

# RAGCrop: Intelligent Crop Recommendation via Multi-Attribute Learning and Expert-Augmented Dialogue

A.M.Z.I.A.A. Sujair Ibrahim, B. Gobihanath, V. Abishethvarman, S. Prasanth and B.T.G.S. Kumara

**Abstract** Agriculture has historically driven human civilization, but modern farming faces new challenges from soil diversity, climate variability, and the erosion of generational knowledge. These factors complicate accurate and timely decision-making, particularly in crop prediction. Current technological approaches often prioritize commodity production while overlooking the value of traditional farming expertise. To address this gap, this study proposes an integrated advisory system that combines statistical machine learning with Retrieval-Augmented Generation (RAG) to bridge data-driven prediction and expert agricultural knowledge. Unlike approaches that treat farmers' input merely as model attributes, our framework incorporates farmers' knowledge as a validated knowledge source through RAG, while more than 80 agronomic, environmental, and management-related features serve as attributes for statistical prediction. The system predicts optimal crops and simultaneously retrieves expert-verified information to provide context-aware, personalized responses. To ensure transparency, model outputs are explained using eXplainable AI (XAI) techniques such as LIME, which highlight the most influential attributes for each prediction. This design enhances both accuracy and trust, while enabling natural language dialogue that allows users to ask follow-up questions and receive expert-grounded explanations. Results demonstrate the feasibility of a scalable, explainable, and farmer-friendly advisory tool that promotes informed and sustainable agricultural practices. Moreover, the proposed framework is adaptable beyond crop prediction, offering potential for broader applications across diverse agricultural domains.

**Index Terms**— Agronomic, Agriculture, Intelligent Crop Framework, Multi-Attribute, Retrieval-Augmented Generation

## I. INTRODUCTION

AGRICULTURE is a key driver of economic growth and food security, particularly in low and middle-income nations where it employs a considerable proportion of the population [1]. In recent years, efforts to modernize agriculture through improved seed varieties, smart irrigation systems [2], and data-driven farming approaches [3] have shown strong potential to enhance productivity and sustainability. These advancements aim to support farmers in making more accurate and efficient decisions under changing environmental conditions [4]. Despite its critical role, the sector faces significant obstacles, including uncertain weather conditions [5], deteriorating soil fertility [6], and the loss of generational knowledge [7], all of

which complicate farm-level decision-making. One of the most persistent and significant challenges is farmers' struggle to decide which crops to grow in the face of environmental uncertainty and inadequate data availability [8]. Making the wrong choice not only threatens household livelihoods but also undermines national food systems.

Over the last two decades, advances in Artificial Intelligence (AI) and Machine Learning (ML) have started to alter agriculture by allowing for data-driven insights and automation. Applications including as yield forecasting, pest and disease detection, soil health monitoring, and precision irrigation have shown how these technologies may improve production and resilience. These developments illustrate the potential of intelligent systems to aid farmers' decisions. However, traditional crop recommendation methods using machine learning models such as decision trees, random forests [9], support vector machines (SVM) [10], k-nearest neighbors (KNN) [11], and rule-based classification systems [12], which often rely on broad extension standards or localized experiential knowledge, are becoming increasingly ineffective in meeting the dynamic and site-specific demands of modern agriculture. Such approaches often fail to account for the growing complexity of agricultural ecosystems, providing farmers with advice that is either too broad to be practical or too rigid to adjust to fast changing conditions [13]. Consequently, there is a widening gap between the decision support farmers need and the tools currently available to them.

A.M.Z.I.A.A. Sujair Ibrahim is undergraduate from the Faculty of Computing, Sabaragamuwa University of Sri Lanka, Sri Lanka.

B. Gobihanath is undergraduate from the Faculty of Computing, Sabaragamuwa University of Sri Lanka, Sri Lanka.

V. Abishethvarman is undergraduate from the Faculty of Computing, Sabaragamuwa University of Sri Lanka, Sri Lanka.

S. Prasanth is from the Faculty of Engineering & Applied Science, Memorial University of Newfoundland, Canada.

B.T.G.S. Kumara is from the Faculty of Computing, Sabaragamuwa University of Sri Lanka, Sri Lanka. (Email: [bgobihanath@std.appsc.sab.ac.lk](mailto:bgobihanath@std.appsc.sab.ac.lk))

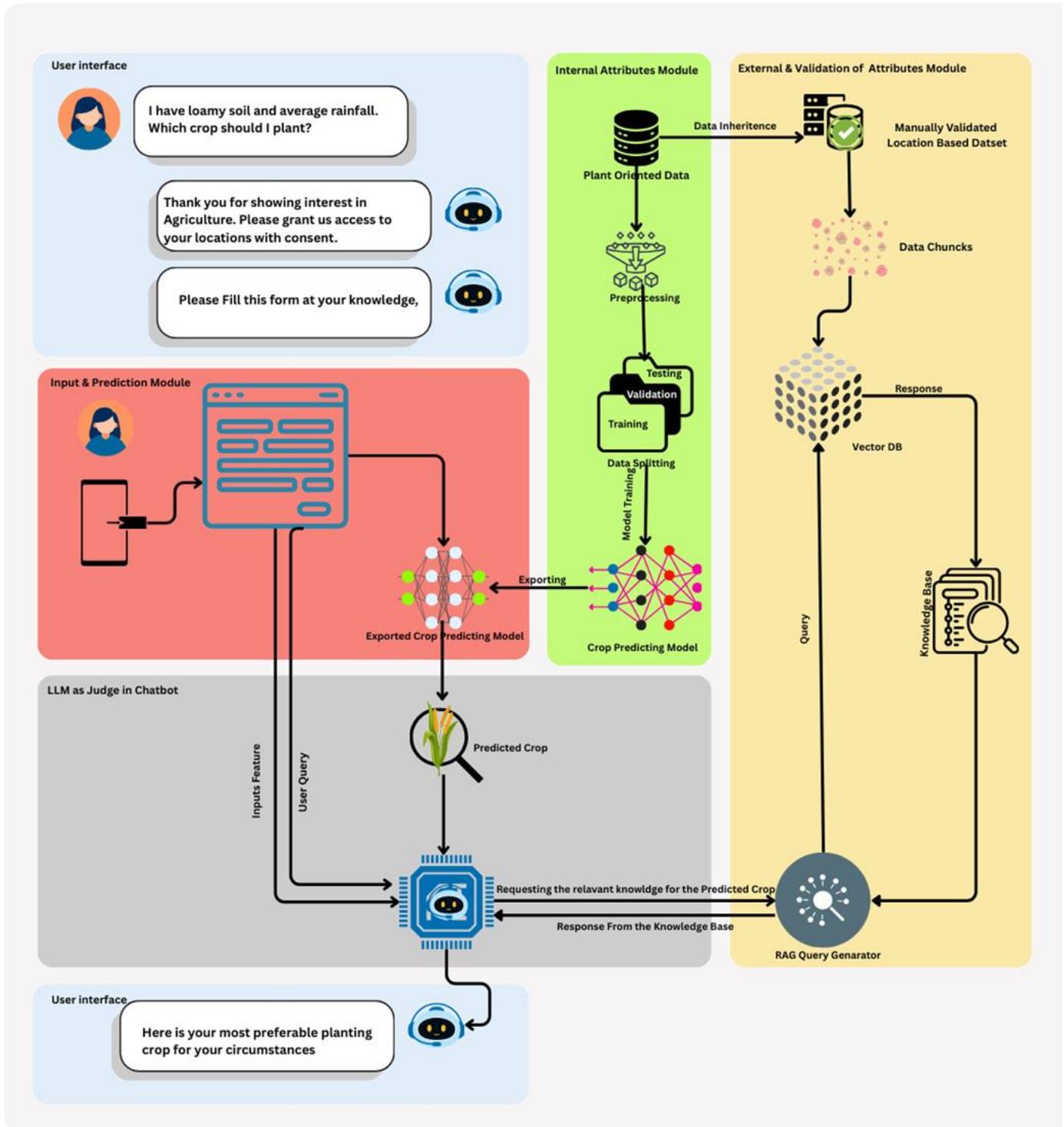


Fig. 1: Workflow Structure

In response to this critical problem, this study proposes an advanced framework for an intelligent crop recommendation system that leverages Machine Learning (ML), Large Language Models (LLMs), and Retrieval Augmented Generation (RAG) to deliver both precise predictions and expert-level, context-aware explanations. As illustrated in Fig. 1, the system follows a multi-layered workflow designed to address both the accuracy and interpretability of recommendations. At its foundation lies

a comprehensive crop prediction model trained on more than 80 features, encompassing soil chemistry, weather patterns, geographical characteristics, and agronomic indicators. Unlike traditional approaches constrained by a narrow set of inputs, this multi-attribute strategy produces richer, data-driven insights into crop suitability and generates personalized recommendations that align with the unique environmental and operational contexts of farmers.

A distinguishing feature of the proposed system is its dual-layer architecture. The first layer applies supervised ML models to structured datasets in order to predict the most suitable crop(s) under specific agricultural conditions. The second layer builds upon these predictions by using a Large Language Model (LLM) embedded within a Retrieval-Augmented Generation (RAG) [14] pipeline. Here, the LLM generates natural language explanations while retrieving relevant, evidence-based knowledge from a curated vector library of expert insights, scientific literature, and validated agricultural practices. Together, this dual approach delivers not only accurate crop forecasts but also transparent, expert-grounded reasoning that enhances trust, usability, and practical value for farmers.

By directly addressing the inadequacies of existing systems, many of which operate as opaque black boxes with little transparency, the proposed solution bridges the gap between algorithmic performance and practical use. Its emphasis on explainability promotes trust and better educated use of AI-powered advisory. Furthermore, the system's architecture is specifically intended for scalability. While the current implementation focuses on crop-related data, it can be expanded to include factors such as plant physiology, environmental health, pest and disease profiles, and more. This versatility prepares the framework for more widespread use in agricultural decision-making domains such as irrigation scheduling, disease diagnostics, and yield forecasts.

Building on this foundation, the scope of the present study is centered on developing a robust, scalable, and explainable decision support system to improve crop selection procedures in the agricultural domain. By addressing the constraints of traditional crop advising systems, which frequently lack specificity, flexibility, and interpretability, the proposed framework intends to assist farmers in making precise, data-driven decisions that are adapted to their individual production conditions.

At its core is a multi-attribute crop prediction model that has been trained on over 80 different characteristics, including soil nutrients, pH level, rainfall, humidity, temperature, location data, and other agronomic indicators. This high-dimensional data method allows the algorithm to make more precise and context-sensitive crop suggestions. Importantly, the system incorporates Explainable AI (XAI) approaches for visualizing and interpreting how each unique variable affects the model's outcome [15]. This component gives transparency to the end user, making the decision-making process more intelligible and trustworthy.

The second layer of the structure includes a conversational, expert-oriented dialogue system driven by a Large Language Model (LLM) with Retrieval-Augmented Generation (RAG). This layer enables the system to dynamically extract expert information from a curated vector library consists of validated agricultural research, farmer-provided best practices, and regional farming guidelines. The LLM synthesizes this information in real time to provide contextualized, natural language replies that explain the reasons behind each suggestion and offer practical, actionable advice regarding crop management, irrigation, fertilizer application, and more.

By merging predictive analytics, domain-specific retrieval, and natural language production, the suggested approach goes beyond static suggestions. It develops an interactive support platform that not only identifies best crops based on environmental and soil conditions but also provides expert-informed explanations and recommendations tailored to the user's requirements and circumstances. This design improves significantly in its capacity to reduce probable flaws in the ML model's predictions. In circumstances where the model's output is incorrect or unreliable, the RAG component serves as a correction layer, gathering and transmitting accurate data from the manually checked expert dataset. This fallback strategy guarantees that the system remains reliable and effective even in cases with low predicted confidence, lowering the risk of misinformation and enhancing overall decision quality.

In summary, the main objective of this research is to provide a generalizable and extendable framework for intelligent agricultural decision support that can process high-dimensional inputs, make visible and explainable predictions, and facilitate human-centered dialog. Although the current focus is on crop prediction, the underlying structure can be extended and scaled to other agricultural decision-making tasks such as pest control, disease testing, and resource optimization. The system's use of ML, RAG, and XAI aims to provide access to expert agronomic knowledge, equip farmers with evidence-based tools, and contribute to long-term agricultural growth and national food security.

## II. LITERATURE REVIEW

Agriculture is currently undergoing unprecedented pressure from climatic variability, soil degradation, and resource limits, while global food demand continues to climb at an unsustainable rate [16], [17]. These issues make farm-level decision-making, particularly crop selection, more complex and high-risk, with significant consequences for food security and rural livelihoods [18]. Traditional advisory approaches, which are frequently based on broad suggestions or specialized expertise [19], are no longer adequate to meet the dynamic, site-specific requirements of modern agricultural systems [20], [21]. In this context, machine learning (ML) and artificial intelligence (AI) have emerged as revolutionary tools, providing predictive, data-driven, and adaptive solutions that can optimize agricultural operations far beyond the limits of traditional methods [22], [23].

In recent years, there has been a growing interest in applying machine learning (ML) approaches to agricultural problems, motivated by concerns such as environmental unpredictability, resource restrictions, and rising food demand [24]. Numerous studies [25], [26], [27], [28] have explored the use of supervised machine learning models to help farmers make decisions, notably when selecting crops. K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Random Forests (RF), and Artificial Neural Networks (ANN) have all been extensively tested for their predictive ability to improve farm-level planning and production.

In a similar research [29], soil data from various geographical zones were used to evaluate the effectiveness of algorithms including Naïve Bayes, Bayes Net, and Instance Based Learner

(IBK) for crop recommendations. The study found that these classifiers can effectively guide farmer decisions based on local soil variables. Another study [30] found Naïve Bayes to be 99.45% accurate in selecting ideal crops using a three-phase methodology that included data preparation, classification, and performance assessment. Although promising, these models lack contextual awareness and flexibility, making them restricted in real-world applications. [31] conducted a comparative analysis of fifteen machine learning algorithms to evaluate their performance in crop prediction tasks. The study used real-time agricultural data collected by IoT-enabled sensors and emphasized the importance of feature selection techniques in enhancing model performance.

In previous studies, Artificial Neural Networks (ANNs) were widely used to estimate crop output by modelling the complicated, non-linear interactions between meteorological conditions, soil qualities, and agricultural productivity. For example, [32] created an ANN-based system that forecasted crop production with high accuracy using numerous environmental and soil characteristics such as pH, nitrogen, phosphorus, potassium, temperature, rainfall, and humidity. The study demonstrated the adaptability of feed-forward backpropagation neural networks in capturing complex relationships between input variables and yield outcomes. By experimenting with various network topologies and optimizing the number of hidden layers, the researchers proved that ANNs may successfully help decision-making in data-rich agricultural environments.

Recent studies on RA-DSSs (Retrieval-Augmented Decision Support Systems) are a significant improvement that combine ML algorithms' inference skills with dynamic access to external knowledge stores. This hybrid design improves system flexibility and relevance for various decision situations including nutrient management, irrigation scheduling, and yield estimate. Empirical studies show that RA-DSSs boost performance and user engagement, especially for high-stakes agricultural choices [33]. The introduction of Large Language Models (LLMs) in these systems has improved the area by providing natural language advisory, improving clarity, and decreasing cognitive burden for end users.

Study [34] conducted a thorough study to critically assess the rising importance of Large Language Models (LLMs) in decision-making frameworks for agricultural surveillance. The authors emphasized that, while LLMs have tremendous promise for automating and improving agrotechnical decision-making, their efficacy is frequently restricted by static training data and domain specialization. To address these constraints, the study emphasizes the use of Retrieval-Augmented Generation (RAG) as a major technique for extending LLM capabilities through real-time access to external, domain-specific information. This reveals the key research possibilities for increasing trustworthiness, explainability, and bias reduction in LLM-based agricultural systems.

Prior research has shown that machine learning methods, particularly supervised classifiers and artificial neural networks, are excellent for crop prediction, but these systems frequently lack contextual flexibility and provide static results. Similarly, while recent advances in Retrieval-Augmented Decision

Support Systems (RA-DSSs) and Large Language Models (LLMs) have offered dynamic, explainable advising capabilities, their incorporation into practical, end-to-end agricultural applications is still restricted. Existing research treats prediction and advice as distinct activities, resulting in a significant gap in the creation of unified frameworks that integrate high-dimensional, data-driven prediction with interactive, expert-oriented conversation.

This work fills that gap by presenting a new architecture that combines a multi-attribute crop prediction model with a RAG-enhanced LLM advice system. It accomplishes this by expanding on the advantages of previous ML-based systems and introducing an explainable, scalable, and conversational layer for real-time, context-aware agricultural decision help. The system not only enhances prediction accuracy and user engagement, but it also provides a fallback mechanism that boosts reliability by accessing manually vetted expert information when the machine learning model's confidence is low.

### III. METHODOLOGY

#### A. Data

To investigate the feasibility of this significant feature integration, an open-source dataset from Kaggle with approximately 2,200 entries was chosen and prepared using a systematic preparation procedure. To reduce measurement inconsistencies, the raw dataset was first cleaned by removing noise and outliers. Continuous variables, such as soil nutrients and temperature, were standardized to a standard scale, while categorical features were encoded using one-hot encoding to ensure interoperability with machine learning techniques. An 80/20 train-test split was used, with 20% of the data kept for final validation. In parallel, real-time soil data was collected using a portable IoT-based soil testing device that measures on-site parameters such as pH, temperature, and nutrient content. To improve dependability, measurements were collected from at least five distinct locations within each field and averaged to create a representative soil profile. The dataset used for training and evaluation was made more robust and contextually relevant by combining open-source and field-level data.

The knowledge foundation for the RAG system was created by curating qualitative information from experienced farmers and agricultural extension sources. This contained domain-specific information on irrigation procedures, fertilizer application, pest and disease control, crop rotation plans, soil preparation methods, care suggestions for different development stages, and water management practices. Before being incorporated in the vector database, agricultural experts carefully reviewed each document or knowledge segment to verify correctness, geographical relevance, and contextual consistency. This rigorous validation verified that, once operational, the RAG module could provide contextually grounded and trusted solutions to queries.

However, due to resources and access restrictions, manual validation by agricultural specialists was not conducted during the present experimental phase of this study. This highlights a significant limitation of the current investigation. Nonetheless,

the proposed architecture emphasizes the importance of expert validation for both the numeric dataset and the textual knowledge base before real-world implementation. Before being embedded in the vector database, the collected records went through a preprocessing phase that included cleaning and normalizing the text, removing stop words, and correcting formatting inconsistencies.

### B. Experimental Setup

This study employs a structured methodology consists of three main phases to create an integrated, and context aware crop advice system. The first step is to build a high-dimensional machine learning (ML) model for crop prediction based on a large number of agronomic, environmental, and historical data. The second phase is on the creation of a Retrieval-Augmented Generation (RAG) system that would produce expert oriented natural language answers based on projected outcomes. The last phase entails combining these two components into a single decision support system that improves both prediction accuracy and contextual interpretability.

The first phase involves the development of a supervised ML model trained on a dataset containing over 80 agronomic, environmental, and management-related variables. These characteristics were deliberately chosen to highlight the complicated nature of crop farming. The dataset covers 80 basic macronutrients: manganese (Mn), copper (Cu), boron (B), molybdenum (Mo), soil salinity, soil texture, sulfur (S), chlorine (Cl), soil respiration rate, porosity (percentage), bulk density, base saturation (percentage), calcium (Ca), magnesium (Mg), silicon (Si), electrical conductivity (EC), cation exchange capacity (CEC), buffer pH, organic carbon (OC), organic matter (OM), soil texture (sand), soil texture (silt), soil texture (clay), water holding capacity, soil moisture content, soil temperature, solar radiation, wind speed, evapotranspiration, crop rotation, fertilizer type, fertilizer application rate, fertilizer timing, irrigation type, irrigation volume, tillage method, cover crops, pesticide use, herbicide use, elevation, slope gradient, slope aspect, microbial biomass, earthworm count, soil enzyme activity, latitude, longitude, soil depth, season, year, historical land use, maximum temperature, minimum temperature, average temperature, maximum rainfall, minimum rainfall, average rainfall, maximum humidity, minimum humidity, average humidity, maximum solar radiation, minimum solar radiation, average solar radiation, maximum wind speed, minimum wind speed, and average wind speed.

Data preprocessing involved normalization of continuous variables, categorical encoding of management practices, and imputation for missing values using k-nearest neighbors (KNN) techniques. Feature selection methods, including correlation analysis and Recursive Feature Elimination (RFE), were employed to reduce redundancy while preserving predictive relevance.

The Random Forest classifier was chosen as the prediction model because it is resistant to overfitting and can handle diverse, high-dimensional datasets. The number of trees, maximum depth, and minimum samples per split were all optimized using grid search and five-fold cross-validation. The dataset was divided into two subsets: training (80%) and

validation (20%), providing balanced representation across crop classes. Model performance was assessed using accuracy, precision, recall, and F1-score, with a focus on generalizability to previously encountered data.

Explainable AI (XAI) approaches were used to increase the prediction pipeline's interpretability. LIME (Local Interpretable Model-agnostic Explanations) was used to create local, instance-specific explanations that demonstrated how each input attribute contributed to a certain prediction. For example, for a predicted crop, the system identifies which elements, such as nitrogen levels, rainfall, and soil pH, had the greatest influence on the suggestion. These visualizations improve transparency and assist farmers and stakeholders understand the agronomic reasoning behind each choice, fostering trust in the system.

The second step of the process involves the creation of the Retrieval-Augmented Generation (RAG) system, which improves the explainability and contextual relevance of crop predictions by delivering expert-oriented natural language answers. This phase begins with the compilation of two distinct knowledge sources: crop-oriented data, which includes agronomic best practices, and location-based farmer knowledge, which includes region-specific techniques and traditional insights into irrigation, soil preparation, crop rotation, pest management, and general care practices.

To optimize retrieval granularity, documents are broken down into cohesive, semantically relevant chunks. ChromaDB was chosen for this vector database creation. The vector database contains both crop-oriented information and location-specific farmer data, allowing the system to return context-aware, expert-informed solutions based on individual user inputs.

The RAG system works using a Large Language Model (LLM) as the generating layer. Importantly, the LLM does not function in isolation; rather, it augments user queries with retrieved expert information before creating responses. This architecture ensures that recommendations are based on real information rather than uncontrolled generating outputs. RAG results were evaluated using both automated metrics (BLEU, ROUGE-L, METEOR) and human expert validation, with an emphasis on factual accuracy, contextual relevance, and explanatory clarity.

The final stage of the process is integrating the machine learning-based crop prediction model with the Retrieval Augmented Generation (RAG) system to create a unified, interactive, and context-aware crop advice application. This integrated system is intended to receive user input, provide data driven crop recommendations, and reply to user enquiries with expert validated, context-aware advice via a conversational interface. The process starts when a user interacts with the program using a form-based interface integrated in a chat app. These input features are then delivered to the exported machine learning model, which processes them and provides the predicted crop that best fits the given conditions.

The anticipated crop, original input features, and user inquiries are then sent into the RAG-enhanced LLM. The technology extracts expert-validated knowledge segments from the vector database and converts them into structured natural language explanations. Along with textual responses, XAI

graphics are used to highlight the most influential aspects driving the forecast. This combined strategy ensures predicted accuracy and interpretability. The built-in validation process improves reliability. When the ML model generates low-confidence predictions, the system compares them to expert-validated knowledge in the vector store. In the event of a discrepancy, the expert-informed retrieval layer acts as a fallback, ensuring that users obtain reliable recommendations even when the model alone is insufficient.

The approach provides interpretable, expert-backed insights as well as reliable crop recommendations due to the seamless integration of predictive modelling and information retrieval. It boosts user confidence, promotes informed decision-making, and fills one of the most crucial deficiencies in traditional agricultural advice tools.

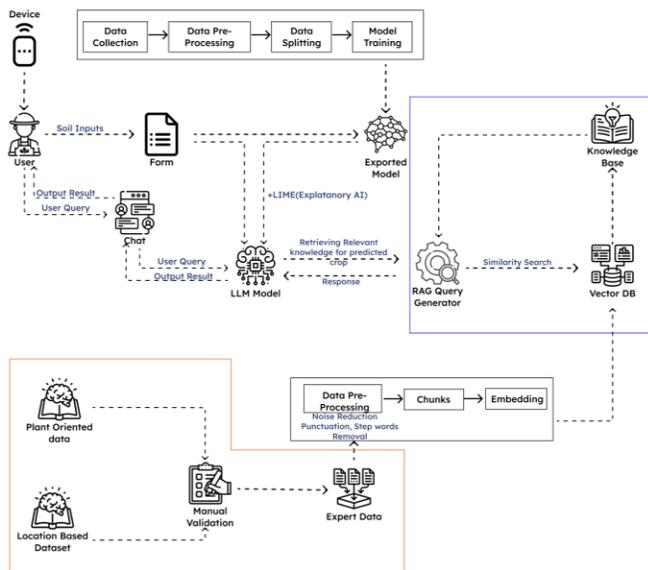


Fig. 2: Methodology

IV. RESULTS AND DISCUSSION

The performance of the integrated crop advisory system was assessed using two basic components: the machine learning based crop prediction model and the Retrieval Augmented Generation response module. The results show that the system can give both accurate predictions and contextually appropriate natural language explanations, proving the viability of a multi-attribute, explainable decision support system for agricultural applications. The machine learning model, which was trained on over 80 carefully selected agronomic, environmental, and managerial parameters, with Random Forest Classifier performed well on the holdout validation dataset. The model's accuracy was 98.86%, with a precision of 98.97%, recall of 98.86%, and F1 score of 98.87%, indicating that the model could generalize well to new data. The high F1 score, which balances both precision and recall, further confirms the model's robustness and reliability in making consistent and accurate crop recommendations across diverse conditions.

TABLE I

MODEL'S PERFORMANCE METRICS	
Metrics	Value
Accuracy	98.86%
Precision	98.97%
Recall	98.86%
F1 Score	98.87%

To tackle the interpretability of the model's results, especially given the large dimensionality of the feature space, Local Interpretable Model-agnostic Explanations (LIME) were used. The use of LIME enabled the creation of local explanations that highlighted the most important elements for individual forecasts [35]. Fig. 3 shows that nitrogen levels over 85.00, moderate potassium concentrations, and particular rainfall, humidity, and pH thresholds all have a significant impact on the forecast for the crop class "banana". By giving such visual and numerical breakdowns, the system provides insight into the model's internal reasoning process, which is crucial for deploying AI solutions in trust-sensitive areas such as agriculture.

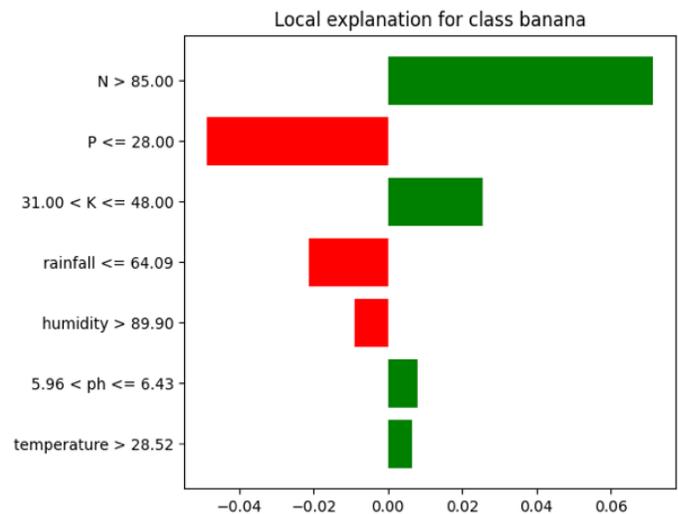


Fig. 3: LIME Explanation

The successful implementation of LIME demonstrates that multi-attribute crop recommendation systems can be created and explained, even when functioning on datasets containing dozens or hundreds of variables. This increases user trust by allowing domain experts to assess or cross-check the model's reasoning process. The quality of the responses generated by the system was evaluated using standard natural language generation metrics. As a baseline, the model was first tested using only the Large Language Model (LLM) without Retrieval Augmented Generation (RAG), achieving a ROUGE-L score of 44, a METEOR score of 28.9, and a BLEU score of 33.78. After integrating the RAG mechanism, the system showed a clear performance improvement, obtaining a ROUGE-L score of 50.5, a METEOR score of 36.5, and a BLEU score of 40.1, which indicates enhanced fluency, relevance, and factual correctness of the generated explanations. This notable increase across all metrics demonstrates the effectiveness of

incorporating expert-validated knowledge through RAG, resulting in stronger alignment with reference expert responses and reinforcing the chatbot interface's trustworthiness and contextual value. Furthermore, the system's average reaction time was observed at 1.45 seconds, proving its capability for near real-time interaction.

## V. CONCLUSION

This study presents an integrated framework for multi-factor crop prediction improved with expert-oriented discussion, integrating machine learning and retrieval-augmented generation (RAG) to handle the increased complexity of agricultural decision-making. The suggested method provides extremely accurate and interpretable crop recommendations compared to Traditional recommendation models by including more than 80 agronomic, environmental, and management factors. The application of explainable AI approaches, such as LIME, promotes transparency by demonstrating the influence of specific attributes on model outputs, hence increasing user trust even in high-dimensional environments.

Beyond prediction, integrating a RAG-based conversational module allows for context-aware, expert-oriented replies that are personalized to each user's individual input and enquiries. This dual-layered solution not only bridges the gap between algorithmic prediction and real-world agricultural knowledge, but it also reduces prediction mistakes by validating against a manually curated knowledge base. The findings demonstrate the viability and dependability of implementing a multi-attribute, expert oriented dialogue driven crop advice system capable of promoting sustainable and informed agricultural practices. Future developments will concentrate on field deployment, extending expert knowledge sources, and adapting the system to new crop domains and locales.

First, the training dataset and the expert knowledge base were not manually validated throughout the experimental phase due to resource and access limitations. The dependability of some responses may be impacted by the lack of validation in this study, despite the fact that it is highly advised for real-world implementation. Furthermore, a comparatively small dataset that was artificially extended was used to evaluate the system, which might not accurately represent the variety of actual agricultural situations. Lastly, the RAG module's evaluation was based on common NLP metrics rather than domain-specific user feedback. Future field-level testing should address this issue to determine the module's usability and practical impact.

In order to verify the suggested system's practical efficacy, future study will concentrate on putting it into practice in actual agricultural environments. This entails gathering data unique to a certain location, incorporating sensor-based live inputs, and enlisting the help of agricultural specialists to manually verify the RAG system's knowledge base and training datasets. The system's versatility will also be increased by broadening the information base to include other crop kinds, pest and disease situations, and regional best practices. In order to further customize the dialogue-based responses, improvements will also concentrate on improving the user interface and gathering input from agronomists and farmers. Lastly, the framework's usefulness across many areas and farming communities can be

expanded by scaling it for multilingual support and distinct agro-climatic zones.

## REFERENCES

- [1] A. Abdimomynova, E. Kolpak, B. Doskaliyeva, D. Stepanova, and V. Prasolov, "Agricultural diversification in low-and middle-income countries: Impact on food security," *Montenegrin Journal of Economics*, vol. 15, no. 3, pp. 167–178, 2019.
- [2] Y. Gamal, A. Soltan, L. A. Said, A. H. Madian, and A. G. Radwan, "Smart irrigation systems: Overview," *Ieee Access*, 2023.
- [3] D. Jiménez et al., "A scalable scheme to implement data-driven agriculture for small-scale farmers," *Glob. Food Sec.*, vol. 23, pp. 256–266, 2019.
- [4] I. Darnhofer, S. Bellon, B. Dedieu, and R. Milestad, "Adaptiveness to enhance the sustainability of farming systems. A review," *Agron. Sustain. Dev.*, vol. 30, no. 3, pp. 545–555, 2010.
- [5] K. E. Ukhrebor et al., "Precision agriculture: Weather forecasting for future farming," in *AI, edge and iot-based smart agriculture*, Elsevier, 2022, pp. 101–121.
- [6] M. A. E. AbdelRahman, M. M. Metwaly, A. A. Afifi, P. D'Antonio, and A. Scopa, "Assessment of soil fertility status under soil degradation rate using geomatics in West Nile Delta," *Land (Basel)*, vol. 11, no. 8, p. 1256, 2022.
- [7] A. A. Melash et al., "Indigenous agricultural knowledge: A neglected human based resource for sustainable crop protection and production," *Heliyon*, vol. 9, no. 1, 2023.
- [8] M. V. K. Sivakumar, "Climate prediction and agriculture: current status and future challenges," *Clim. Res.*, vol. 33, no. 1, pp. 3–17, 2006.
- [9] M. Keerthana, K. J. M. Meghana, S. Pravalika, and M. Kavitha, "An ensemble algorithm for crop yield prediction," in *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, 2021, pp. 963–970.
- [10] N. Gandhi, L. J. Armstrong, O. Petkar, and A. K. Tripathy, "Rice crop yield prediction in India using support vector machines," in *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2016, pp. 1–5.
- [11] R. Kumar, M. Gupta, and U. Singh, "Precision agriculture crop recommendation system using knn algorithm," in *2023 International Conference on IoT, Communication and Automation Technology (ICICAT)*, 2023, pp. 1–6.
- [12] C. Musanase, A. Vodacek, D. Hanyurwimfura, A. Uwitonze, and I. Kabandana, "Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices," *Agriculture*, vol. 13, no. 11, p. 2141, 2023.
- [13] P. Khatri, P. Kumar, K. S. Shakya, M. C. Kirlas, and K. K. Tiwari, "Understanding the intertwined nature of rising multiple risks in modern agriculture and food system," *Environ. Dev. Sustain.*, vol. 26, no. 9, pp. 24107–24150, 2024.
- [14] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive nlp tasks," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 9459–9474, 2020.
- [15] R. Dwivedi et al., "Explainable AI (XAI): Core ideas, techniques, and solutions," *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–33, 2023.
- [16] H. Mishra, "Environmental degradation and impacts on agricultural production: A challenge to urban sustainability," in *Sustainable urban environment and waste management: Theory and practice*, Springer, 2025, pp. 53–92.
- [17] A. Dubey, P. K. Singh, S. Singh, U. Manzoor, C. Sahoo, and A. Saini, "Impact of Climate Change and Land Degradation on Agriculture," in *Eco-Resilience: Climate Change, Land Degradation and Sustainable Solutions*, Springer, 2025, pp. 49–95.
- [18] S. E. Abebaw, "A Global Review of the Impacts of Climate Change and Variability on Agricultural Productivity and Farmers' Adaptation Strategies," *Food Sci. Nutr.*, vol. 13, no. 5, p. e70260, May 2025, doi: 10.1002/FSN3.70260.
- [19] S. Adla et al., "Participatory development of mobile agricultural advisory driven by behavioural determinants of adoption," *J. Environ. Manage.*, vol. 374, p. 124140, 2025, doi: <https://doi.org/10.1016/j.jenvman.2025.124140>.
- [20] E. Mugaru et al., "Influence of public agricultural extension services on sustainable land management practice adoption among smallholder farmers in Fetakgomo Tubatse Local Municipality, South Africa," *Front. Sustain. Food Syst.*, vol. 9, p. 1618938, 2025.
- [21] M. Nkosi, A. I. Agholor, and O. D. Olorunfemi, "Examining the Sustainability of Precision Agriculture and the Role of Advisory Services in South Africa: An Integrative Review of Potentials, Challenges, And Strategies,"

- South African Journal of Agricultural Extension (SAJAE), vol. 53, no. 4, pp. 56–85, Oct. 2025, doi: 10.17159/2413-3221/2025/v53n4a17426.
- [22] H. Gunasekaran, D. Kanmani, and R. Krishnamoorthi, “Optimized Ensemble Learning for Enhanced Crop Recommendations: Leveraging ML for Smarter Agricultural Decision-Making,” *Engineering Proceedings*, vol. 82, no. 1, 2024, doi: 10.3390/ecsa-11-20366.
- [23] F. S. Prity et al., “Enhancing Agricultural Productivity: A Machine Learning Approach to Crop Recommendations,” *Human-Centric Intelligent Systems*, vol. 4, no. 4, pp. 497–510, 2024, doi: 10.1007/s44230-024-00081-3.
- [24] S. O. Araujo, R. S. Peres, J. C. Ramalho, F. Lidon, and J. Barata, “Machine learning applications in agriculture: current trends, challenges, and future perspectives,” *Agronomy*, vol. 13, no. 12, p. 2976, 2023.
- [25] M. G. Ps and R. Bhargavi, “Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms,” *Applied Artificial Intelligence*, vol. 33, no. 7, pp. 621–642, 2019.
- [26] M. Hasan et al., “Ensemble machine learning-based recommendation system for effective prediction of suitable agricultural crop cultivation,” *Front. Plant Sci.*, vol. 14, p. 1234555, 2023.
- [27] S. Rani, A. K. Mishra, A. Kataria, S. Mallik, and H. Qin, “Machine learning-based optimal crop selection system in smart agriculture,” *Sci. Rep.*, vol. 13, no. 1, p. 15997, 2023.
- [28] I. Attri, L. K. Awasthi, and T. P. Sharma, “Machine learning in agriculture: a review of crop management applications,” *Multimed. Tools Appl.*, vol. 83, no. 5, pp. 12875–12915, 2024.
- [29] M. Chandraprabha and R. K. Dhanaraj, “Soil based prediction for crop yield using predictive analytics,” in 2021 3rd international conference on advances in computing, communication control and networking (ICAC3N), 2021, pp. 265–270.
- [30] Ishak, M. S. Rahaman, and T. Mahmud, “FarmEasy: An intelligent platform to empower crops prediction and crops marketing,” in 2021 13th international conference on information & communication technology and system (ICTS). IEEE, 2021.
- [31] E. Elbasi et al., “Crop prediction model using machine learning algorithms,” *Applied Sciences*, vol. 13, no. 16, p. 9288, 2023.
- [32] S. S. Dahikar and S. V Rode, “Agricultural crop yield prediction using artificial neural network approach,” *International journal of innovative research in electrical, electronics, instrumentation and control engineering*, vol. 2, no. 1, pp. 683–686, 2014.
- [33] C. JAMES, “Retrieval-Augmented Decision Support Systems for Precision Agriculture,” 2024.
- [34] A. Vizniuk, G. Diachenko, I. Laktionov, A. Siwocha, M. Xiao, and J. Smolag, “A comprehensive survey of retrieval-augmented large language models for decision making in agriculture: unsolved problems and research opportunities,” *Journal of Artificial Intelligence and Soft Computing Research*, vol. 15, 2025.
- [35] J. Dieber and S. Kirrane, “Why model why? Assessing the strengths and limitations of LIME,” arXiv preprint arXiv:2012.00093, 2020.