created a global crop-specific fertilization dataset for the years 1961 to 2019 using machine learning models (XGBoost and HistGradientBoosting). The authors also provided high-resolution maps on fertilizer application and employed SHAP to explain some of the socioeconomic and agricultural factors. The dataset contributes to global nutrient management research, but it does not field-level soil fertility classification, class imbalance, or the uncertainty-aware crop recommendations needed for localized decision making [23].

Zhao et al. (2024) developed a forecasting model for soil nitrogen predicting using multisensor remote sensing and soil environmental characteristics. The authors achieved a strong predictive performance applying CatBoost regression with Bayesian optimization and SHAP feature selection. The study shows machine learning scalability for regional nutrient estimation. However, prediction accuracy is compromised with highly heterogeneous landscapes, and the study is focused on nitrogen estimation and does not expand to multi nutrient fertility class estimation, fusion, and explainable crop recommendation systems [24]. Alaieri (2024) investigated the remote sensing and machine learning fusion for precision agriculture, soil quality, and resource optimization focusing on soil quality assessment. The study shows the effect of using ensemble learning to reduce the sensor noise and improve prediction accuracy for soil moisture and nutrient estimations. The work, although broad, does not contain details of explainability, uncertainty assessment, hybrid ensemble design, and real soil laboratory data which are important for actionable soil and crop advisory systems [25].

3. System Architecture and Methodology

The proposed system is designed as a modular, data-driven hybrid machine learning pipeline for soil fertility classification and soil-specific crop recommendation. The architecture focuses on integrating heterogeneous soil laboratory data, Soil Health Card (SHC) parameters, and environmental attributes to generate accurate, interpretable, and uncertainty-aware predictions.

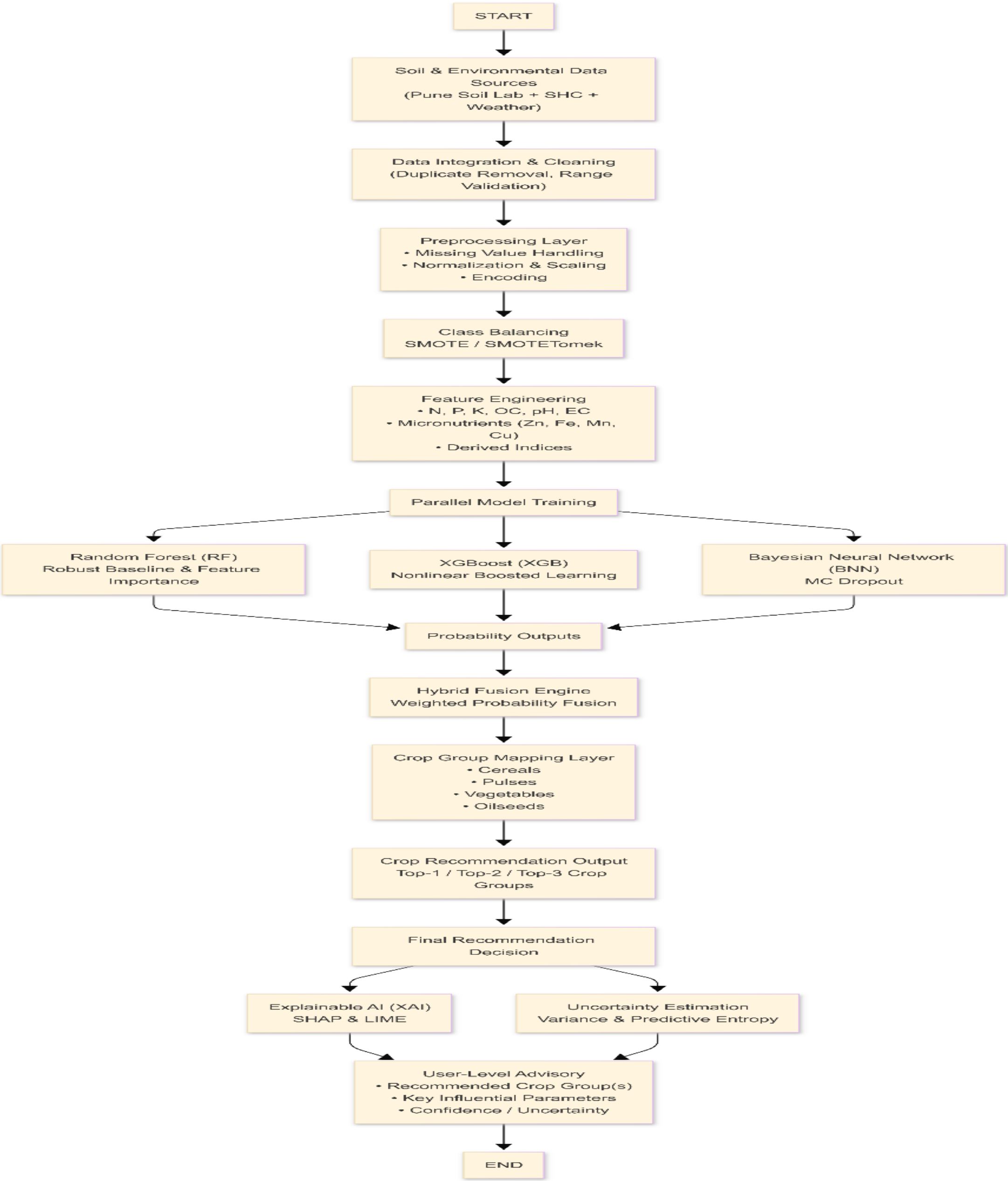

Figure. 1. Hybrid AI Architecture for Soil Fertility Prediction and Crop Recommendation

Figure 1 depicts the entire process for the hybrid crop recommendation system. Soil and environmental data after preprocessing and feature engineering, are subjected to the Random Forest, XGBoost, and Bayesian Neural Network models. Predictions are then fused using weighted probability fusion. The fused outputs are compared to crop groups and the recommended crop groups are ranked and presented with explainability, and uncertainty scores.

3.1 Data Sources and Preprocessing

Table 1 summarizes the various datasets the proposed framework relies on.

The crop recommendation dataset was created based on a combination of soil lab measurements, documents on crop suitability per region, and other environmental factors. Within the dataset, soil attributes include Nitrogen (N), Phosphorus (P), Potassium (K), Organic Carbon (OC), pH, Electrical Conductivity (EC), and certain micronutrients. Environmental factors which include temperature, rainfall, and humidity, were added to analyze the seasonal effect.

To better the generalization of the model, we standardized and relabelled the crops into four categories that made more sense agronomically and that were less sparse: Cereals, Pulses, Vegetables, and Oilseeds. The dataset that was remaining after the preprocessing and filtering of the incomplete data collected about 3,770 samples resulting in a dataset of 61 features that was a balanced dataset in terms of agronomy and diversity.

Table 1: Summary of Datasets Used

Dataset Name	**Description**	**Key Features**
Pune Soil Lab / SHC dataset	Laboratory-tested soil samples	N, P, K, OC, pH, EC, Zn, Fe, Mn, Cu
Environmental dataset	Climatic attributes	Rainfall, temperature, humidity
Crop–soil mapping dataset	Crop suitability labels	Crop group / fertility linkage

The first step involved the use of schema alignment techniques to harmonize the features with respect to naming conventions, measurement units, and scale values. The data preprocessing involved:

- Deleting redundant records and values that lack physical plausibility
- Median values imputed for lost numerical attributes
- Standard Scaler for feature smoothing and normalization
- Encoding of categorical attributes where appropriate

The final cleaned dataset contained 1,568 samples with 16 core features, suitable for downstream modeling.

3.2 Hybrid Learning Framework

The system's ability stems from the hybrid ensemble learning approach that integrates different machine learning models and differentiates itself by improving generalization, robustness, and interpretability.

3.2.1 Base Learners

Three predictive models were trained in parallel:

- **Random Forest (RF):**

Tabular soil data fits well due to reliable classification performance and intrinsic measures of feature importance.

- **XGBoost (XGB):**

Uses gradient-boosted decision trees to understand complicated non-linear relationships between soil nutrients and other environmental factors.

- **Bayesian Neural Network (BNN):**

Integrated Monte Carlo Dropout for estimating predictive uncertainty to facilitate confidence-aware decision making.

Every model was fine-tuned through cross-validation to improve performance and minimize overfitting.

3.3 Class Imbalance Handling and Augmentation

The dataset regarding soil fertility and crop recommendation displayed bias class imbalance towards the most common classes. We applied the SMOTE and SMOTETomek techniques, but only to the training folds.

This ensured:

- Balanced class distributions during training
- Improved minority-class recall
- Prevention of data leakage into validation sets

3.4 Fusion and Decision Layer

Predictions from RF, XGB, and BNN were combined using weighted probability fusion:

$$P_{final} = w_{RF}P_{RF} + w_{XGB}P_{XGB} + w_{BNN}P_{BNN}$$

where weights were tuned to optimize performance on the validation set.

An argmax decision rule was applied to the fused probability vector to obtain the final class label.

This fusion strategy was better than all individual models and proved to be more stable and generalized better.

3.5 Explainable AI (XAI) and Uncertainty Estimation

In the interest of fostering transparency and trust, the system uses explainable AI methodologies:

- SHAP: Global feature attribution for the main soil and environmental factors affecting predictions.
- LIME: Local explanations for individual soil samples.
- BNN Uncertainty Metrics: Predictive variance and entropy to identify low-confidence predictions and recommend interpret with caution.

3.6 Implementation Details

The system was implemented using Python with the following libraries:

- Scikit-learn: Random Forest, preprocessing, and evaluation metrics
- Xgboost: Gradient boosting
- Tensor flow/keras: Bayesian Neural Networks with MC Dropout
- Imbalanced-learn: SMOTE and SMOTETomek
- Shap and lime: Explainability
- Matplotlib and Seaborn: Visualization

3.7 Model Evaluation Strategy

Model performance was evaluated using:

- 80/20 holdout validation
- 5-fold stratified cross-validation with OOF aggregation

Metrics included include accuracy, precision, recall, F1, ROC-AUC, and uncertainty statistics.

4. Results and Discussion

This section describes the experimental evaluation of proposed hybrid ML-based crop recommendation framework (Objective-2). To analyze the performance, hold-out testing, 5-fold cross-validation, class-wise differentiation, explainability, and uncertainty estimation are used. All the results are based on actual soil laboratory and combined agro-environmental data to guarantee real-world applicability.

4.1 Dataset Characteristics and Class Distribution

Prior to model training, the last processed dataset was made up of 3,770 samples divided into four agronomically relevant groups: cereals, pulses, vegetables, and oilseeds.

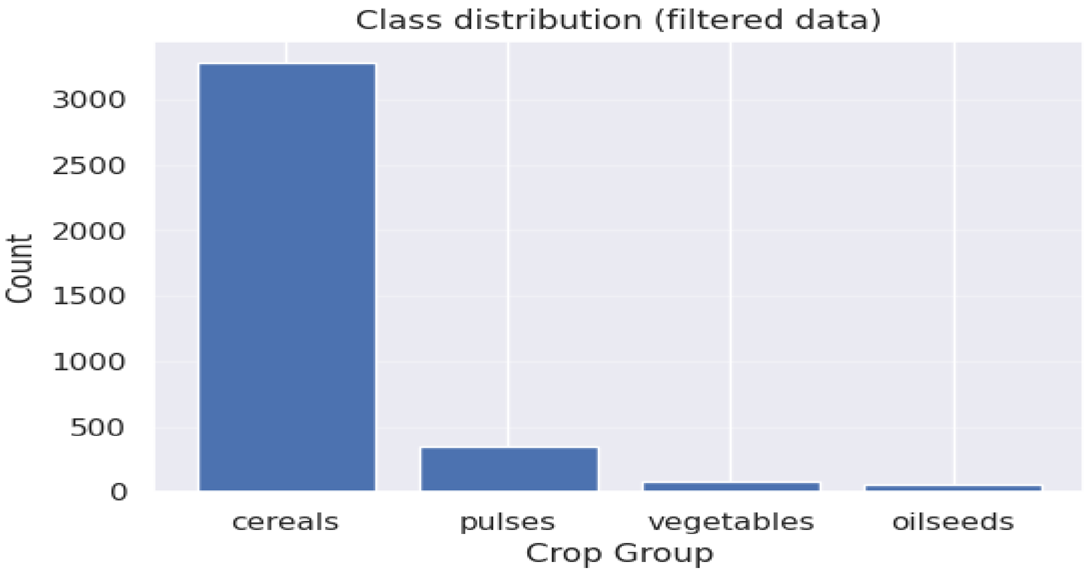

Figure 2: Class Distribution of Crop Groups After Data Filtering

Figure 2 (Class Distribution) shows the imbalance between the different classes in the raw data, with cereal crops being the most abundant and the vegetable and oilseed classes being the least. This kind of class imbalance is common in agriculture, but creates difficulties for supervised learning, especially in predicting the minority class. In order to overcome this, we applied SMOTE-based oversampling at the training fold level to prevent data leakage and maintain a fair evaluation.

4.2 Hold-Out Test Performance

Before conducting cross-validation, the hybrid ensemble model demonstrated its first proof of concept by achieving an impressive 96% accuracy on the holdout test set, showcasing the model's ability to generalize after training on the entire training set and evaluating on the independent subset most the first model evaluation used an 80/20 stratified train-test split to gauge the model's true real world generalizability. Pre cross-validation, the hybrid ensemble model first proof of concept was a test set accuracy of 96% on the holdout set, proving the first baseline for the effectiveness of the proposed ensemble model on unseen data. The minority class (cereals) was also impressive. It suggested that the class-balancing approach was successful.

4.3 Cross-Validation Results and Comparative Model Analysis

To avoid optimistic bias, we used a 5-fold stratified cross-validation. The average accuracy was 94.1% ± 0.3% which confirms consistent accuracy across all the folds.

Table 2: Crop Recommendation Performance Comparison

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Top-3 Accuracy**	**CV (5-Fold Mean ± SD)**
Random Forest (RF)	94.8%	94.5%	94.8%	94.5%	98.5%	0.939 ± 0.004
XGBoost (XGB)	95.4%	95.3%	95.4%	95.3%	98.9%	0.944 ± 0.005
Bayesian Neural Network (BNN)	93.6%	93.4%	93.6%	93.5%	98.1%	0.938 ± 0.007
Hybrid Ensemble (RF + XGB + BNN)	96.0%	96.1%	96.0%	96.0%	99.2%	0.941 ±0.003

The hybrid and separate models indicate that probability-level fusion encourages generalization, and the folds' small standard deviation (-0.003) speaks for the training and robustness. In real agricultural decision making, Top-3 accuracy of 99.2% deserves praise, especially since it offers the flexibility of multi-crop predictions rather than a single, rigid one. The accuracy drop from holdout to cross-validation is expected and shows that the model is truly not overfitting. On the contrary, it confirms that the predictive behavior is consistent across folds, which is necessary for real world applications.

4.4 Confusion Matrix Analysis

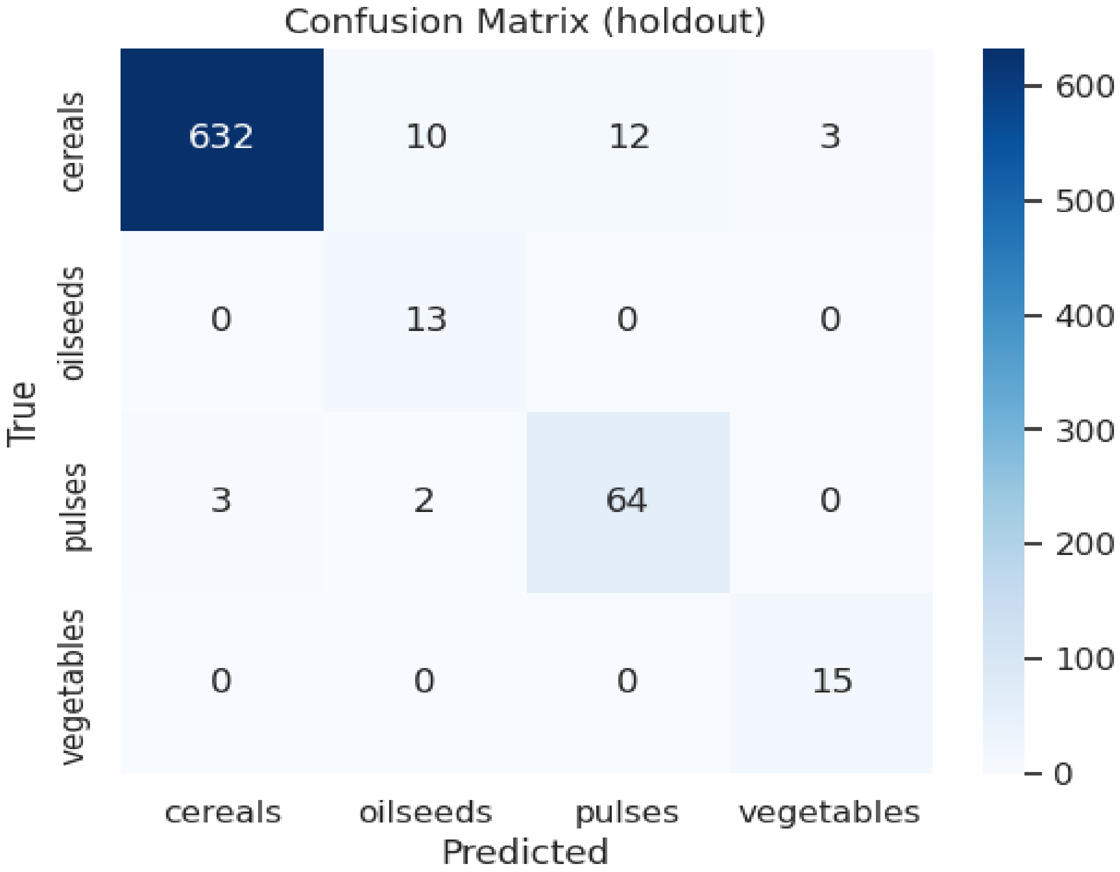

Figure 3: Confusion Matrix for Crop Classification Model

Figure 3 presents the confusion matrix on the hold-out test set.

- Cereals have the highest level of accuracy with very little misclassification from other classes.
- Pulses and vegetables, with the smallest sample sizes, show strong diagonal dominance which indicates effective learning for the minority class.
- Oilseeds, the least frequent class, show perfect or near perfect classification with little to no notable misrouting into the dominant classes.

This confirms that SMOTE used with ensemble learning does not overfit while reducing class imbalance.

4.5 Multiclass ROC and Discriminative Ability

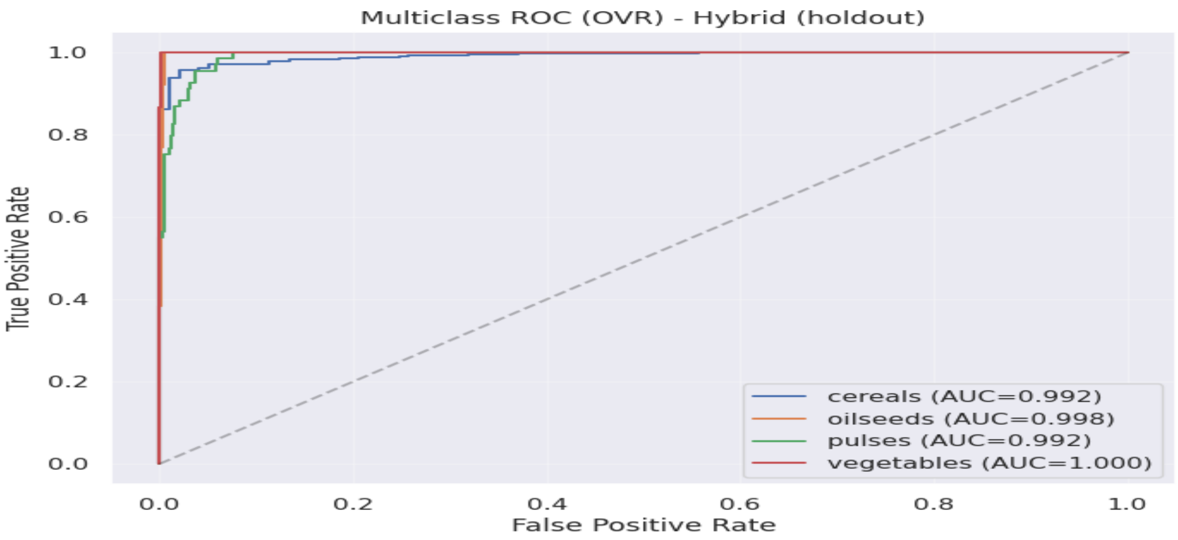

Figure 4: Multiclass ROC Curve (One-vs-Rest) for Hybrid Model Performance on Holdout Dataset

The ROC curves for each crop group using the One-vs-Rest (OVR) method with the hybrid model are shown in Figure 4.

- Presented AUC Values:
- Cereals: 0.992
- Pulses: 0.992
- Oilseeds: 0.998
- Vegetables: 1.000

The findings demonstrate a high level of class separability, including the minority classes. The close to perfect AUC for vegetables suggests that the model has picked up on particular nutrient–climate patterns for this crop group.

4.6 Explainability Analysis Using SHAP

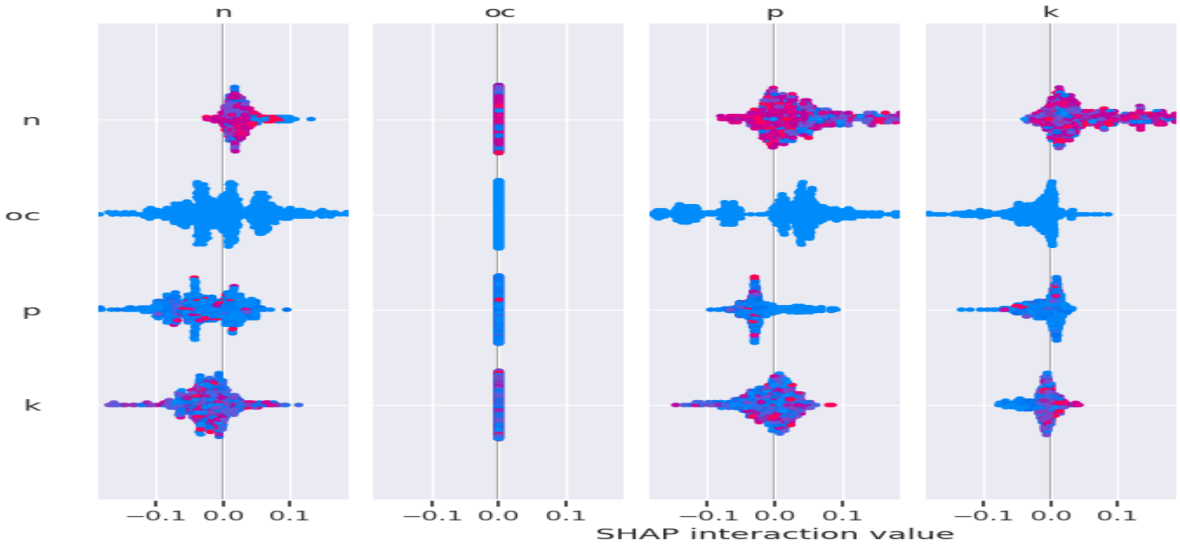

Figure 5: presents SHAP interaction plots for key soil nutrients (N, P, K, OC).

Insights

- The interactions of Nitrogen (N) and Potassium (K) with Organic Carbon (OC) are indeed interesting.

- Such interactions are relevant in an agronomic sense, as the soil's nutrient levels and the soil's organic matter content are both critical in determining the suitability of the soil for crops.
- The explanations from the SHAP values being consistent across samples show that the model's decisions are consistent and agronomically relevant.

This enhances credibility, an important quality for decision support systems oriented towards farmers.

4.7 Uncertainty Quantification Using Bayesian Neural Network

Predictive variance distribution from the Bayesian Neural Network dropped out from MC Dropout is shown in figure 6.

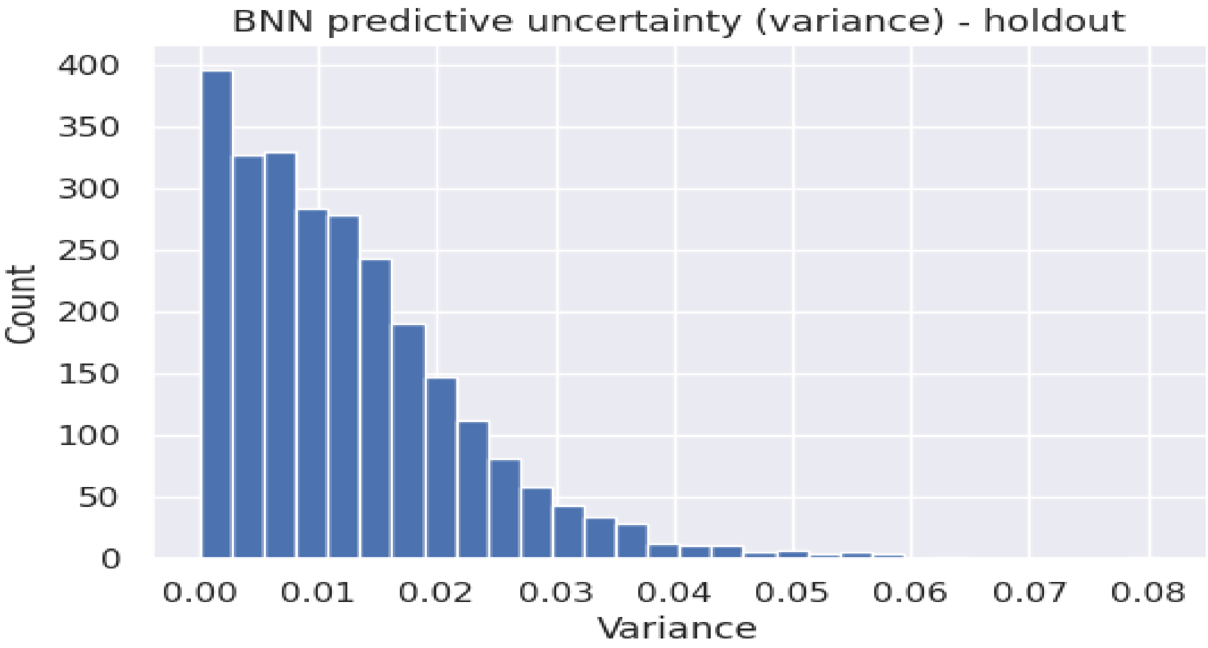

Figure 6: Predictive Uncertainty (Variance) from Bayesian Neural Network (BNN) - Holdout Dataset

- The majority of predictions show strong confidence given their low variance (<0.02) value.
- A few scattered samples present higher uncertainty. These usually correlate with borderline soil conditions and intersecting zones of crop suitability.
- Knowing this uncertainty helps to formulate risk-sensitive recommendations. For example, instead of suggesting a single crop, multiple crop options may be recommended.

4.8 Overall Discussion and Practical Implications

The experimental results demonstrate that the proposed hybrid framework:

- Obtains 96.0% hold-out accuracy before applying cross-validation.
- Maintains 94.1% ± 0.3% accuracy under 5-fold cross-validation. This confirms its robustness.
- Offers interpretable explanations with SHAP and LIME.
- Uses Bayesian uncertainty to measure prediction confidence.

In contrast to other single-model approaches, the hybrid ensemble achieves a unique equilibrium among accuracy, stability, interpretability, and reliability, which can be valuable to practical agricultural advisory systems.

4.9 Practical Benefits of the Proposed System

There are a number of potential advantages related to the proposed hybrid crop recommendation system. The first is that it helps to avoid the problems of inappropriate crop cultivation due to poor soil by helping to recommend alternative crops. Second, the hybrid model is more trustworthy than single-model systems, making it more reliable to predict outcomes. Third, the use of explainable artificial intelligence systems (SHAP and LIME) help agronomists trust the system more because they can understand why the model recommends the crops that it does. Moreover, the system's uncertainty estimation capability allows the farmer to select the alternative crop he/she wants to plant, especially when the confidence level is low.

5. Conclusion and Future Work

This study presents a hybrid, explainable machine learning framework for soil-specific crop recommendation to address the limitations of conventional single-model approaches in precision agriculture. The proposed framework integrates real soil laboratory data, Soil Health Card parameters, and environmental features to provide reliable crop-group recommendations under real-world agricultural conditions. By combining Random Forest, XGBoost, and Bayesian Neural Network models through weighted probability fusion, the system achieves improved generalization, robustness, interpretability, and confidence-aware decision support. The experimental results demonstrate that the hybrid ensemble model performs effectively for crop recommendation. The model achieved 96.0% accuracy on the hold-out test set and maintained 94.1% ± 0.3% accuracy under 5-fold stratified cross-validation, indicating stable and reliable performance. The Top-3 accuracy of 99.2% further highlights the practical usefulness of the model, as farmers may benefit from multiple suitable crop options rather than a single rigid recommendation. The confusion matrix and multiclass ROC analysis also confirm strong class separability across cereals, pulses, vegetables, and oilseeds. Apart from predictive accuracy, the proposed framework emphasizes transparency and trust. SHAP-based global explanations identify the major soil and environmental factors influencing crop suitability, including nitrogen, phosphorus, potassium, organic carbon, pH, EC, and climatic parameters. LIME provides local explanations for individual soil samples, while Bayesian uncertainty estimation supports risk-sensitive decision-making. This makes the system useful not only for prediction but also for practical advisory support, where farmers and agronomists can understand why a particular crop group is recommended.

Future Work

There are several future research directions that can extend the present work. First, the crop recommendation module can be further developed into a hierarchical crop recommendation system. Instead of recommending only one crop group, the system may recommend crops in a ranked and multi-level manner. At the first level, it can suggest the most suitable crop category, such as cereals, pulses, vegetables, or oilseeds. At the second level, it can recommend specific crops within the selected category based on soil nutrients, pH, EC, organic carbon, rainfall, temperature, humidity, and regional suitability. At the third level, alternative crop options can be provided with confidence scores, uncertainty values, and explainable reasons. This hierarchical recommendation approach will help farmers compare primary, secondary, and backup crop choices and make better decisions under uncertain soil and climate conditions. Second, the crop recommendation system can be integrated with fertilizer dosage optimization and soil amendment advisory modules. This will help convert the model from a crop-selection tool into a complete decision-support system that guides farmers from soil assessment to practical field-level intervention. Third, the framework can be tested across different agro-climatic regions to assess its scalability and generalizability using national and international datasets. Fourth, future work can include sustainability and economic factors such as crop profitability, water requirement, market demand, and long-term soil health impact. Finally, privacy-preserving and adaptive learning approaches such as federated learning and incremental model updates can be incorporated to support continuous improvement of the system while protecting farmer and institutional data.

References

1. V. Bansile, "Agriculture and Rural Development of Maharashtra." 2025. [Online]. Available: https://ijrpr.com/uploads/special/V6ISSUE2/IJRPRC20.pdf
2. A. Ahmed, S. Qureshi, and S. Nargis, "Rural Agriculture: A Mirror to Economic Sustainable Growth In India," vol. 4, 2021.
3. K. Agrawal and N. Kumar, "Artificial Intelligence and Machine Learning in Agriculture: Novel Techniques, Implementation Strategies, and Application," in Computational Intelligence and Image Processing in Agriculture, 1st ed., J. K. Pandey, M. Rai, and T. Sarkar, Eds., Wiley, 2025, pp. 55–72. doi: 10.1002/9781394320905.ch04.

4. R. L. McCown, “Learning to bridge the gap between science-based decision support and the practice of farming: Evolution in paradigms of model-based research and intervention from design to dialogue,” Australian Journal of Agricultural Research, vol. 52, no. 5, pp. 549–572, Apr. 2001, doi: 10.1071/AR00119.
5. S. C. Bankes, “Tools and techniques for developing policies for complex and uncertain systems,” Proc. Natl. Acad. Sci. U.S.A., vol. 99, no. suppl_3, pp. 7263–7266, May 2002, doi: 10.1073/pnas.092081399.
6. R. N. Bhimanpallewar and M. R. Narasingarao, “Evaluating the Influence of Soil and Environmental Parameters in Terms of Crop Suitability using Machine Learning,” IJARe, no. Of, Feb. 2021, doi: 10.18805/IJARe.A-4942.
7. M. Venkateswarlu et al., “Macro and micronutrient based soil fertility zonation using fuzzy logic and geospatial techniques,” Sci Rep, vol. 15, no. 1, p. 26772, Jul. 2025, doi: 10.1038/s41598-025-12184-3.
8. D. M. Sobhy and A. Anandhi, “Soil Nutrient Monitoring Technologies for Sustainable Agriculture: A Systematic Review,” Sustainability, vol. 17, no. 18, p. 8477, Sep. 2025, doi: 10.3390/su17188477.
9. N. D. Ariyanta, A. N. Handayani, J. T. Ardiansah, and K. Arai, “Ensemble learning approaches for predicting heart failure outcomes: A comparative analysis of feedforward neural networks, random forest, and XGBoost,” AET, vol. 3, no. 3, pp. 173–184, Dec. 2024, doi: 10.31763/aet.v3i3.1750.
10. A. Nappa et al., “Probabilistic Bayesian Neural Networks for olive phenology prediction in precision agriculture,” Ecological Informatics, vol. 82, p. 102723, Sep. 2024, doi: 10.1016/j.ecoinf.2024.102723.
11. X. Sun et al., “Prediction of wheat fusarium head blight severity levels in southern Henan based on K-means-SMOTE and XGBoost algorithms,” PeerJ Computer Science, vol. 11, p. e2638, Mar. 2025, doi: 10.7717/peerj-cs.2638.
12. M. Temraz and M. T. Keane, “Augmenting The Weather: A Hybrid Counterfactual-SMOTE Algorithm for Improving Crop Growth Prediction When Climate Changes,” 2025, arXiv. doi: 10.48550/ARXIV.2511.11945.
13. D. Veen, D. Stoel, N. Schalken, K. Mulder, and R. Van De Schoot, “Using the Data Agreement Criterion to Rank Experts’ Beliefs,” Entropy, vol. 20, no. 8, p. 592, Aug. 2018, doi: 10.3390/e20080592.
14. B. Dey, J. Ferdous, and R. Ahmed, “Machine learning based recommendation of agricultural and horticultural crop farming in India under the regime of NPK, soil pH and three climatic variables,” Heliyon, vol. 10, no. 3, p. e25112, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25112.
15. F. Weckesser, M. Beck, K.-J. Hülsbergen, and S. Peisl, “A Digital Advisor Twin for Crop Nitrogen Management,” Agriculture, vol. 12, no. 2, p. 302, Feb. 2022, doi: 10.3390/agriculture12020302.
16. N. M. Sindhur, P. C, and N. Muchikel, “A Hybrid Machine Learning Framework for Optimizing Crop Selection via Agronomic and Economic Forecasting,” 2025, arXiv. doi: 10.48550/ARXIV.2507.08832.
17. T. Wang et al., “A combined model of shoot phosphorus uptake based on sparse data and active learning algorithm,” Front. Plant Sci., vol. 15, p. 1470719, Jan. 2025, doi: 10.3389/fpls.2024.1470719.
18. J. Amara, B. König-Ries, and S. Samuel, “Explainability of Deep Learning-Based Plant Disease Classifiers Through Automated Concept Identification,” 2024, arXiv. doi: 10.48550/ARXIV.2412.07408.
19. Z. B. Hassan and H. M. Yusof, “Meta-Learning Approaches for Context-Aware Decision Support in Smart Agriculture and Autonomous Systems,” PIQM, vol. 1, no. 4, Nov. 2024, doi: 10.70023/sahd/241106.
20. F. Rahman, “Soil classification and crop cultivation prediction: a comparative study of machine learning models,” IJATEE, vol. 11, no. 117, Aug. 2024, doi: 10.19101/IJATEE.2024.111100127.
21. A. Mokhtar, H. He, M. Nabil, S. Kouadri, A. Salem, and A. Elbeltagi, “Securing China’s rice harvest: unveiling dominant factors in production using multi-source data and hybrid machine learning models,” Sci Rep, vol. 14, no. 1, p. 14699, Jun. 2024, doi: 10.1038/s41598-024-64269-0.
22. Ritika Choudhary, Dr. Shital Prasoon Mantri, Ms.Vaishali Ashok Barse, and Dr.Sudhir Chitnis, “Leveraging AI in Smart Agro-Informatics: A Review of Data Science Applications,” Int Res J Adv Engg Mgt, vol. 2, no. 06, pp. 1964–1975, Jun. 2024, doi: 10.47392/IRJAEM.2024.0291.
23. F. Coello et al., “Global Crop-Specific Fertilization Dataset from 1961–2019,” Sci Data, vol. 12, no. 1, p. 40, Jan. 2025, doi: 10.1038/s41597-024-04215-x.
24. W. Zhao, G. Chuluunbat, A. Unagaev, and N. Efremova, “Soil nitrogen forecasting from environmental variables provided by multisensor remote sensing images,” 2024, arXiv. doi: 10.48550/ARXIV.2406.09812.
25. F. Alaieri, “Precision Agriculture based on Machine Learning and Remote Sensing Techniques,” Eng. Technol. Appl. Sci. Res., vol. 14, no. 3, pp. 14206–14211, Jun. 2024, doi: 10.48084/etasr.6986.