Grenze International Journal of Engineering and Technology, January Issue

Crop Recommendation System for Konkan Region of Maharashtra using Interpretable Machine-Learning with XAI

Janhavi Vadke[1] and Sachin Bhoite[2]
[1-2]Vishwanath Karad MIT-World Peace University Department of Computer Science and Applications, Pune, India
Email: janhavi.vadke@mitwpu.edu.in, sachin.bhoite@mitwpu.edu.in

***Abstract*— Machine Learning has proved beneficial in many domains positively impacting various aspects of our daily lives. Agriculture in India is a major occupation and nations economy is deeply affected by agriculture. This study suggests a Crop Recommendation System for the Konkan region of Maharashtra utilizing Machine Learning combined with Explainable Artificial Intelligence (XAI) in light of growing agricultural uncertainties and the urgent need for sustainable farming techniques. The Konkan region, a crucial agricultural zone in Maharashtra, is the direction of the dataset, which includes important agronomic metrics such as soil nutrients (nitrogen, potassium and phosphorus), pH level, temperature, humidity, and rainfall. The prediction performance of a number of supervised machine learning models, such as Random Forest, KNN, Support Vector Machine (SVM), Decision Tree and XGBoost, in suggesting the best crops based on these characteristics was assessed. SHAP (SHapley Additive exPlanations) values were used to illustrate feature contributions and model decisions in or-der to improve interpretability and confidence among local farmers and policymakers. LIME is used to explain the reason for suggestion of a specific crop for a given land. In addition to promoting openness, the incorporation of XAI enables stakeholders to make knowledgeable agronomic choices. By enhancing crop output, soil health, and financial returns, this technology can make a substantial contribution to precision agriculture in Maharashtra.**

***Index Terms*— Crop Recommendation, farming, agronomic metrics, classification, machine learning models, shap, lime, Explainable Artificial Intelligence, XAI.**

I. Introduction

Agriculture continues to be a vital component of the Indian economy with a sizable section of the people depending on it for their living. One of the biggest states in India, Maharashtra, has a varied agroclimatic environment that presents potential and difficulties for the development of sustainable crops. Because of its distinctive topography, diverse soil types, and heavy rain-fall, the Konkan region in particular need detailed and region-specific agricultural planning. In this area, traditional crop selection techniques frequently rely on farmers' instincts and knowledge, which might not always coincide with ideal agronomic and meteorological circum-stances. Suboptimal yields, soil deterioration, and financial losses may result from this.
New directions in precision agriculture have been made possible by recent developments in machine learning (ML), which allow for decision-making based on data that boosts crop productivity and resource efficiency.

Grenze ID: 01.GIJET.12.1.194

In order to recommend the best crops for a particular area, machine learning algorithms can examine intricate, non-linear correlations between a variety of variables, including soil nutrients, weather patterns, and environmental conditions. The "black box" character of ML models, however, is a significant obstacle to their use in agri-culture since it reduces end-user confidence and transparency, particularly for farmers and ag-ricultural officers.

This study fills this gap by developing a Crop Recommendation System specifically for Maha-rashtra's Konkan region by combining Explainable Artificial Intelligence (XAI) methods with conventional machine learning models. We hope to improve the interpretability of model pre-dictions by using SHAP (SHapley Additive exPlanations), which will foster trust and facilitate better decision-making. In order to train and assess multiple ML classifiers, this study uses a region-specific dataset that includes characteristics like temperature, humidity, pH, rainfall, potassium (K), phosphorus (P) and nitrogen (N). Finding the model that offers clear and useful insights in addition to being the most accurate is the aim. Traditional agricultural methods in Maharashtra might be completely transformed by combining XAI with crop recommendation, which would maximize resource usage, encourage sustainability, and boost farmer profits. By presenting a scalable, interpretable, and data-driven method for crop planning in a climate-sensitive area, this work advances the expanding subject of smart agriculture

II. Problem Domain

Crop recommendation using predictive modelling and explainable artificial intelligence (XAI) in the context of regional agriculture in the Konkan region of Maharashtra. It can help farmers make decision for agriculture by recommending suitable crop for successful cultivation.

III. Problem Definition

Crop selection is difficult in Maharashtra's Konkan region because of the region's diverse soil composition, erratic climate, and dearth of regional decision-support tools. Conventional crop recommendation techniques mostly rely on general recommendations and farmer experience, which frequently ignore the intricate relationships between soil nutrients, rainfall, tempera-ture, and other environmental variables. By evaluating big datasets, machine learning pro-vides a reliable answer by recommending the best crops according to local characteristics. However, many ML models' "black box" character can limit their usability and credibility, par-ticularly among farmers and local agricultural officials. Explainable AI (XAI) tools like as SHAP or LIME enables more accurate, transparent, and trustworthy support for sustainable ag-riculture in the Konkan region.

IV. Related Work or Literature Studies

Ref. [1] created a crop recommendation model Using Random Forest that performed well in Tamil Nadu's agricultural zones. Even while the models performed well in terms of prediction, they frequently lacked openness, which is essential for acceptance in farming communities that are semiliterate and rural.

Ref. [2] Multiple machine learning models, including as Naive Bayes, Support Vector Ma-chine, Random Forest, AdaBoost, Logistic Regression, Decision Tree, XGBoost, Bagging and Gradient Boost, were investigated in the study's initial phase. After that, each model's perfor-mance was evaluated. Accuracy is a frequently used metric to evaluate the performance of a classification model. It can be observed that random forest outperformed the others in terms of accuracy, achieving about 91.88%.

Ref. [3] The model has been examined by making use of specific styles of learning system and set of rules. Bagged tree and KNN indicates exact accuracy but amongst all the classifiers, SVM performed better than other classifiers, achieving the highest accuracy and the shortest processing time. By suggesting appropriate crops based on soil, environmental, and financial considerations, the system increases output while using less fertilizer.

Ref. [4] The provided Kaggle datasets, along with rainfall and actual meteorological data, are used to test the predictive accuracy of the machine learning system (Random Forest). A crop recommendation system that uses Random Forest is a helpful method for providing farmers and stakeholders with data-driven guidance on the best crops for specific environmental con-ditions. Although it takes longer to execute than Decision Tree, the random forest approach has a higher accuracy than Naïve bias, SVM, Decision Tree, and Logistic Regression.

Ref. [5] Demonstrated better performance when using Naive Bayes and Support Vector Ma-chines (SVM) to select crops based on soil nutrients and environmental data as compared to expert-based advisory systems.

Ref. [6] A number of popular techniques, such as logistic regression, SVM, random forests, na-ïve bayes, and decision trees. The idea of supervised learning is the basis of each and every algorithm. It will boost output for

every method, with SVM and random forest producing more efficiently than other models. For our dataset, we find that the Random Forest algorithm yields the best level of precision. Based on weather data such as temperature, humidity, rainfall, and sun hours, the Random Forest algorithm was used to predict the yield.
Ref. [7] The yield prediction model that was created by utilizing machine learning techniques offers reliable crop suggestions and direction for the best fertilizer application, with Random Forest attaining a peak accuracy of 95%. This user-friendly technology gives farmers data-driven insights to increase yields and ease financial strains, eventually promoting the agricul-tural sector's sustainability.
Ref. [8] Natural elements including N, P, K, temperature, humidity, pH, rainfall, and so forth were taken into consideration when using a two-fold model, also known as the data balancing & classification model, which consists of a family of machine learning algorithms over the balanced data set. This study confirms the possible effectiveness of using machine learning (ML) algorithms in a decision-making system that employs an intelligent crop recommenda-tion system to suggest suitable crops. Boosting (Cboost), one of the 14 classifiers and six per-formance metrics used by the ML algorithm family, yields the best outcomes with an accuracy of 99.15.
Ref. [9] Created XAI-CROP, a crop recommendation system that makes use of eXplainable ar-tificial intelligence and machine learning methodologies. We showed the better performance of XAI-CROP over other machine learning models through training and evaluation on a da-taset for crop cultivation in India. The lower Mean Squared Error (MSE), Mean Absolute Error (MAE) as well as higher R-squared (R2) value in the data demonstrated its higher accuracy.
Ref. [10] The system provides precise crop recommendations based on a range of variables, such as soil quality, climate, and environmental influences, by utilizing advanced machine learning algorithms and data-driven assessments. The usage of Explainable AI (XAI) ensures that the decision-making process is transparent, allowing users to understand the rationale be-hind the suggestions.

V. Research Gaps and Challenges

This study seeks to close the gap by creating a crop recommendation system tailored for the Konkan region of Maharashtra using ML models supplemented with SHAP-based XAI for im-proved interpretability. The project aims to increase technology and practical applicability in precision agriculture by fusing predicted accuracy with model transparency.

VI. Research Methodology and Analysis

A. Research Workflow

The research workflow for this study was designed to systematically build and evaluate a re-gion-specific, interpretable crop recommendation system. The methodology derived from Ref. [11] is a structured ML pipeline that includes data preprocessing, model selection based on prior accuracy benchmarks, and training with robust evaluation. While originally designed for image-based garbage categorization, this method can be used for classification tasks in crop recommendation. The structured pipeline allows for data-driven optimization, and by combining SHAP and LIME, the model outputs become interpretable, ensuring transparency in identifying significant water quality drivers.
The steps involved are as follows:

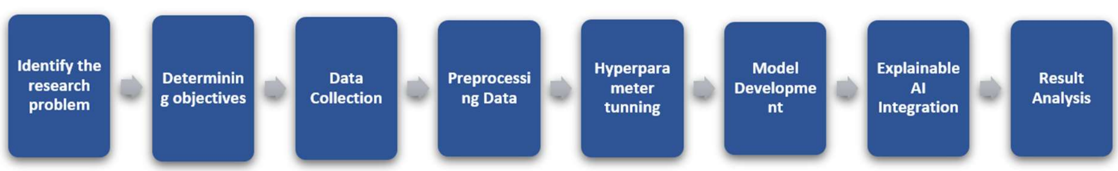

Figure 1. Research Workflow

B. Identify research problem and objective

The majority of research has been on regionally generic or pan-Indian datasets. There aren't many detailed, region-specific models that take into account the distinct agroclimatic traits of particular areas, like the Konkan belt of Maharashtra. This area need a targeted approach to crop planning because of its lateritic soils, significant rainfall, and coastal humidity. This study fills a gap in the majority of generic agricultural models by introducing a region-specific crop recommendation system designed to the particular agro-climatic conditions of Maharashtra's Konkan area. To enhance the interpretability and transparency of machine learning predic-tions, this work includes Explainable AI (XAI) methodologies, notably SHAP (SHapley Addi-tive Explanations), in contrast to traditional systems. To improve accuracy and relevance, the model makes use of local soil and climate data.

C. Data Collection

- A dataset "Crop_recommendation_konkan_maharashtra" specific to the Konkan re-gion of Maharashtra was used, containing key agricultural features such as nitrogen, phosphorus, potassium levels, pH, temperature, humidity, and rainfall.
- It contains more than 111214 records.
- The Kaggle is the dataset's source.

Data fields are
N stands for ratio of Nitrogen content in soil
P stands for ratio of Phosphorous content in soil
K stands for ratio of Potassium content in soil
Temperature: gives the temperature in degree Celsius
Humidity: gives the relative humidity in percentage
ph stands for ph value of the soil
Rainfall: gives the overall rainfall in millimeters

D. Data Preprocessing

In order to prepare raw data for machine learning models, data preprocessing is a crucial step. The data needs to be appropriately organized, cleansed, and transformed. This dataset was proper without any missing values.

E. Hyperparameter tunning

Hyperparameter tuning is the step for finding the best combination of hyperparameters for a machine learning model to improve its performance. Hyperparameter tuning is essential for improving model accuracy, training efficiency, and generalizability.

F. Model Development

Multiple machine learning models (Support Vector Machine, Random Forest, KNN, Decision Trees and XGBOOST) were developed and tested. Performance was evaluated using metrics such as accuracy and F1-score.

G. Explainable AI Integration

SHAP (SHapley Additive Explanations) was implemented to interpret the model predictions, providing insights into how each feature contributed to the recommended crop in Konkan re-gion of Maharashtra.

H. Result Analysis

Model outputs and SHAP visualizations were analysed to evaluate both predictive perfor-mance and interpretability.

VII. IMPLEMENTED METHODOLOGIES

Firstly, dataset is imported, and the first 5 rows are as follows

TABLE I: DATASET SAMPLE OF FIRST 5 ROWS

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

Feature Correlation Heatmap: Features Correlation heatmap is a visual aid used in machine learning and data analysis that illustrates the connections between various variables or fea-tures in a dataset. Quickly determining which variables are highly correlated with one another will help you choose features, cut down on redundancy, and enhance model performance.
The correlation coefficient, usually the Pearson correlation, which quantifies the linear rela-tionship between two variables, sits at the center of the heatmap. The range of values is -1 to +1:
A perfect positive correlation is shown by a value of +1, meaning that as one increases, the other does too. 0 denotes no linear association. A perfect negative correlation is denoted by a value of -1, meaning that when one increases, the other decreases.

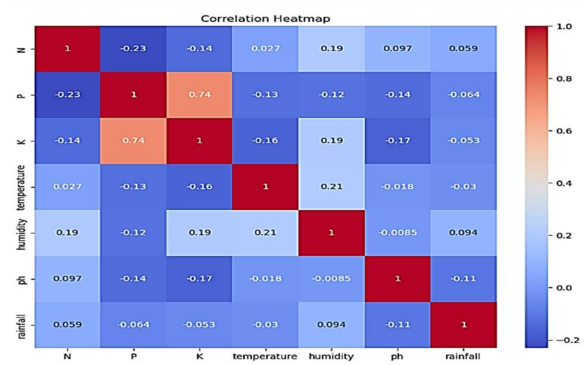

Figure 2. Feature Correlation Heatmap

Further dataset undergoes modelling process where multiple mentioned models are imple-mented and evaluated. Hyperparameter grids are applied. Following results were obtained.

TABLE II. RESULT OF DIFFERENT MODELS

	Model	Accuracy	F1 Score
0	SVM	0.988636	0.988595
1	KNN	0.981818	0.981690
2	Decision Tree	0.979545	0.979423
3	Random Forest	0.993182	0.993178
4	XGBoost	0.990909	0.990843

Best Model identified is Random Forest: Random Forest is a powerful, flexible, and widely used ensemble machine learning algorithm used for classification and regression tasks. This method reduces the risk of overfitting, increases accuracy, and handles large datasets with higher dimensionality well.

```
Best model: Random Forest

              precision    recall  f1-score   support

       apple       1.00      1.00      1.00        20
      banana       1.00      1.00      1.00        20
   blackgram       1.00      0.95      0.97        20
    chickpea       1.00      1.00      1.00        20
     coconut       1.00      1.00      1.00        20
      coffee       1.00      1.00      1.00        20
      cotton       1.00      1.00      1.00        20
      grapes       1.00      1.00      1.00        20
        jute       0.95      1.00      0.98        20
 kidneybeans       1.00      1.00      1.00        20
      lentil       1.00      0.95      0.97        20
       maize       0.95      1.00      0.98        20
       mango       1.00      1.00      1.00        20
   mothbeans       0.95      1.00      0.98        20
    mungbean       1.00      1.00      1.00        20
   muskmelon       1.00      1.00      1.00        20
      orange       1.00      1.00      1.00        20
      papaya       1.00      1.00      1.00        20
  pigeonpeas       1.00      1.00      1.00        20
 pomegranate       1.00      1.00      1.00        20
        rice       1.00      0.95      0.97        20
  watermelon       1.00      1.00      1.00        20

    accuracy                           0.99       440
   macro avg       0.99      0.99      0.99       440
weighted avg       0.99      0.99      0.99       440
```

Figure 3. Best Model Evaluated

Confusion Matrix: This confusion matrix is a tool to evaluate performance for classification models. It shows how well the model's predictions match the actual outputs by breaking them down into four cat-egories:
Structure of a Confusion Matrix -For binary classification, it looks like this:

TABLE III

	Positive (Predicted)	Negative (Predicted)
Positive (Actual)	True Positive (TP)	False Negative (FN)
Negative (Actual)	False Positive (FP)	True Negative (TN)

From Confusion Matrix we can calculate Accuracy, Precision, Recall (Sensitivity) and F1 Score

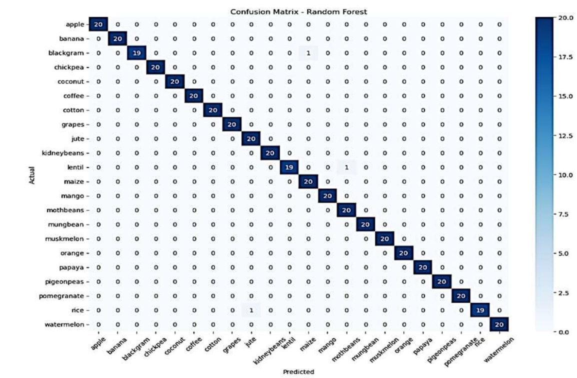

Figure 4. Confusion Matrix

SHAP value is a technique from Explainable AI (XAI) called SHAP (SHapley Additive exPlanations) aids in the explanation of how machine learning models generate predictions. It indicates the relative contribution of each input feature (such as temperature, pH, rainfall, etc.) to a particular model fore-cast.

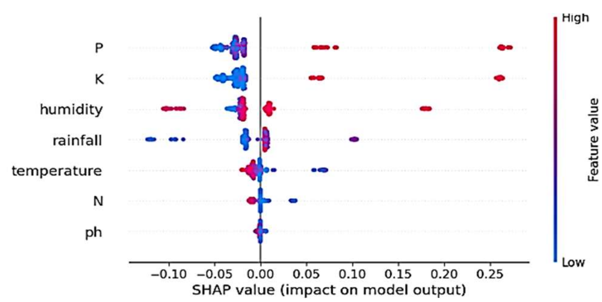

Figure 5. Bee-swarm plot for SHAP

Figure 5. Bee-swarm plot displays SHAP values of features impacting the model output, the blue plots indicate negative influence whereas the red plots indicate positive influence.

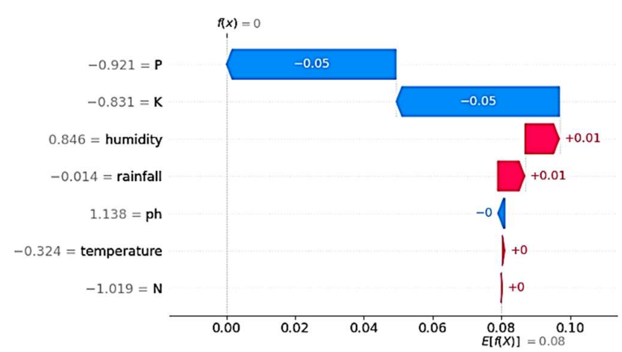

Figure 6. SHAP Water fall plot

Figure 6. Waterfall plot also displays SHAP values of features impacting the model output, the blue bars indicate negative influence whereas the red bars indicate positive influence on overall imple-mentation of the model.

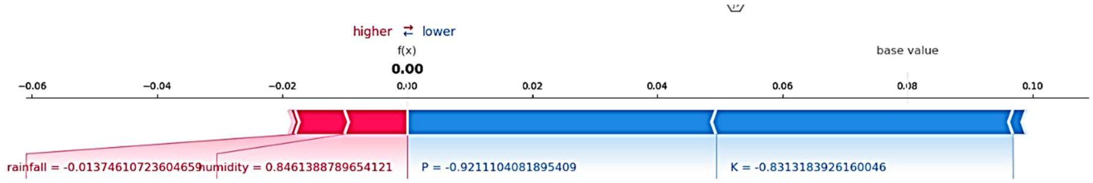

Figure 7. SHAP Force plot

LIME is the Local Interpretable Model-Agnostic Explanations. It is an Explainable AI (XAI) technique that can be used to explain the predictions of any black-box machine. learning mod-el, such as SVM, Random Forest, or XGBoost. LIME can be used to explain why a specific crop, like banana, orange, rice, cashew, or turmeric, was suggested for a given land based on soil and environmental characteristics in the Konkan region of Maharashtra's crop recommen-dation systems.

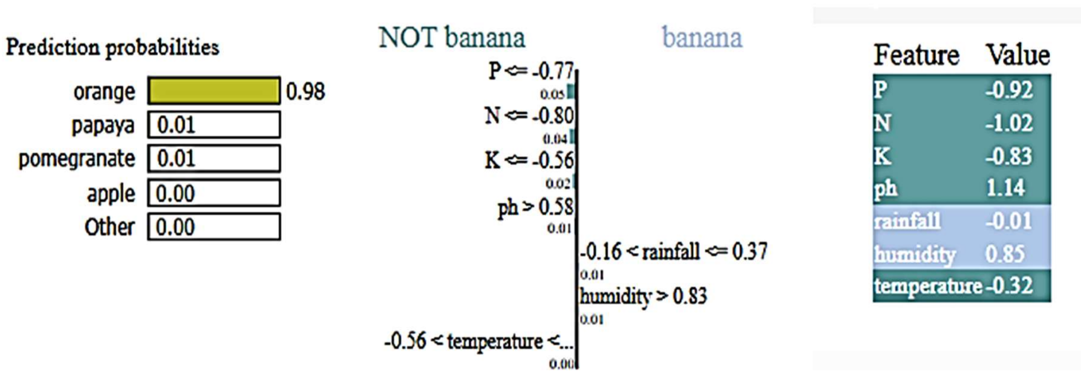

Figure 8. Lime Implementation

Figure 8. displays LIME implementation on the model for a single instance specifying features impact-ing the model performance.

VIII. Conclusion

This work uses machine learning techniques supplemented with Explainable AI (XAI) to develop a crop recommendation system tailored to the Konkan region of Maharashtra. In addition to achieving high prediction accuracy, the system offers transparency into how each environmental and soil parameter affects crop selection by using SHAP for model interpretability. Random For-est performed the best among the studied algorithms in terms of accuracy and efficiency. By ad-dressing a significant drawback of conventional black-box models, XAI promotes increased use-fulness and trust among farmers and agricultural officers. All things considered, this research helps make better, data-driven agricultural decisions and could boost regional sustainable farming methods, decrease resource waste, and increase production.

Future Work

Although the suggested crop suggestion method shows promise for the Konkan area, there is still room for improvement and growth:

- Mobile Application Deployment: The system may be made into an intuitive smartphone appli-cation that supports multiple languages, making it simple for farmers with varying levels of literacy to access it.
- Image-Based Crop and Soil Analysis: By automating soil classification and disease detection through the use of computer vision techniques, field photographs may be analysed, enhancing the diagnostic capabilities of the system.
- Real-Time Data Integration: Improvising data set based on satellite imagery, IoT sensor data, and real-time weather updates can improve the recommendation's precision and responsive-ness.

REFERENCES

[1] Pandian, S., Rajendran, P., & Ravi, M. (2019). A crop recommendation model using Random Forest for Tamil Nadu's agricultural zones. Journal of Agricultural Technology and Innovation, 8(2), 45-58.

[2] Yashashree, M., Rani, K., & Sharma, N. (2024, June). Comparative analysis of machine learning models for crop recommendation. International Journal of Agricultural Data Science, 10(3), 105-118.

[3] Soni, N., Singh, R., & Mehta, A. (2023, April). An evaluation of machine learning models for crop predic-tion based on soil and environmental factors. Agricultural Systems Research, 15(1), 72-88.

[4] Pradip Mukundrao Paithane, M., Deshmukh, P., & Joshi, V. (2023, October). A crop recommendation sys-tem using Random Forest for accurate predictions based on environmental data. Proceedings of the 6th In-ternational Conference on Agricultural Data Science, 100-112.

[5] Sharma, R., & Suryavanshi, A. (2017). Crop prediction using Naive Bayes and Support Vector Machines for nutrient-based advisory systems. International Journal of Agricultural Informatics, 4(1), 25-36.

[6] Saikat, K., & Dr. Abhoy, P. (2023, July). Supervised learning algorithms for crop yield prediction: A com-parison of SVM, Random Forest, and Naive Bayes. International Journal of Machine Learning in Agricul-ture, 12(2), 78-92.

[7] Annaji Kuth, A., Rao, S., & Verma, N. (2024, December). Yield prediction and fertilizer optimization using machine learning: A case study in crop recommendation systems. Agricultural Technology & Sustainability Journal, 18(4), 145-158.

[8] Shraban Kumar, S., Roy, M., & Patil, N. (2022, October). Intelligent crop recommendation system using a two-fold model: Data balancing and classification algorithms. International Journal of Intelligent Systems in Agriculture, 20(5), 233-249.

[9] Mahmoud Y. Shams, M., Ali, K., & Hasan, R. (2024, January). XAI-CROP: A crop recommendation system using explainable AI and machine learning. Journal of Agricultural Machine Learning, 5(1), 12-29.

[10] Divya S., & Dr. S. Lavanya. (2025, March). Crop recommendation using advanced machine learning algo-rithms and Explainable AI (XAI). Journal of Agricultural AI & Data Science, 9(2), 67-81.

[11] S. J. Buchade and S. Bhoite, “Comparative Study of ML Algorithms for Garbage Classification,” Jan. 30, 2024. doi: 10.21203/rs.3.rs-3903806/v1.