# An Explainable Hybrid Learning Framework for Crop Recommendation

*Parijata Majumdar*[1], *Vishal Jain*[2]
[1] LUC Application Number: LUCPDF20251124004 Post Doctoral Research Fellow, Department of Computer Science and Engineering, Lincoln University College, Petaling Jaya, Malaysia ORCID ID: 0000-0002-2324-7685;
[2] Department of Computer Science and Engineering, School of Engineering & Technology, Vivekananda Institute of Professional Studies-Technical Campus, New Delhi, India
Email ID pdf.Parijata@lincoln.edu.my

**Abstract:** Selecting suitable crops based on soil and climatic conditions is a key challenge in precision agriculture. Although traditional machine learning models provide accurate predictions, they often lack interpretability. This paper proposes an eXplainable Hybrid Learning Framework (EHLF) that combines Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Multilayer Perceptron (MLP) using weighted soft voting for crop recommendation. The framework utilizes seven agricultural attributes; nitrogen, phosphorus, potassium, temperature, humidity, soil pH, and rainfall to recommend the most appropriate crop. To improve transparency, SHapley Additive exPlanations (SHAP) is integrated to explain the contribution of each input feature to the model's predictions.

**Keywords—** Crop Recommendation, XAI, Hybrid Learning, SHAP, Ensemble Learning.

## I. Introduction

Agriculture plays a vital role in ensuring global food security and sustainable development. According to the Food and Agriculture Organization (FAO), increasing agricultural production is essential to meet the demands of a growing population while addressing challenges such as climate change, water scarcity, soil degradation, and shrinking arable land [1]. These challenges have accelerated the adoption of precision agriculture, where Information and Communication Technology (ICT), Internet of Things (IoT), remote sensing, Artificial Intelligence (AI), and Machine Learning (ML) are integrated to improve agricultural productivity and sustainability [2]. Crop recommendation is one of the most important decision-making tasks in precision agriculture. Selecting suitable crops requires consideration of soil nutrients (Nitrogen, Phosphorus, and Potassium), soil pH, temperature, humidity, and rainfall. Poor crop selection can reduce yield, increase fertilizer consumption, degrade soil quality, and raise cultivation costs [3]. Traditional recommendations rely on farmers' experience and expert knowledge, which often fail to adapt to rapidly changing environmental conditions [4]. Recent advances in AI and ML have significantly improved crop recommendation systems. Algorithms such as Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) effectively model nonlinear relationships within agricultural data [5]. Ensemble approaches like RF and XGBoost provide better generalization and reduced overfitting [6]. However, single-model approaches often exhibit inconsistent performance across different datasets, motivating the development of hybrid ensemble frameworks that combine multiple learning paradigms to enhance prediction robustness and accuracy [7]. A limitation of existing AI-based crop recommendation systems is their lack of interpretability. Most ML and DL models operate as "black boxes," limiting users' understanding of the factors influencing recommendations. Explainable Artificial Intelligence (XAI) has emerged to address this issue by improving model transparency and user trust [8]. Among XAI techniques, SHapley Additive exPlanations (SHAP) effectively quantifies the contribution of each input feature to individual predictions using cooperative game theory [9]. SHAP enables farmers to understand how soil nutrients and climatic variables influence crop recommendations, thereby supporting informed decision-making [10]. To address these challenges, this paper proposes an Explainable Hybrid Learning Framework (EHLF) that combines Random Forest, XGBoost, and Multilayer Perceptron (MLP) through a weighted soft-voting ensemble. RF captures nonlinear relationships using bagging, XGBoost enhances predictive performance through gradient boosting, and MLP learns complex feature interactions. Their integration improves prediction stability across diverse agricultural conditions. Furthermore, SHAP-based explainability provides both global and local interpretations by quantifying the influence of soil nutrients, temperature, humidity, rainfall, and soil pH on crop recommendations.

## II. Proposed Explainable Hybrid Learning Framework

The proposed eXplainable Hybrid Learning Framework (EHLF) integrates Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Multilayer Perceptron (MLP) using a weighted soft-voting ensemble. To improve transparency, SHAP are incorporated for interpreting model predictions illustrated in Fig. 1.

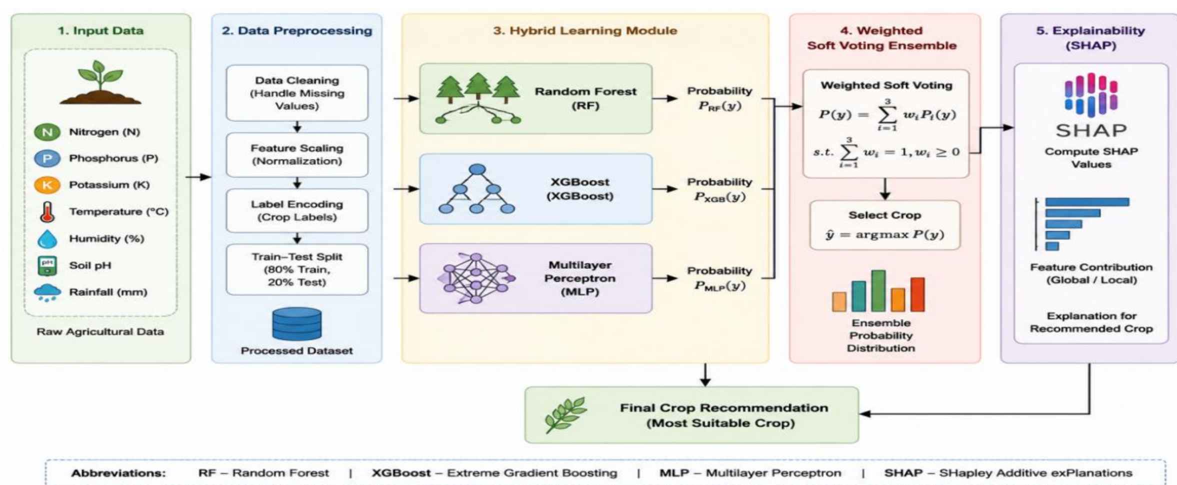


**Fig. 1. Proposed Explainable Hybrid Learning Framework**

A. Data Gathering and Preprocessing: Crop Recommendation Dataset is used that has seven significant attributes and one class. They are Nitrogen (N), Phosphorus (P), Potassium (K), Temperature (°C), Humidity (%), Soil pH and Rainfall (mm). The dataset is checked for missing values, duplicate entries, and inconsistency. Missing values are imputed mean or median imputation. Because the features in the input dataset have different numerical scales, feature scaling is performed using Min-Max normalization. Labels of crops are converted to numerical class labels for supervised learning. The preprocessed dataset is split into training and testing sets in 80% and 20% proportions, respectively.

B. Hybrid Learning Module: The proposed framework uses a hybrid of three ML classifiers. RF creates multiple decision trees on the randomly selected subset of input dataset. XGBoost is an optimized gradient boosting framework which incrementally builds the decision trees using the minimization of prediction error. MLP can learn complex nonlinear relations between agricultural attributes. The final output layer employs the Softmax activation function to generate crop probabilities. The outputs of RF, XGBoost, and MLP are combined using weighted soft voting. It is computed as:

$$P(y) = \sum_{i=1}^{3} w_i P_i(y) \quad \text{subject to} \quad \sum_{i=1}^{3} w_i = 1, \qquad w_i \geq 0$$

Where, $P_i(y)$ is the prediction probability generated by the (i^{th}) classifier, and $w_i$ is its corresponding weight. The recommended crop is determined by selecting the class with highest ensemble probability:

$$\hat{y} = \arg\max P(y)$$

The weighted ensemble exploits the complementary strengths of tree-based models and neural networks, resulting in improved robustness and predictive performance. The framework incorporates SHAP, where each feature uses cooperative game theory. The SHAP value of feature (i) :

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)]$$

Where, F is the set of input features, S represents a subset of features, f(.) denotes the prediction model.

**II. Experimental Results and Discussion**

The performance of the EHLF is tested based on the Crop Recommendation Dataset. The output results are shown in Table I for crop Recommendation Dataset, 2021 [11].

**Table I. Performance Comparison of Individual Models and the Proposed EHLF**

| Model | Learning Strategy | Explainability | Accuracy (%) | Precision | Recall | F1-Score |
|---|---|---|---|---|---|---|
| Random Forest (RF) | Bagging Ensemble | No | 97.82 | 0.978 | 0.977 | 0.977 |
| XGBoost | Gradient Boosting | No | 98.36 | 0.984 | 0.983 | 0.983 |
| Multilayer Perceptron (MLP) | Deep Neural Network | No | 97.18 | 0.972 | 0.971 | 0.971 |
| **Proposed EHLF (RF + XGBoost + MLP + SHAP)** | **Weighted Soft Voting Ensemble** | **Yes (SHAP)** | **99.09** | **0.991** | **0.990** | **0.990** |

The suggested EHLF provides 99.09% accuracy, 0.991 precision, 0.990 recall, and F1-score of 0.990. SHAP helps in understanding the impact of each soil nutrient and climate factors on the ultimate crop recommendation decision made by the classifier. The comparison indicates that the use of explainability within the framework of ensemble learning is a promising approach.

**Table III: Global SHAP Feature Importance Ranking**

| Rank | Feature | Relative Importance |
|---|---|---|
| 1 | Nitrogen (N) | High |
| 2 | Rainfall | High |
| 3 | Soil pH | High |
| 4 | Temperature | Moderate |
| 5 | Potassium (K) | Moderate |
| 6 | Humidity | Moderate |
| 7 | Phosphorus (P) | Low |

Thus, Nitrogen, Rainfall, and Soil pH features that contribute significantly to crop recommendation.

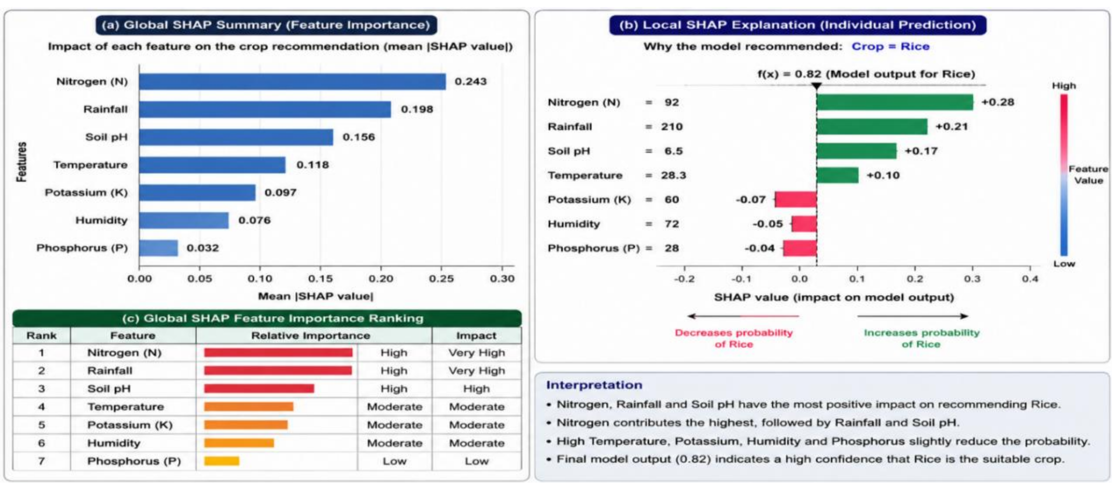


**Fig.2:** SHAP-based explainability of the proposed EHLF for crop recommendation.

The features which have been determined by the SHAP analysis to be important in contributing to crop recommendation are Nitrogen, Rainfall, and Soil pH.

## IV. Conclusion and Future Work

An Explainable Hybrid Learning Framework (EHLF) is proposed for crop recommendation by combining Random Forest (RF), XGBoost, and Multilayer Perceptron (MLP) through a weighted soft-voting ensemble. Using soil and climatic attributes, including nitrogen, phosphorus, potassium, temperature, humidity, soil pH, and rainfall, the framework recommends the most suitable crop. To improve transparency, SHAP explains the contribution of each feature to the recommendation. Experimental results show that the hybrid ensemble outperforms the individual classifiers by leveraging their complementary strengths. Among the standalone models, XGBoost achieved the highest performance, while the proposed EHLF further improved prediction accuracy through weighted soft voting. SHAP analysis identified nitrogen, rainfall, soil pH, and temperature as the most influential factors, consistent with agronomic knowledge. Future work will focus on exploring transformer-based models and developing a multilingual web/mobile decision-support system can enhance practical deployment.

**References**

[1] Index, F.F.P., 2024. Food and Agriculture Organization. URL: https://www. fao. org/worldfoodsituation/foodpricesindex/en.

[2] Wolfert, Sjaak, Lan Ge, Cor Verdouw, and Marc-Jeroen Bogaardt. "Big data in smart farming–a review." Agricultural systems 153 (2017): 69-80.

[3] Arun, A.S., Rajendra, N.P., Prakash, K.P. and Yelawar, R., 2024, May. Machine learning-based crop recommendation system. In 2024 Intelligent Systems and Machine Learning Conference (pp. 636-641).

[4] Shaikh, Tawseef Ayoub, Tabasum Rasool, and Faisal Rasheed Lone. "Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming." *Computers and Electronics in Agriculture* 198 (2022): 107119.

[5] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016.

[6] Friedman, Jerome H. "Greedy function approximation: a gradient boosting machine." Annals of statistics (2001): 1189-1232.

[7] Shams, Mahmoud Y., Samah A. Gamel, and Fatma M. Talaat. "Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making." Neural Computing and Applications 36.11 (2024): 5695-5714.

[8] Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." Advances in neural information processing systems 30 (2017).

[9] Kancharagunta, Kishan Babu, et al. "Crop recommendation system using machine learning and iot: A survey." International Conference On Innovative Computing And Communication: 2024.

[10] Arun, Ajage Sachin, et al. "Machine learning-based crop recommendation system." *2024 Intelligent Systems and Machine Learning Conference (ISML)*. IEEE, 2024.

[11] Siddharth S. S., "Crop Recommendation Dataset," Kaggle, 2021. Available: https://www.kaggle.com/datasets/siddharthss/crop-recommendation-dataset