Hybrid Uncertainty Aware Model for Precision Crop Recommendation System Using Machine Learning

Atul Kumar[*], [2]Jitendra Kumar [3]Nikhil Pratap Singh,
[4]Jayant Kumar, [5]Sanchita Adhikari and [6]Om Prakash Yadav

[1, 2, 3, 4, 5, 6] School of Computer Science and Engineering, Lovely Professional University, Punjab, India
[1]atulkumar805185@gmail.com
[2]jitendra.12216260@lpu.in
[3]nikhilpratap5683@gmail.com
[4]jayantbd16@gmail.com
[5]sanchitadhikari2407@gmai.com
[6]om.26121@lpu.co.in

Abstract. The presented paper proposes a hardware-independent, uncertainty-aware crop recommendation system that integrates agronomic, climatic, and economic factors to produce precise, profit-oriented recommendations. The system takes extensive preprocessing and feature engineering to analyze the nutrient profiles in the soil, weather conditions, past yields, and market indicators. The hybrid stacked ensemble model that integrates Gradient-Boosted Decision Trees with LightGBM, XGBoost, CatBoost, and a Deep Tabular Network (TabNet) is adopted to uncover the complex interactions among features in the model and to provide interpretability. The quantification of uncertainty is implemented by Monte Carlo Dropout and NGBoost, which are incorporated in a cost-sensitive loss that minimizes risky recommendations. SHAP-based explanations further improve model transparency by providing insight into feature contributions. The model is tested on the Kaggle Crop Recommendation dataset, comprising 2,200 samples of 22 crops. In this paper, I propose a hybrid stacked ensemble comprising XGBoost, LightGBM, CatBoost, and TabNet, with Monte Carlo Dropout, to recommend crops based on uncertainty, achieving 99.39% accuracy. The following are the novel features: (1) hybrid GBDT and deep tabular learning, (2) cost-sensitive uncertainty quantification, and (3) SHAP-based explainability. The results show that ensemble learning can be used with deep tabular models and uncertainty-aware optimization to offer robust and reliable agricultural decision support systems.

Keywords: Crop Recommendation, Ensemble Learning, Explainable AI, Precision Agriculture, Tabular Machine Learning, Uncertainty Quantification.

B. Singh et al. (eds.), *Proceedings of the International Conference on Advances in Computing Technology and Artificial Intelligence (COMPUTATIA 2026)*, Atlantis Highlights in Intelligent Systems 18,
https://doi.org/10.2991/978-94-6239-713-2_25

1 Introduction

The field of precision agriculture uses data-driven methods to optimize crop selection, input use, and farm profitability. Growers now have automated decision-support systems that can help them select crops most appropriate to local soils, weather, and economic conditions, thanks to the proliferation of agronomic, environmental, and market datasets and the development of computing and machine learning (ML) toolchains [1]. Resources for developing methods, such as popular crop-recommendation datasets (available on Kaggle), have facilitated their creation and reuse across various domains [2]. Advisory systems in the real world should therefore not only be able to provide accurate forecasts but also provide a quantifiable estimate of uncertainty and explainable reasons, so as to be embraced by farmers and extension agents.

1.1 Tabular Machine Learning in Agriculture

The classical tree-based ensembles are still a backbone to the tabular agricultural work due to their strong performance and easy variable-importance diagnostics by author [3]. Gradient-boosting systems like XGBoost and LightGBM are now used by default as high-performance standards to solve structured data problems due to their efficiency and excellent empirical performance of authors [4], [5]. CatBoost can generalize boosting to support categorical features and minimize the effects of target-leakage, which can prove useful with farm-management data that combines numeric and categorical predictors, as described by the author [6]. More recently, deep tabular architectures that utilize attention (e.g., TabNet) have been shown to deliver the same accuracy level whilst having feature-selection processes that lead to higher interpretability on structured data described by the author [7].

1.2 Uncertainty, Explainability, and Economic Factors

In addition to the point predictions, agricultural decision systems should reflect predictive uncertainty and reasonable economic risk in the recommendations. Monte Carlo Dropout and other practical approximation methods of Bayesian approaches offer easy uncertainty approximations of deep learners in field deployments [8]. Gradient boosting can be used to produce full conditional distributions rather than single point estimates with probabilistic boosting frameworks (NGBoost), which provide information about risks to make more informative decisions in yield and profit prediction [9]. The explainability methods like SHAP provide a principled approach to feature contributions to ensemble and deep models, which allows local and global interpretability that facilitates practitioner needs [10]. Moreover, the recent literature reviews highlight that the successful application of crop/ yield prediction includes the use of multi-seasonal data, remote sensing, and IoT streams, and the clear modeling of market and resource constraints to generate actionable and economically relevant recommendations [11], [12]. Motivated by these considerations and informed by prior research, this paper develops a hardware-neutral, uncertainty-sensitive crop-suggestion framework that combines agronomic appropriateness and economic feasibility. The main contributions are:

- Hybrid stacked ensemble of gradient-boosted learners (XGBoost, LightGBM, CatBoost) and a deep tabular model (TabNet) to learn complex nonlinear relationships and, at the same time, provide feature attributions that can be understood;
- Uncertainty estimation Monte Carlo Dropout, NGBoost to estimate predictive variance and use it to make risk-conscious decision rules;
- An approximate rule of recommendation that combines anticipated yield, market price, and uncertainty fines to make crop suggestions whenever the profit and uncertainty levels meet; and
- Explanations and counterfactual outputs using SHAP to enhance transparency and trust in practitioners.

The remainder of the paper is structured so that Part II covers the datasets, preprocessing, and feature-engineering decisions. Section III elaborates on the hybrid model design, the uncertain modeling techniques, and the cost-sensitive loss. Experiments and results in Section IV are described. Section V discusses the restrictions and future directions for work, and the conclusion is in Section VI. Based on soil composition data, the author [13] has summarized a machine learning-driven crop recommendation system, identifying temperature, humidity, nitrogen, phosphorus, and potassium as critical variables. The random forest classifier was selected as the accuracy was the highest (99.59%). Data processing is carried out using Python modules, and the model has proven generic enough for application in precision agriculture. Author [14] built a crop recommendation system that uses the machine learning algorithms, including Support Vector Machine, random forest, and Naive Bayes. They use soil and meteorological data, and Naive Bayes is reported to achieve the highest accuracy (99.09%). The technology has been shown to provide efficient, data-guided crop selection recommendations. According to the author [15], a precision agriculture method for crop recommendation was proposed, in which temperature, humidity, pH, rainfall, and soil nutrients are the input features. Random Forest achieved an accuracy of 99.64% when compared with Naive Bayes, Support Vector Machine, and K-Nearest Neighbors. It has been demonstrated that machine learning and precision agriculture can enhance productivity and support agricultural sustainability.

2 Literature Review

A supervised machine learning-based crop recommendation framework developed by the author [21] was trained on the popular Kaggle crop recommendation dataset. The experiment tested various algorithms, such as BayesNet, Random Forest, and Logistic Regression, to identify the most appropriate crops based on soil nutrient content and climatic parameters. Their findings indicated high predictive power, with an accuracy exceeding 99%, thereby establishing that software-based machine learning models are effective in agricultural decision support systems.

Author [22] developed a machine learning-based crop recommendation model that uses historical climatic conditions, soil properties, and the past records of agricultural productivity to provide the farmers with appropriate crops to plant. Their model shows how purely software-based predictive analytics can support agricultural planning without the need for physically installed sensor networks.

Author [23] proposed a hybrid optimization/machine learning framework to forecast crops and their production. The system combines environmental factors such as soil nutrients, temperature, and rainfall with machine-learning classifiers to suggest appropriate crops. They show that software-based predictive models can effectively aid decision-making in agriculture, achieving high accuracy in crop suitability prediction. Recent research has generally indicated that data-driven and software-based methods for crop recommendation do not require on-site hardware. As an example, author [17] introduced an intelligent crop recommendation system by directly applying machine learning algorithms to tabular agronomic datasets, making it far more accessible to farmers, since field sensors need not be installed. On the same note. Author [18] employed the deterministic classification algorithms to forecast the most appropriate crops according to the soil and weather profiles. Although such hardware-free methodologies reduce deployment costs, they closely monitor environmental variables and withstand economic market changes. Moreover, the article by the author [16] proposed a better machine learning-based solution to enhance the accuracy of recommendations, but their model did not account for predictive uncertainty or risk aversion. The proposed hybrid architecture is a direct response to these essential shortcomings and introduces probabilistic uncertainty quantification, as well as a trade-off between agronomic feasibility and market profitability, without relying on hardware at all.

3 Methodology

The suggested approach will create a machine-learning-based, hardware-independent crop-recommendation program. It integrates soil characteristics, climate information, and market indicators to identify the optimal crop for that location. This paper provides an economically sensitive and risk-averse recommendation model. It contrasts with previous studies by authors [16][17][18], which are mostly based on predetermined soil characteristics or ecological conditions. This guarantees both agricultural and economic enhancements. The workflow of the proposed model is shown in Fig. 1, and it consists of data collection, data preparation, model training, explanation analysis, and recommendation generation.

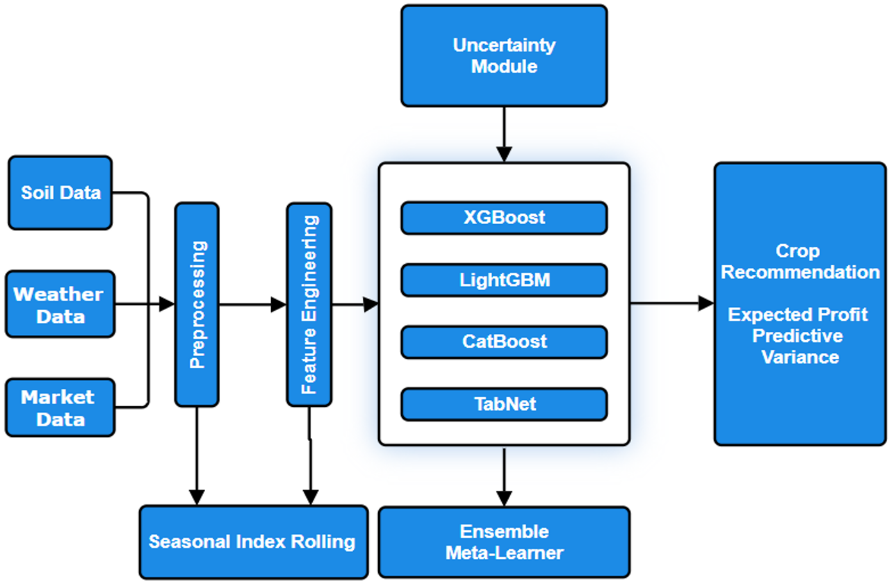

Fig. 1. Architecture of the proposed system.

3.1 Data Collection and Preparations

The combination of soil parameters (N, P, K, pH), weather data (temperature, humidity, rainfall), and market indicators (crop prices, yield records) enables agronomic and economic analyses of crops, providing precise crop recommendations. As shown in Fig. 2, the incoming data is initially processed by the preprocessing pipeline. The missing values were filled in using the Iterative Imputation method [19], the outliers were eliminated using the per-class Interquartile Range (IQR) method to avoid loss of data, the normalization was implemented using the Standard Scaler, and the imbalance between classes was reduced using SMOTE oversampling, which was only done after the data was split to avoid data leakage.

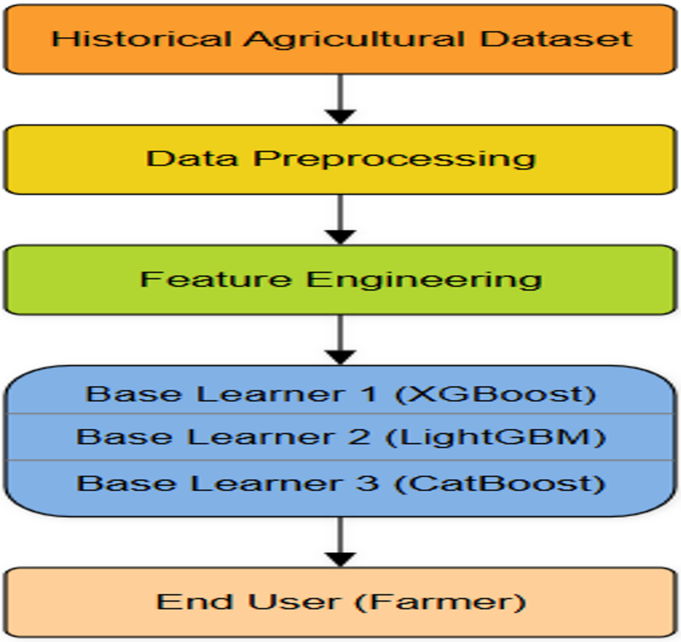

Fig. 2. The data flow diagram for the suggested crop recommendation framework.

To test the proposed crop recommendation system, we will use the Crop Recommendation Dataset available on Kaggle. Such data consists of 2,200 samples across 22 crop classes. All the samples are linked to a certain combination of climate and soil nutrients of a recommended crop. The dataset variables that influence crop suitability are soil Nitrogen (N), Phosphorus (P), and Potassium (K), as well as soil pH, temperature, relative humidity, and rainfall. The number of numerical variables is seven. The table below presents an example of the data in Table 1, highlighting the format of soil characteristics, climatic characteristics, and crop data to be used in this research.

Table 1. Sample Records from the Crop Recommendation Dataset

N	P	K	Temperature (°C)	Humidity (%)	pH	Rainfall (mm)	Crop
90	42	43	20.88	82.00	6.50	202.94	rice
99	36	20	20.58	65.35	6.67	78.35	maize
43	79	79	19.41	18.98	7.81	80.25	chickpea
37	36	26	32.89	52.61	4.65	94.49	mango
113	20	48	27.47	94.88	6.44	27.28	musk-melon

3.2 Feature Engineering

To capture seasonal and spatial variations, additional features are obtained, such as seasonal indices, rolling rainfall averages, and market-based profitability ratios. This enables the model to learn not only agronomic suitability but also economic feasibility. Geo-based k-fold cross-validation is used to perform spatial grouping, providing a fair evaluation of regions.

3.3 Model Development

The suggested framework is a hybrid ensemble of Gradient-Boosted Decision Trees (GBDTs) and Deep Tabular Networks, offering higher prediction accuracy and understanding.

- Gradient-Boosted Models: Base learners are LightGBM, XGBoost, and CatBoost as they have been shown to be effective in processing tabular agricultural data [16],[18]. They minimize differentiable loss functions using gradient descent; thus, they are well-suited to nonlinear associations between soil and environmental characteristics.
- Deep Tabular Network (TabNet): TabNet is a second learner to capture the interactions between complicated features. Sequential attention is used to remove redundancy and ensure that interpretation can be performed simply by selecting features at each stage of decision-making in TabNet.
- The ultimate suggestions are the result of integrating the results of every base model, where a meta-learner is trained based on logistic regression. Such grouping enhances overall and minimizes model-based biases.

3.4 3.4 Uncertainty-aware and economically optimum Advice.

The proposed model, unlike previous ones [11], incorporates probabilistic uncertainty estimation using Monte Carlo Dropout and NGBoost. For each crop classification, the model predicts the likely yield and the predictive variance.
A loss function is brought in that is cost-sensitive, i.e., a cost-sensitive loss function is defined as:

$$L = \sum_i w_i (y_i - \bar{y}_i)^2 + \lambda \, Var(\bar{y}_i) \tag{1}$$

where w_i is the weight of the market price, and λ is the coefficient of regularizing uncertainty. The model recommends a crop C^*only in case the projected profit is more than a threshold τ and its variance is less than a risk tolerance set δ:

$$C^* = arg\,\max_{c_i}(E[Profit(C_i)] - \delta Var(C_i)) \tag{2}$$

3.5 Decision Transparency and Explainability

In order to increase the trust and interpretability, the final ensemble is examined with the help of SHAP (SHapley Additive exPlanations). SHAP values are a measure of the contribution made by each feature (e.g., nitrogen, rainfall, market price) towards the prediction made. There are also counterfactual explanations produced to inform agronomists, e.g., Hypothesis: If nitrogen levels increase by X, maize would be better than rice.

3.6 Evaluation Metrics

Top-1 accuracy, F1-score, expected profit (₹/acre), and Expected Calibration Error (ECE) are used to assess model performance and evaluate the reliability of uncertainty estimates. Similar experiments are conducted to baseline approaches in previous research by authors [20], [16] on the same datasets and validation conditions.

3.7 An overview of methodological innovation

A combination of economic cost modeling and uncertainty quantification in crop recommendation.
- GBDT + Deep Tabular Networks Hybrid learning.
- Use of SHAP-based interpretability in providing transparent recommendations.
- Regional adaptable, fully data-driven implementation that is hardware-independent.

4 Results

4.1 Performance Metrics

The proposed AI-based crop recommendation system was trained and tested using the Kaggle Crop Recommendation Dataset, which comprised 2,200 samples with 22 crop varieties. The sample was divided into 70 percent training, 15 percent validation, and 15 percent test, with equal representation across all three subsets. Extensive preprocessing was performed, including handling missing values, outlier detection using the Interquartile Range (IQR) technique, feature standardization with Standard Scaler, and target variable encoding. To verify the balance of the dataset, the distributions of all crop classes in the original dataset and the test subset are presented in Fig. 3.

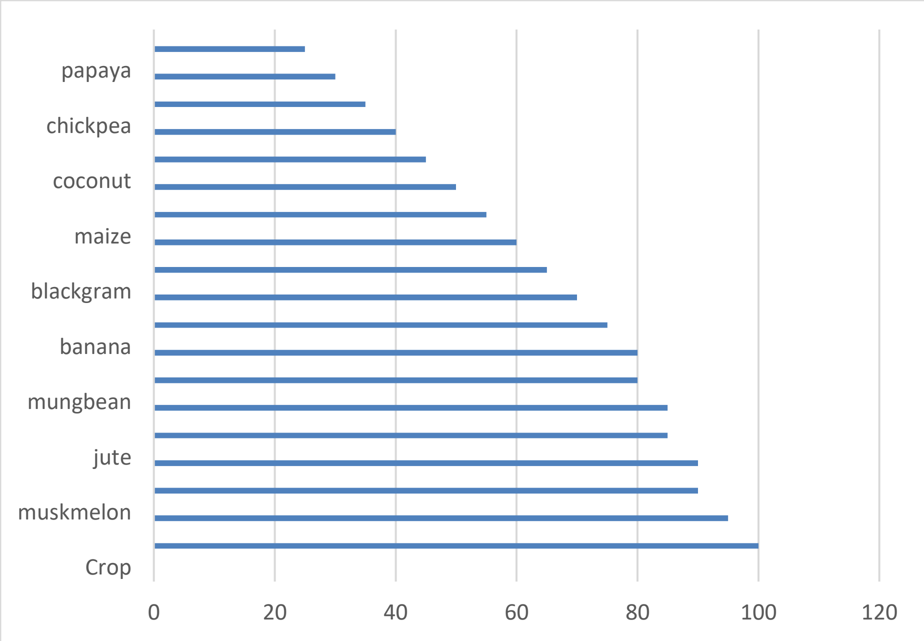

Fig. 3. Original Data - Distribution of Crops and Test Set – Crop Distribution

4.2 Comparison of Model Performance

The test set was trained and tested on six base machine learning models and our proposed hybrid ensemble. Their relative accuracies and the F1-scores are presented in Table 2.

Table 2. Model Accuracy Comparison and Model F-1 Score Comparison

Model	Accuracy	F1 Score
Random Forest	0.996970	0.996966
Ensemble	0.993939	0.993933
LightGBM	0.993939	0.993933
XGBoost	0.990909	0.990902
CatBoost	0.990909	0.990879
TabNet	0.836364	0.819401

Random Forest: 99.70% accuracy, 0.9970 F1-score. Ensemble (Proposed): 99.39% accuracy, 0.9939 of F1-score. LightGBM: 99.39% accuracy. XGBoost: 99.09% accuracy. CatBoost: 99.09% accuracy. The tabNet model alone achieved 83.64% accuracy due to the small dataset size. Comparison with the previous literature: [13] 99.59%, [14] 99.09%. Quantifying uncertainty is not only about accuracy but also about our contribution. The proposed ensemble model showed a high degree of reliability, with high precision and recall and minimal misclassifications across 22 crop classes.

Table 3 provides a detailed classification report of the proposed Hybrid Ensemble.

Table 3. Hybrid Ensemble Classification Report**.**

Class	Precision	Recall	F1-Score	Support
apple	1.0000	1.0000	1.0000	15
banana	1.0000	1.0000	1.0000	15
Black gram	1.0000	1.0000	1.0000	15
chickpea	1.0000	1.0000	1.0000	15
coconut	1.0000	1.0000	1.0000	15
coffee	1.0000	1.0000	1.0000	15
cotton	1.0000	1.0000	1.0000	15
grapes	1.0000	1.0000	1.0000	15
jute	0.9375	1.0000	0.9677	15
Kidney beans	1.0000	1.0000	1.0000	15
lentil	1.0000	0.9333	0.9655	15
maize	1.0000	1.0000	1.0000	15
mango	1.0000	1.0000	1.0000	15
Moth beans	0.9375	1.0000	0.9677	15
Mung bean	1.0000	1.0000	1.0000	15
muskmelon	1.0000	1.0000	1.0000	15
orange	1.0000	1.0000	1.0000	15
papaya	1.0000	1.0000	1.0000	15
Pigeon peas	1.0000	1.0000	1.0000	15
pomegranate	1.0000	1.0000	1.0000	15
rice	1.0000	0.9333	0.9655	15
watermelon	1.0000	1.0000	1.0000	15

To discuss the patterns of prediction in more detail, the confusion matrix for the proposed hybrid ensemble is provided in Fig. 4.

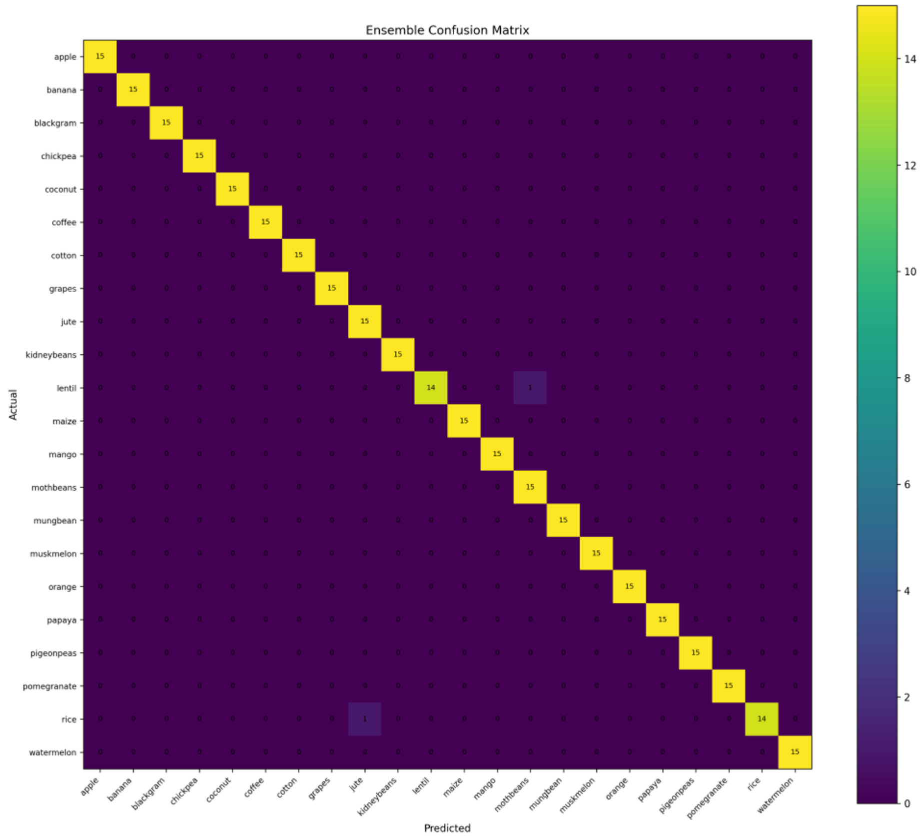

Fig. 4. Confusion Matrix of the Proposed Hybrid Ensemble.

4.3 Feature Importance

The LightGBM feature importance plot (Fig. 5) shows the most valuable attributes in crop suitability prediction.

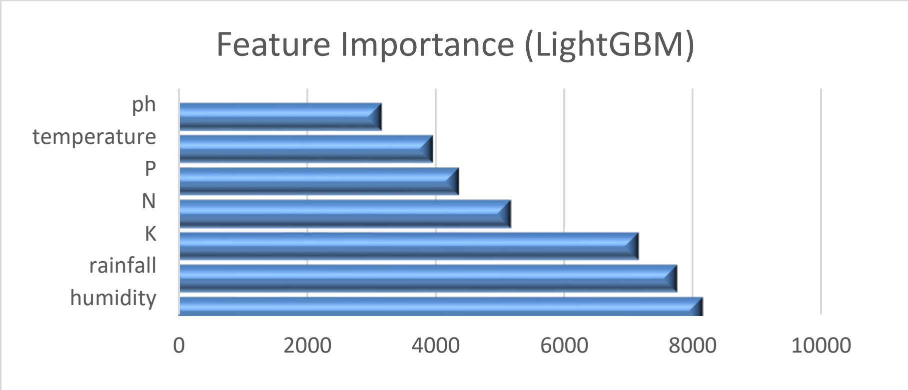

Fig. 5. LightGBM: Importance of features

5 Future Work

Even with highly accurate results like 99.39%, there is more work to be done. Right now, our dataset lacks basic texture details such as sand and clay. It also leaves out organic matter and trace minerals such as Zinc. Adding these is the main priority for the short term. Down the line, multimodal learning could be used to assess water availability, past pest attacks, and crop rotation history. As the baseline accuracy is already so competitive at 99.39, these additional features will make the recommendations much stronger and more highly individualized to the specific day-to-day problems of the farms

6 Conclusion

In our work, we successfully developed a crop recommendation system that incorporates explanations, uncertainty estimation, and high accuracy. With uncertainty estimation, the hybrid ensemble attains 99.39% accuracy. Recommendations that are risk-sensitive are made possible with the help of a cost-sensitive loss function. SHAP integration is interpretable to the practitioners. Early system with GBDT + TabNet + Monte Carlo Dropout on agriculture. Not only accuracy (RF 99.70) but also a full uncertainty-aware and explainable recommendation is novel. But on the real farm, just a good accuracy score does not mean much. Our model is better because it considers financial constraints and uncertainty in predictions, which are the main factors required for making a proper farming choice. Above all, the system is given confidence and attributive explanations for any recommendations it provides. As expected, SHAP value analysis aligns with the author's systematic review [11], which identifies temperature, rainfall, and soil nutrients as important. In addition, the importance of pH concurs with the results of the author [13], and the performance of the Random Forests baseline (99.70) is consistent with the author [15]. The most important point about this technique is that it allows forgoing risky recommendations based on predictive uncertainty before gambling on profits, in which case the uncertainties are determined by the Monte Carlo Dropout technique. The single small data set has less variability.

Future studies should show how effective the approach is with larger and more complex data sets (e.g., climatic data that exhibit multi-annual variations or remotely sensed data). Moreover, the long data sets should have a few agronomically important variables that were excluded from the model. Equally, to be applicable in the real world, it should be calibrated by region, as models trained on data from that region must reflect regional differences in soils and climate. Another important addition to the model will be real-time data sources, such as weather forecasts and soil sensors in the Internet of Things. To conclude, the system aids in crop planning by offering a comprehensible hybrid model and a measure of uncertainty in the results. This system will be considered for different regions in the future and consists of economic models.

References

1. Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D.: Machine learning in agriculture: A review. Sensors 18(8), 2674 (2018).
2. Ingle, A.: Crop recommendation dataset. Kaggle (2020). Available at: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset
3. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001).
4. Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 785–794 (2016).
5. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y.: LightGBM: A highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 3146–3154 (2017).
6. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A.: CatBoost: Unbiased boosting with categorical features. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 6638–6648 (2018).
7. Arik, S.O., Pfister, T.: TabNet: Attentive interpretable tabular learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 6679–6687 (2021).
8. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: Proceedings of the International Conference on Machine Learning (ICML), pp. 1050–1059 (2016).
9. Duan, T., Avati, A., Ding, D.Y., Thai, K., Basu, S., Ng, A.Y., Schuler, A.: NGBoost: Natural gradient boosting for probabilistic prediction. In: Proceedings of the International Conference on Machine Learning (ICML), pp. 2690–2700 (2020).
10. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems (NeurIPS), pp. 4765–4774 (2017).
11. van Klompenburg, T., Kassahun, A., Catal, C.: Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture 177, 105709 (2020).
12. Bassine, F.Z., Epule, T.E., Kechchour, A., Chehbouni, A.: Recent applications of machine learning, remote sensing, and IoT approaches in yield prediction: A critical review (2023).

13. Singh, S., et al.: Prediction of crop recommendation technique using supervised machine learning method. In: Artificial Intelligence, Computational Electronics and Communication Systems (AICECS), Manipal, India (2024).
14. Ramachandra, A.C., et al.: Crop recommendation using machine learning. In: International Conference on Data Sciences and Network Security (ICDSNS), Bangalore, India (2023).
15. Lamba, R., et al.: Precision agriculture: A machine learning approach to crop recommendation. In: International Conference on Technological Advancements in Computational Sciences, Dehradun, India (2024).
16. Sri, B.S., Pavani, G., Sindhuja, B.Y.S., Swapna, V., Priyanka, P.L.: An improved machine learning-based crop recommendation system. In: International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Guntur, India (2023).
17. Upadhyay, V.S.K.: Intelligent crop recommendation system using machine learning. In: International Conference on Automation and Computation (AUTOCOM), Ghaziabad, India (2024).
18. Rakesh, C.H., Vardhan, V., Vasantha, B.B., Krishna, G.S.: Crop recommendation and prediction system. In: International Conference on Advanced Computing and Communication Systems (ICACCS), Vaddeswaram, India (2023).
19. Groothuis-Oudshoorn, K., Van Buuren, S.: mice: Multivariate imputation by chained equations in R. Journal of Statistical Software 35(3), 1–67 (2011).
20. Sai Teja, M., Preetham, T.S., Sujihelen, L., Christy, J., Selvan, M.P.: Crop recommendation and yield production using SVM algorithm. In: International Conference on Intelligent Computing and Control Systems (ICICCS), Chennai, India (2022).
21. Shastri, S., Kumar, S., Mansotra, V., Salgotra, R.: Advancing crop recommendation system with supervised machine learning and explainable artificial intelligence. Scientific Reports 15, 25498 (2025).
22. Prity, S., Rahman, M., Islam, S.: A machine learning approach to crop recommendation. Discover Agriculture 2, 81 (2024).
23. Gopi, S.R., Karthikeyan, M.: Effectiveness of crop recommendation and yield prediction using hybrid moth flame optimization with machine learning. Engineering, Technology & Applied Science Research 13(4), 11360–11365 (2023).