2024 27th International Conference on Computer and Information Technology (ICCIT)
20-22 December 2024, Cox's Bazar, Bangladesh

Smart Crop Recommendation System Using Machine Learning and Explainable AI

Syed Imran Ertaza[1], and SK Alamgir Hossain[2]
[1,2]Computer Science and Engineering Discipline
Khulna University, Khulna-9208, Bangladesh
Email: imranertaza12@gmail.com, alamgir@cseku.ac.bd

***Abstract*—In many countries of the world, especially in South Asian countries, most of the people's livelihood and food habits are based on agriculture, so it is very important to produce different types of crops profitably and in large quantities. However many farmers are still using primitive methods of cropping and do not understand exactly which crop is best for which land at which time due to climate variability, soil degradation, and environmental instability. The AI-based automated system is a crying need for the farmers for this problem. This study proposes a smart crop recommendation model using Machine Learning (ML) and Explainable AI to help farmers make data-driven decisions on what crops are ideal to grow. This investigation utilizes a standard crop recommendation dataset from Kaggle, which includes key soil characteristics such as nitrogen, phosphorus, potassium, and pH level; and climate parameters including temperature, humidity, and rainfall. The proposed work makes the best use of several machine learning algorithms including Support Vector Machine, Decision Tree, Random Forest, Gaussian Naive Bayes, and a boosting ensemble classifier called Adaptive Boosting (AdaBoost). Among these models, AdaBoost reveals the best performance of 99.35% with a Decision Tree oriented base learner ID3. In addition, the SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) from explainable AI are used to detect the responsible soil and climate features for the classification of the AdaBoost ensemble model.**

***Index Terms*—precision agriculture, recommendation model, machine learning, AdaBoost, explainable AI**

I. Introduction

Agriculture is crucial to the livelihoods of millions of people worldwide, especially in South Asian countries, where it drives economic growth and sustains communities [1]. Choosing the right crop is paramount in agriculture where livelihoods depend on the soil's bounty and the weather's whims [2]. However, traditional farming methods often lack the skills and finesse to navigate current challenges. Farmers, especially in South Asian countries, struggle with uncertainty related to climate variability and soil properties, resulting in poor crop selection and low productivity [3], [4]. These challenges emphasize the critical need for advanced technological solutions to assist farmers in establishing informed conclusions about crop selection, maximizing yield, and sustainability.

Integrating machine learning and artificial intelligence (AI) offers an unprecedented opportunity to revolutionize crop recommendation systems [2]. It has emerged as a powerful tool in the domain of precision agriculture with its capability to examine large amounts of historical examples. Using factors such as soil quality, weather, and past crop success, these algorithms can give farmers personalized advice, helping them make better decisions [5]. In this context, the emergence of explainable AI methodologies becomes crucial. Explainable AI techniques such as model distillation and feature importance analysis deliver crucial understanding into the recommendation procedure of machine learning algorithms [6]. By distilling complex models into simpler ones and identifying the most influential features, stakeholders gain a clear understanding of why certain recommendations are made. This transparency helps build trust and gives stakeholders the confidence to make informed decisions using the model's results.

This study aims to address the complexities of modern agriculture by providing farmers with actionable recommendations based on the comprehensive analysis of soil attributes, climate data, and historical crop performance. Our empirical study considers using the crop recommendation dataset from Kaggle [7], utilizing several machine learning algorithms as well as providing explanations of the factors for which a certain recommendation is made. Adaptive Boosting (AdaBoost), a boosting ensemble version of the decision tree algorithm [8], is utilized in this investigation. This study examines AdaBoost solely with a series of weak learners such as Support Vector Machine (SVM) or Iterative Dichotomiser 3 (ID3) or C4.5 or Classification and Regression Trees (CART) or Gaussian Naive Bayes (GNB) algorithm to ensure a thorough understanding of crop recommendations. The following is a summary of the main contributions of this study.

1) Examines the Adaboost algorithm using different base learners including the SVM, ID3, C4.5, CART, and GNB, which achieves a marvellous accuracy of 99.35%.
2) Utilizes SHAP and LIME visualizations from explainable AI to investigate which features are mostly responsible for recommending crop suggestions.

The article is constructed as follows: section II provides the pros and cons of existing literature in the field of crop recommendation systems, section III outlines the proposed research methodology, Section IV describes the analysis of the experimental results, and Section V terminates the study with insights and future directions.

II. Related Works

This section discusses the recent studies on the pros and cons of crop recommendation systems.

TABLE I: Summary of the Recent Studies for Crop Recommendation System

Name	Year	Dataset Used	No. of Crops	Used Features	Preprocessing	Used Models	Evaluation Accuracy (%)	Advantages	Disadvantages
A. Shwetha et al. [2]	2024	Not Specified	Not Specified	Crop length, type of soil, season, pH, location, temperature, soil water content, pesticide need (8)	Data cleaning, noise removal, replace missing values with mean	MLR ANN RF KNN **NB**	60.00 86.00 90.00 89.00 **95.00**	Building a web and mobile application for crop recommendation	Quality of the dataset is questionable, and the lack of interpretability of the deployed models
R. Ajoodha et al. [1]	2023	Kaggle (2200) Used (1000)	10	N, P, K, pH (4)	None	RF SVM KNN **K-Star** NB	91.10 88.40 91.00 **91.30** 90.60	Proposing region-based crop recommendation system for South Africa	Limited to soil attributes only. May not consider all factors affecting crop growth.
D. Dahiphale et al. [3]	2023	Kaggle (2200)	22	N, P, K, pH, Temperature, Humidity, Rainfall (7)	Data cleaning, removal of outliers	LR DT **RF** KNN **NB** SVM NN	95.55 98.68 **99.50** 98.04 **99.50** 98.68 97.73	Discussing about the challenges and key future ideas in the field of agriculture.	Longer training time, Lack of interpretability of the implemented models
C. Rajesh et al. [4]	2023	Kaggle (2200)	22	N, P, K, pH, Temperature, Humidity, Rainfall (7)	Data cleaning, normalization	LR DT **RF**	66.00 98.54 **99.27**	The proposed RF model is very fast (training time: 0.06s)	Data preprocessing step is ignored
M. Patel et al. [9]	2023	Kaggle (2200)	22	N, P, K, pH, Temperature, Humidity, Rainfall (7)	Noise removal, normalization	**DT** SVM LR **GNB**	**99.30** **99.30**	Enhances productivity and profitability of agriculture.	Advanced algorithms and explainability with AI are not considered
G. Deng et al. [10]	2023	Kaggle (2200)	22	N, P, K, pH, Temperature (5)	Standardization	KNN XGBoost RF **Ensemble**	88.36 93.27 93.64 **94.36**	Ensemble learning combines strengths of multiple models	Risk of overfitting with complex ensemble methods
A. Lokhande et al. [11]	2022	Kaggle (2200)	22	N, P, K, pH, Temperature, Humidity, Rainfall (7)	Noise removal, normalization, outlier removal	DT SVM LR **RF**	90.00 96.08 95.22 **99.09**	Providing a GUI based application to predict which crop to grow	Explainable AI is not utilized to interpret which features are the most responsive

A web and mobile application-based smart crop recommendation system was proposed by Shwetha et al. [2] in 2024 creating a new dataset. The authors utilized 8 features including the crop length, soil type, season, pH, location, temperature, soil water content, and pesticide need. They implemented five ML models, of which Naive Bayes was found to be the best performer revealing an accuracy score of 95%. In 2023, Ajoodha et al. [1] presented a ML-based approach using only soil attributes data. The authors considered the standard Kaggle dataset [7] of 2200 instances, but they used only 1000 instances from them while taking only 4 soil features (nitrogen, phosphorus, potassium, and soil pH). They examined five ML algorithms and achieved the best accuracy of 91.3% using the K-star model. A smart farming mechanism was introduced by Dahiphale et al. [3] using ML techniques and the standard Kaggle dataset of [7] in 2023. Initially, the authors preprocessed the dataset using data cleaning and removal of outliers; furthermore implemented various ML models and acquired the highest results of 99.50% from the SVM and Naïve Bayes classifiers. This study discusses 13 key challenges and future ideas for agriculture and related sub-fields. Rajesh et al. [4] proposed another approach based on ML in 2023 addressing modern agricultural challenges by leveraging technology and precision agriculture techniques. They utilized various ML algorithms including Logistic Regression (LR), Decision Tree (DT), and Random Forest (RF), trained on a dataset obtained from the Kaggle [7] containing 7 features. RF outperformed others with 99.27% accuracy, followed by DT with 98.54% accuracy and LR with 66% accuracy. An efficient ML-based system was suggested by Patel et al. [9] in 2023 utilizing a Kaggle dataset [7] containing 2200 instances of historical agricultural data. The authors applied several preprocessing tasks including noise removal and normalization, and afterward executed several ML techniques such as DT, SVM, LR, and GNB. Among these models, DT and GNB revealed the best accuracy level of 99.30%. Deng et al. [10] designed an ensemble learning model for crop recommendation in 2023. The authors used a Kaggle public dataset [7] containing 2200 samples and then made use of feature selection using the genetic algorithm and finally reduced 2 out of 7 features. The ensemble learning model achieved 94.36% accuracy with 5 features, so the computational cost of the model is slightly reduced compared to 7 features. In 2022, Lokhande et al. [11] introduced a GUI-based system using ML, leveraging a dataset from Kaggle [7]. The authors employed various preprocessing chores including noise removal, outlier removal, and normalization, etc., and afterward examined several algorithms such as DT, SVM, LR, and RF. Through experimenting with these models, RF emerged as the most accurate model with 99.09% accuracy.

A. Observations from Related Works

The overall summary of the recent existing literature is demonstrated in Table I. By having a very close look at this table we have outlined several observations. These are-

1) Most studies used the standard dataset [7] from Kaggle containing 2200 instances.
2) Feature selection reduced the performance of the models (compare [1], [10] with other methods in Table I).
3) Ignoring data preprocessing causes a deduction in performance (compare [1] and other methods in Table I).
4) Explainable AI is not considered by most of the studies.

III. Research Methodology

The research methodology of the proposed system is given in Fig. 1.

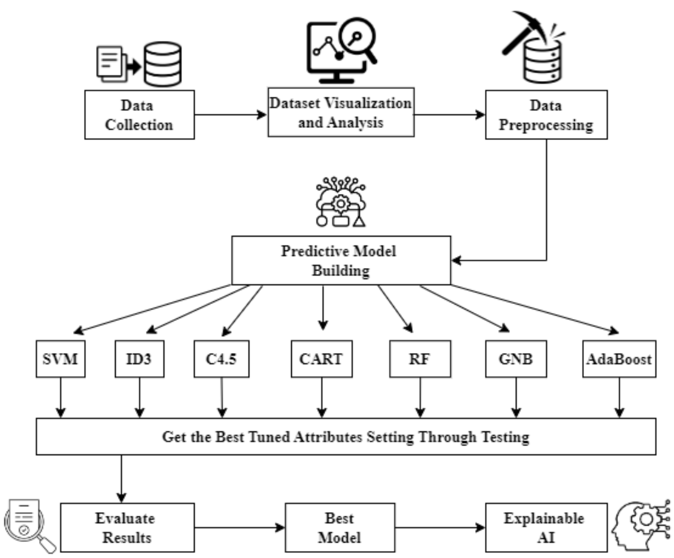

Fig. 1: Research methodology of the crop recommendation system

A. Data Collection

The initial movement of this study involved the collection of the dataset. This investigation accessed the crop recommendation dataset [7] from Kaggle containing 2200 examples for 22 crops, with 100 examples for each crop.

B. Dataset Visualization and Analysis

Table II briefly describes the features included in the dataset. The correlation matrix depicts a visual representation of the correlation between the features of the dataset and it is shown in Fig. 2. Correlation values between two features range from -1 to 1 where a value of -1 means a strong negative correlation, and a value of 1 illustrates a strong positive correlation. The correlation value between P and K is 0.74, reflecting a strong positive correlation while negative correlation is seen between the features N and P.

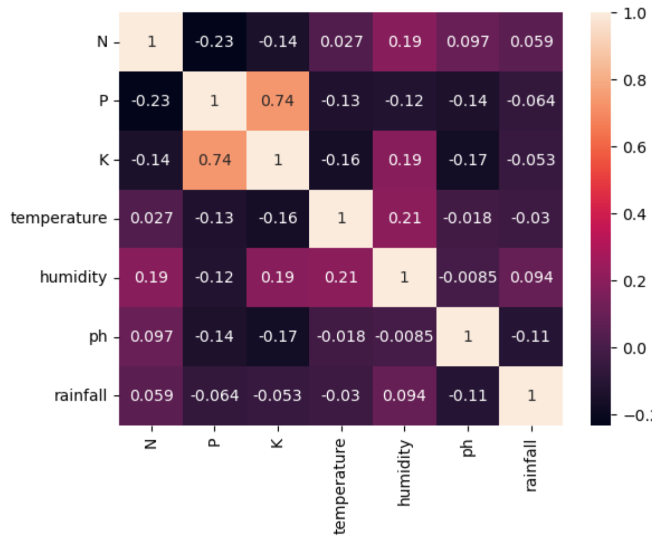

Fig. 2: Correlation among features of the dataset

TABLE II: Features of the Dataset

SN	Attribute	Data Type	Description	Mean	Min	Max
1	N	Numeric (Integer)	Nitrogen	50.55	0	140
2	P	Numeric (Integer)	Phosphorous	53.36	5	145
3	K	Numeric (Integer)	Potassium	48.14	5	205
4	Temperature	Numeric (Double)	Celsius scale temperature	25.61	8.82	43.67
5	Humidity	Numeric (Double)	Humidity as a percentage	71.48	14.25	99.98
6	pH	Numeric (Double)	pH value of soil	6.46	3.50	9.93
7	Rainfall	Numeric (Double)	Rainfall in millimeter	103.46	20.21	298.56
8	Target Label	Categorical	Grown crop	-	-	-

C. Data Preprocessing

Since raw data often contains incomplete, redundant, or inconsistent information, preprocessing is crucial for ensuring the success of a classification system [4].

1) Label Encoding: The process of converting labels into a numerical format is referred to as label encoding [12]. This transformation ensures that the model can effectively process categorical data, which is essential for many ML algorithms.

2) StandardScaler: The $StandardScaler$ from $scikit-learn$ is utilized for feature scaling via standardization [12]. This process ensures that each feature is transformed to possess a mean of 0 and a standard deviation of 1. The standardization formula is expressed as:

$$standard\ score, ss = \frac{x - \overline{x}}{sd(x)} \tag{1}$$

In (1), the symbols x, $\overline{x}$, and $sd(x)$ demonstrate the original value, mean, and standard deviation respectively.

D. Predictive Model Selection

This empirical study explores several ML algorithms including SVM, ID3, C4.5, CART, RF, GNB, and a boosting ensemble learner AdaBoost. The SVM functions by detecting the optimal hyperplane that best differentiates among the classes in the feature space [13]. ID3, a decision tree (DT) algorithm, selects the best attribute at each node based on information gain to split the dataset into subsets [14]. ID3 recursively builds the decision tree until all data instances are correctly classified or a stopping criterion is met. C4.5 is an extension of the ID3 algorithm which utilizes the gain ratio metric to select the best attribute [15]. The feature space is recursively partitioned into subsets based on the values of input features, optimizing Gini impurity or entropy, by another DT algorithm called CART. [16]. RF, an ensemble learning technique, generates multiple DTs during the training phase where each tree declares its prediction individually, and the final prediction is determined by majority voting from the individual predictions of all trees [17]. GNB is a Bayes theorem [observe (2)] based classification technique that relies on probabilities and follows a Gaussian distribution [observe (3)]; assuming that features are independent [18].

$$Bayes\ Theorem, P(R_i|S) = \frac{P(S|R_i)P(R_i)}{\sum_{i=1}^{m} P(R_i)P(S|R_i)} \quad (2)$$

$$P(R_i \mid S) = \frac{1}{\sqrt{2\pi\sigma_S^2}} \exp\left(-\frac{(R_i - \mu_S)^2}{2\sigma_S^2}\right) \quad (3)$$

In (2), $P(R|S)$ states the probability of appearance of R given Y has occurred, $P(R)$ denotes the prior probability, $P(S)$ demonstrates the probability of appearance of action S, $P(S|R)$ is the probability of appearance of S given R has already occurred. A boosting strong learner is created by combining multiple weak learners through an ensemble learning method called AdaBoost. It enhances the performance of weak classifiers by focusing on training incorrectly classified samples and iteratively trains models by giving more weight to misclassified data points, allowing subsequent models to focus on the difficult instances [19]. The architecture of the developed AdaBoost model is given in Fig. 3 where WL stands for the weak learner.

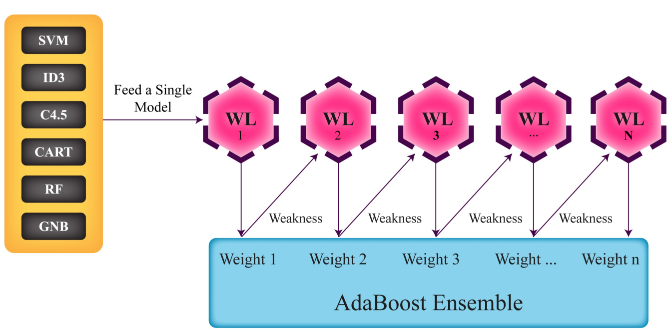

Fig. 3: Structure of the AdaBoost ensemble learner

TABLE III: Fine Tuned Parameters of the Implemented Algorithms

Base Classifier	Parameters	Ensemble Classifier	Parameters
SVM	kernel = rbf gamma=auto	AdaBoost+SVM	random_state=800 n_estimators=70 learning_rate=2
ID3	random_state=2 max_depth=5	AdaBoost+ID3	
C4.5		AdaBoost+C4.5	
CART		AdaBoost+CART	
RF	n_estimators=20 random_state=0	AdaBoost+RF	
GNB	default	AdaBoost+GNB	

IV. Analysis of Experimental Results

This section discusses the performance metrics, tuning of parameters, results analysis, and explainable AI respectively to evaluate the proposed crop recommendation system.

A. Performance Metrics

Recent studies suggest that accuracy, precision, recall, F1-score, area under the curve (AUC) value, and mean squared error (MSE) have become very crucial metrics to evaluate the efficiency of a machine learning model.

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Percision + Recall} \quad (6)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

$$MSE = \frac{1}{c}\sum_{i=1}^{c}(ex_i - pr_i)^2 \quad (8)$$

Precision and Recall are measured according to (4) and (5), and the harmonic mean of precision and recall is used to calculate the F1-score [observe (6)]. The ratio of correctly classified occurrences to all test occurrences is defined as accuracy. MSE is demonstrated by (8) where ex_i and pr_i denote the expected and predicted values out of c observations.

B. Tuning of Parameters

Through parameter fine-tuning, as detailed in Table III, each algorithm was fine-tuned to make the best predictions. These parameters are crucial for optimizing the algorithms and enhancing their predictive capabilities.

C. Results Analysis

The results of the implemented machine learning algorithms are summarized in Table IV where GNB produced the best precision score of 99.20, RF and GNB achieved the highest recall value of 0.9909, while RF outperformed the others with an F1-score of 0.9907, and RF acquired an MSE value of 1.2568. From the ML domain, GNB and RF are found to be the best performers exhibiting an outstanding accuracy of 99.09%.

TABLE IV: Results of the Implemented ML Algorithms

Algorithm	Precision	Recall	F1-Score	Accuracy (%)	AUC	MSE
SVM	0.9839	0.9818	0.9810	98.18	1.0000	1.9340
ID3	0.8572	0.9000	0.8707	90.00	1.0000	7.3727
C4.5	0.8572	0.9000	0.8707	90.00	1.0000	7.3727
CART	0.8416	0.8568	0.8491	85.63	0.9972	8.4386
RF	0.9917	**0.9909**	**0.9907**	**99.09**	1.0000	**1.2568**
GNB	**0.9920**	**0.9909**	0.9906	**99.09**	1.0000	1.3090

TABLE V: Results of the AdaBoost Ensemble Model with Different Base Classifiers

Algorithm	Precision	Recall	F1-Score	Accuracy (%)	AUC	MSE
AdaBoost+SVM	0.9195	0.8568	0.8485	85.68	0.9999	9.6795
AdaBoost+ID3	**0.9924**	**0.9944**	**0.9934**	**99.35**	1.0000	**0.7978**
AdaBoost+C4.5	0.9914	0.9909	0.9909	99.09	1.0000	0.8977
AdaBoost+CART	0.9864	0.9864	0.9863	98.64	1.0000	1.9113
AdaBoost+RF	0.9864	0.9864	0.9863	98.64	1.0000	1.9113
AdaBoost+GNB	0.9597	0.9455	0.9329	94.55	1.0000	5.6954

Table V illustrates the results of the AdaBoost ensemble model with six different base classifiers. The precision, recall, F1-score, and accuracy of both AdaBoost+ID3 and AdaBoost+C4.5 models all exceeded 99%, highlighting the effectiveness of AdaBoost in improving decision tree-based models. However, AdaBoost+SVM and AdaBoost+GNB show slightly lower performance.

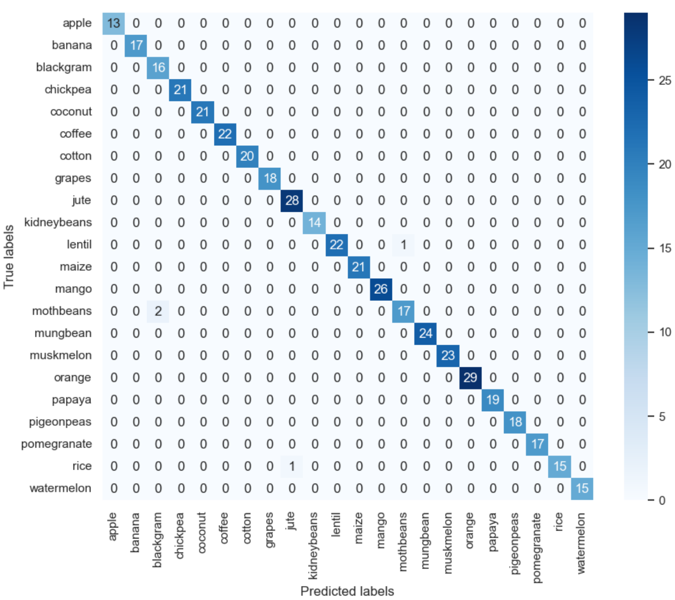

Fig. 4: Confusion matrix of the (AdaBoost+ID3) Model

Fig. 4 demonstrates the confusion matrix of the overall best-fitted model, AdaBoost+ID3. The confusion matrix provides a graphical depiction of the classification accuracy. This visual tool facilitates a detailed classification performance evaluation across diverse crop categories. The effect of AdaBoost model on different base classifiers is given in Fig. 5. The AdaBoost model significantly improves the performance of three DT-based classifiers (ID3, C4.5, and CART), while degrading the performance of the SVM, RF, and GNB models.

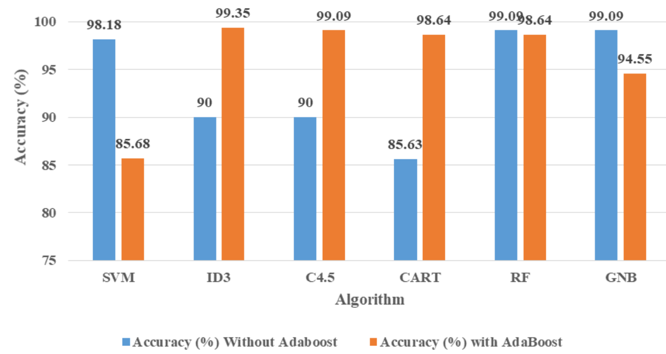

Fig. 5: Effect of AdaBoost model on base classifiers

D. Explainable AI

Interpretability of a black-box classifier is one of the trending questions of recent times [20] [21]. Fig. 6 shows a Force plot of Shapley values for the 1494th index of the dataset that truly belongs to the "muskmelon" crop class. The binary target is yes the crop recommendation belongs to "maskmelon" (maskmelon = 1), or no it does not (no maskmelon = 0). The score of the AdaBoost+ID3 model for the 1494th observation is 18.67. Higher scores predict 1, while lower scores predict 0 by the AdaBoost+ID3 model. The features N, P, and $humidity$ are responsible for pushing the model to score higher while the feature K scores lower. Features that are located closer to the dividing line between red and blue are represented as having a larger influence, and the size of that influence is determined by the size of the bar. So, N is the most impactful feature and $humidity$ is the least impactful feature. So this particular observation is finally classified by the AdaBoost+ID3 as "muskmelon" (1) because they push higher by all the factors shown in red (N, P, and $humidity$). Fig. 7 depicts a LIME explanation for the 2069th index of the dataset where the responsible features of classification are highlighted, and each probable predicted class of the model is shown with their corresponding prediction probabilities. The test instance is classified by the AdaBoost+ID3 model as "lentil" with 66% prediction probability.

E. Results Comparison

To understand how successful a study is, it is crucial to compare various key metrics with contemporary recent state-of-the-art studies on the same topic and achieve consistent results. A comparison of the results with the existing work and the proposed work is given in Table VI.

V. Conclusion

The smart crop recommendation system developed by this study exhibits the potential of machine learning and explainable AI in agriculture, particularly for crop selection. This empirical study uses a dataset that combines environmental, soil, and historical data to provide field-level recommendations and uses a technique called explainable AI by which the farmers can understand why it makes its decisions for

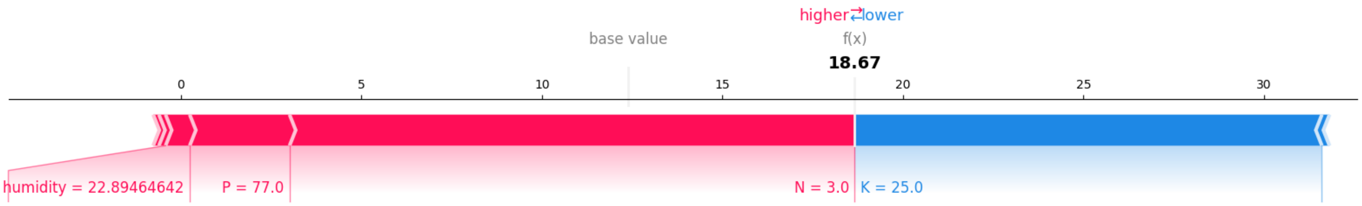

Fig. 6: SHAP explanation of index 1494 for the (AdaBoost+ID3) model

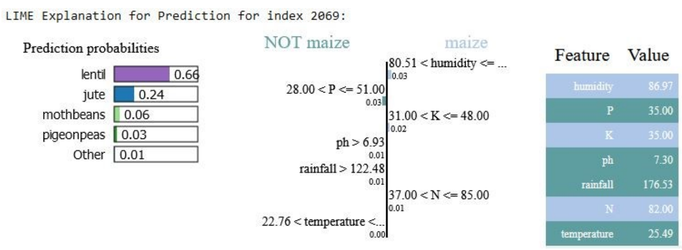

Fig. 7: LIME explanation of index 2069 for the (AdaBoost+ID3) model

TABLE VI: Results Comparison with Existing Works

Name	Year	Dataset Used	No. of Crops	No. of Features	Best Model	Accuracy (%)	Explainable AI
Shwetha et al. [2]	2024	N/A	N/A	8	NB	95.00	No
Ajoodha et al. [1]	2023	Kaggle (2200)	10	4	K-ster	91.30	No
Dahiphale et al. [3]	2023	Kaggle (2200)	22	7	RF	99.50	No
Rajesh et al. [4]	2023	Kaggle (2200)	22	5	RF	99.27	No
Patel et al. [9]	2023	Kaggle (2200)	22	7	DT	99.30	No
Deng et al. [10]	2023	Kaggle (2200)	22	5	Ensemble	94.36	No
Our Proposed	**2024**	**Kaggle (2200)**	**22**	**7**	**AdaBoost**	**99.35**	**Yes (SHAP & LIME)**

recommendations. This transparency is critical in establishing farmer confidence, without which the system's recommendations will not be effectively adopted. This research introduces the AdaBoost ensemble model with ID3 base learners for crop recognition, which achieves a marvellous accuracy of 99.35%. Visualizations from the SHAP and LIME modules are utilized to explain the crop recommendation decisions of the AdaBoost model. One limitation of this study is ignoring the feature selection step. In the future, our target is to develop a new dataset for crop recommendation based on Bangladesh's perspective, and a visualization software using explainable AI for the crop recommendation system. Overall, we are hopeful that this study will serve as a monument of progress in agriculture and the farmers, traders, governments, and private organizations can use it for their needs.

References

[1] R. Ajoodha and T. O. Mufamadi, "Crop recommendation using machine learning algorithms and soil attributes data," *Proceedings of 3rd International Conference on Artificial Intelligence: Advances and Applications: ICAIAA 2022*, pp. 31–41, 2023.

[2] A. Shwetha and P. PrabodhC, "Machine learning based smart crop recommendation system," in *2023 4th International Conference on Intelligent Technologies (CONIT)*. IEEE, 2024, pp. 1–6.

[3] D. Dahiphale, P. Shinde, K. Patil, and V. Dahiphale, "Smart farming: Crop recommendation using machine learning with challenges and future ideas," *Authorea Preprints*, 2023.

[4] C. Rajesh, V. Praveen, G. Deepak, K. Harshith, and M. V. S. Reddy, "Smart crop recommendation system using machine learning," *IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE)*, vol. 18, no. 3, pp. 49–55, 2023.

[5] A. Palakshappa, S. K. Nanjappa, P. Mahadevappa, and S. Sinchana, "Smart irrigation with crop recommendation using machine learning approach," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 3, pp. 1952–1960, 2024.

[6] Z. J. Peya, M. S. Islam, and M. K. N. Chumki, "Thyroid disease prediction based on feature selection and machine learning," in *2022 25th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2022, pp. 495–500.

[7] https://www.kaggle.com/datasets/atharvaingle/crop-recommendation dataset.

[8] A. Mayr, H. Binder, O. Gefeller, and M. Schmid, "The evolution of boosting algorithms," *Methods of information in medicine*, vol. 53, no. 06, pp. 419–427, 2014.

[9] M. Patel and A. Rane, "Crop recommendation system," 2023.

[10] G. Deng, C. Dong, J. Chen, R. Kong, J. Gao, and C. Li, "Crop planting recommendation algorithm based on ensemble learning," *Second International Symposium on Computer Applications and Information Systems (ISCAIS 2023)*, vol. 12721, pp. 337–343, 2023.

[11] A. Lokhande and M. Dixit, "Crop recommendation system using machine learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 9, no. 5, pp. 2395–0056, 2022.

[12] P. Kaware, S. Rithe, N. Kathane, and K. R. Gavhale, "Crop info web application using machine learning," *International Journal of Innovations in Engineering and Science*, vol. 7, no. 8, pp. 1–6, 2022.

[13] S. Suthaharan and S. Suthaharan, "Support vector machine," *Machine learning models and algorithms for big data classification: thinking with examples for effective learning*, pp. 207–235, 2016.

[14] A. E. Khedr, A. M. Idrees, and A. I. El Seddawy, "Enhancing iterative dichotomiser 3 algorithm for classification decision tree," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 6, no. 2, pp. 70–79, 2016.

[15] S. Ruggieri, "Efficient c4. 5 [classification algorithm]," *IEEE transactions on knowledge and data engineering*, vol. 14, no. 2, pp. 438–444, 2002.

[16] D. G. Denison, B. K. Mallick, and A. F. Smith, "A bayesian cart algorithm," *Biometrika*, vol. 85, no. 2, pp. 363–377, 1998.

[17] S. J. Rigatti, "Random forest," *Journal of Insurance Medicine*, vol. 47, no. 1, pp. 31–39, 2017.

[18] A. H. Jahromi and M. Taheri, "A non-parametric mixture of gaussian naive bayes classifiers based on local independent features," *2017 Artificial intelligence and signal processing conference (AISP)*, pp. 209–212, 2017.

[19] D. D. Margineantu and T. G. Dietterich, "Pruning adaptive boosting," *ICML*, vol. 97, pp. 211–218, 1997.

[20] M. Y. Shams, S. A. Gamel, and F. M. Talaat, "Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making," *Neural Computing and Applications*, vol. 36, no. 11, pp. 5695–5714, 2024.

[21] M. Habibullah, M. S. Islam, F. T. Jahura, and J. Biswas, "Bangla document classification based on machine learning and explainable nlp," in *2023 6th International Conference on Electrical Information and Communication Technology (EICT)*. IEEE, 2023, pp. 1–6.