JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-21, No.-11, November (2026) pp 1-32 ISSN (Print) 0973-8975

AGRISENSE AI: A PRECISION AGRICULTURE PLATFORM LEVERAGING RAG, MULTI-MODEL AI ENSEMBLES AND MULTILINGUAL LOCALIZATION

Sandeep Sharma[1], Ruhi Saxena[2], Navin Garg[3], Vanita Kshirsagar[4], Bipin Sule[5], Tanveer Ahmad Wani[6], Ashish Rayal[7]

[1]School of Computer Science and Engineering, Lovely Professional University (LPU), Jalandhar-Delhi G.T. Road, Phagwara, Punjab 144411, India.

[2]School of Engineering & Technology, CGC University, Mohali, Punjab 140307, India.

[3]Department of Computer Science Engineering, Graphic Era Deemed University, Dehradun, Uttarakhand 248002, India.

[4]Department of Computer Engineering, MIT Academy of Engineering Alandi, Pune, Maharashtra 412105, India.

[5]Department of Engineering, Science and Humanities, Vishwakarma Institute of Technology, Pune, Maharashtra, 411037, India.

[6]School of Sciences, Noida International University, Uttar Pradesh 203201, India.

[7]Department of Mathematics, School of Applied and Life Sciences, Uttaranchal University, Dehradun, Uttarakhand 248007, India.

Email : [1]sanintel123@gmail.com , [2]ruhi.saxena2011@gmail.com , [3]navin.garg6@gmail.com, [4]vanita.kshirsagar@gmail.com , [5]bipin.sule@vit.edu , [6]tanveer.ahmad@niu.edu.in , [7]ashish1989rayal@gmail.com

Corresponding Author: **Sandeep Sharma**

https://doi.org/10.26782/jmcms.2026.11.00001

(Received: August 13, 2026; Revised: October 29, 2026; Accepted: November 12, 2026)

Abstract:

Indian agriculture has a high level of spatial, climatic, soil, market, and linguistic diversity, which has a strong impact on agricultural decision-making. AgriSense AI is an integrated precision agriculture decision support platform that integrates Retrieval-Augmented Generation (RAG), real-time weather intelligence, agronomic reasoning, financial analytics, multilingual generation, conversational assistance, and image-based crop-health analysis. It fetches historical knowledge on agricultural topics from a Qdrant vector database and fuses it with Gemini 1.5 Flash-based weather intelligence and GPT-4o-mini to provide structured recommendations and constrained agronomic reasoning. AgriSense AI does not give one set of

recommendations, but rather four different sets of recommendations based on climate fit, strategic opportunity, historical performance, and diversification. The recommendations include "actionable" estimates of NPK requirements, water usage, investment, expected profit, and return on investment (ROI). The multilingual module generates the advice directly in the selected language of India without any post hoc translation. In the controlled evaluation reported, the LLM with RAG-based models obtained around 87% accuracy in yielding the same responses as "ground-truth" data with less than 6% hallucination rates, versus 55% for the vanilla LLM, and less than 28% for the vanilla LLM. State-wise accuracy was 80 to 92%, and overall scores for the multilingual evaluation by the agricultural extension workers were in the range of 4.4–4.8 out of 5. The study covers the architecture, the retrieval and the recommendation mechanism, the evaluation results, case studies, limitations and future directions.

Keywords: Precision Agriculture, Retrieval-Augmented Generation, Large Language Models, Crop Recommend, Multilingual NLP, Agricultural Decision Support, Plant Disease Detection, Weather Intelligence, Qdrant, Generative AI.

I. Introduction

Agriculture is a high-stakes domain of decisions where outcomes of an incorrect recommendation are linked directly to household income, water use, environmental sustainability, and food security. The document supplied by AgriSense AI has also made it clear that the suggested platform is specifically intended for Indian farmers and that the decision about crops cannot be made only at the national or state level due to the wide disparity in soils, rainfall, crop calendars, and irrigation systems in India and market conditions. The source outlines an architecture that integrates historical records from agriculture and meteorological data in real time with reasoning from language models to generate field-specific recommendations. It agrees with the general literature on precision agriculture, where the agricultural decision-making process is viewed as data-driven, using heterogeneous observations, temporal variability, and spatial context [XXX], [XI].

The need for localized decision support is particularly important because the suitability of a crop depends on interactions among soil characteristics, nutrient availability, rainfall, temperature, humidity, water access, previous crop, and cultivation season. Machine-learning studies have repeatedly used temperature, rainfall, and soil variables as important predictors for crop-yield and crop-suitability tasks [XXVIII], [XVI], [XXVII]. However, a conventional classifier normally returns a class label rather than a complete decision package. A farmer does not only need to know which crop is likely to be suitable; the farmer may also need to understand the expected investment, resource requirements, irrigation burden, expected return, alternatives, and reasons behind the recommendation. The AgriSense design therefore extends prediction into decision support by presenting a ranked portfolio of four alternatives and by attaching resource and economic indicators to each alternative.

The second challenge is the rapid adoption of large language models. Despite the ability to generate natural language explanations, heterogeneous instructions, and the ability to combine multiple instructions, the parametric knowledge that LLM models possess is not sufficient in place of a verified agricultural database. Lewis et al. proposed the

RAG framework, which can be used to integrate parametric generation with non-parametric retrieval, enabling a language model to leverage retrieved evidence for its response [XXIII]. This difference is significant in agriculture because the yields of the crops and values of rainfall can change from year to year and from place to place, and fertilizer recommendations and local production statistics can change. AgriSense AI thus performs semantic retrieval from a vector database and then calls the reasoning model. The source uses 5 records as top-k retrieval and asks the LLM to consider the retrieved records as context for a quantitative recommendation.

A third challenge is the emergence of tool-using and agentic LLM architectures. ReAct showed that language models can combine reasoning with actions against external knowledge sources, while Toolformer demonstrated that language models can learn to invoke external tools through APIs [I], [VII]. AgriSense AI applies the same architectural principle at the system level rather than relying on a single model. A weather-specific component supplies current meteorological context; a vector retrieval component supplies historical evidence, and an agronomic LLM synthesizes the information. The resulting architecture is therefore better characterized as a multi-source AI ensemble than as a single predictive model. Separates concerns for maintainability: weather is data that can be changed without affecting historical agricultural information or language reasoning. Linguistic inclusion is the fourth challenge. The agricultural information must be understandable to the farmers, and they should be able to put it into practice; India is a multilingual environment. Previous works on multilingual language-model research have shown that the quality of language representations significantly influences the performance of the models, particularly for languages that are less common in general multilingual models [XII] - [XXI].

Indian-language resources were developed precisely because generic multilingual models can underrepresent Indian linguistic and transliteration patterns. The source document specifies a strategy where the LLM is clearly tasked with generating the recommendation, justification, quantities, and next steps in the target script. This differs from the concepts of generating in English and then translating the results of the generation, in that the generation is itself considering local syntax, terminology, and scripts.

The fifth challenge is the transition from advisory text to operational considerations on the farm. The source system features an analytical dashboard with profit and ROI visualizations as well as NPK visualizations and daily-water visualization; a structured next-step panel; a conversational assistant; crop-image upload pathway. The image module is used as an additional diagnostic tool to complement agronomic knowledge. Studies in the area of plant disease detection have demonstrated that convolutional and deep-learning models can be effective on curated datasets and that there is a significant performance gap between models trained on curated and field images (including those collected in the field) [XXIV], [V]. Therefore, the incorporation of image analysis, as in AgriSense AI, is a potentially valuable feature of the system, but should be made available as an assistive screening tool instead of an independent clinical plant diagnosis.

The sixth challenge is about trust and evaluation. The assessment of a recommendation system for agriculture should not only focus on classification accuracy, but should also take into account factual support, hallucination rate, regional consistency, language

quality, precision of the resources, and usability. A controlled comparison between a vanilla LLM and the RAG-grounded AgriSense pipeline is provided, along with measurements of yield agreement and hallucination rates. It also provides results by state and multilingual ratings by agricultural extension workers. These results would be helpful in writing a journal paper, but the manuscript should differentiate between reported system measurements and independently reproduced experiments. The source reports 200 queries in the controlled RAG comparison, with the state-wise table adding up to 216 queries. The discrepancy should be balanced with the original experiment logs before publication. This paper thus has four aims. First, it introduces the concept of the AgriSense AI architecture as an integrated precision-agriculture decision-support pipeline. Second, it introduces a literature review of machine learning applications in agriculture, precision agriculture, RAG, tool-using LLMs, multilingual Indian-language NLP, and plant disease detection based on images. Thirdly, it records the method, data flow, strategy for recommendations, interface modules, and the results of the evaluation reported in the provided system document. Fourthly, it recommends a repeatable evaluation protocol for subsequent releases. The main research argument posited is that a structured, agronomic reasoning, multi-lingual interaction component is more beneficial when integrated into a single auditable workflow than being passed to an unconstrained LLM.

i. Problem Statement

Most of the current agricultural recommendation systems focus on a specific prediction target. A natural language explanatory system might be used to describe farming practices, and a classifier could determine a proper crop based on soil and weather characteristics. The difference between these two is that farmers need local, evidence-based, economically meaningful, and linguistically friendly advice. AgriSense AI fills this gap by considering crop recommendation as a multi-stage decision support problem that consists of retrieval, weather context, agronomic reasoning, resource accounting, financial projections, and localized communication. The study investigates a few key research questions. First, it investigates whether retrieval-grounded generation is able to boost the factual consistency and reliability of agricultural recommendations over a conventional vanilla large language model (LLM). Second, it looks into whether a multi-source architecture that enables the integration of historical agricultural data and real-time weather context can be applied to create more context-aware and geographically specific recommendations for farmers. Third, the study assesses whether a 4-crop portfolio message is more effective in conveying agricultural risk, resource needs, and trade-offs than a single crop label message. Fourth, it explores how much the proposed framework maintains grammatical correctness, technical accuracy, and comprehension while generating sentences in multiple Indian languages. Finally, it investigates the potential of embedding the image-based crop-health analysis into the same decision-support process and ways to appropriately illustrate uncertainties and lack of diagnostic confidence.

ii. Contributions

I. An end-to-end architecture with a formal structure that incorporates RAG, weather intelligence, LLM reasoning, and structured analytics.

II. Four distinct slot recommendation strategies: climate fit, strategic choice, historical performance, and diversification. Designed and evaluated in a native Indian language – multilingual output.

III. A multimodal extension to accept crop images and integrate the visual analysis with a disease knowledge base and advisory workflow. A single evaluation part-quantitative and qualitative results was presented in the given source file.

IV. A reproducibility and limitation analysis that helps to pinpoint inconsistencies that need to be confirmed before publication in a journal.

II. Literature Review

AgriSense AI is not just a single strand of the literature, but rather it is a collection of research streams. Crop-recommendation topic. These streams encompass precision agriculture and digital farming, machine learning, and AI in agriculture. Models for yield and crop prediction, computer vision for plant health, and retrieval-augmented language models. Multilingual language models and farmer-facing conversational systems, generation and agents using tools to work. A review of these streams is then synthesized, and the specific gap addressed by it is identified. AgriSense AI.

i. Precision Agriculture and Digital Farming

Precision agriculture emerged from the need to manage within-field variability rather than applying identical inputs to every part of a farm. Early work emphasized location-aware management and variable-rate decisions, while later work incorporated remote sensing, GNSS, GIS, IoT, cloud computing and farm-management information systems [XIV], [XI], [X], [XVII]. The resulting paradigm is data-rich but also integration-heavy: the value of a sensor, weather feed or satellite image depends on whether the information can be converted into an actionable decision.

Wolfert et al. reviewed big-data applications in smart farming and highlighted both the potential of heterogeneous data and the associated challenges of data ownership, integration, interoperability and business models [XXVIII]. Similarly, Klerkx et al. showed that digital agriculture is not only a technical shift but also alters the relationships between the farmers, the advisors, the technology providers and the institutions[X]. The results for AgriSense AI validate the design decision of separating the various parts of the system, such as data retrieval, weather context, reasoning, and farmer interaction, from a single model. According to machine-learning reviews, supervised learning, deep learning, and hybrid supervised/deep learning techniques have become a growing trend in agriculture for crop monitoring, yield estimation, disease recognition, and resource management, etc.[VI],[XIX], [XVI], [XXVII]. However, good in the box does not necessarily mean good in the field. Distribution shifts can occur due to changes in crop type, location, season, and data acquisition conditions. AgriSense thus relies on retrieval and contextual inputs to suggest rather than infer, based on available local evidence, rather than relying exclusively on a globally trained classifier.

ii. Crop Recommendation and Yield Prediction

Typical variables used in crop recommendation studies include N, P, K, pH, rainfall, temperature, and humidity. Random Forest, support-vector machines, K-nearest

neighbors, neural networks and ensemble methods are frequently compared. van Klompenburg et al. completed a systematic search of 50 crop yield prediction studies, and found that temperature, rainfall, and soil type were among the most commonly used features [XVI]. More recent reviews have also confirmed that machine learning and deep learning are key technologies for making predictions of yield, pointing to the importance of geographical variation, data quality, and temporal generalization [XXVII], [III], [II]. One drawback of traditional recommendation systems is that the system cannot explain the recommendations. Not every model will generate a crop label, but not every model will give a traceability pathway for quantitative claims. AgriSense overcomes this with RAG in that it fetches historical records before generation. The retrieved fields include crops, year, season, area, production, yield, and rainfall, with a reference to the source system. This design aims to take a black-box recommendation of a text-only response and provide a response with an evidence layer. More importantly, agricultural decision-making has several objectives. Maximizing expected yield may not coincide with minimizing water demand, investment/risk requirements. AgriSense can therefore be considered a multi-objective recommendation interface, with the 4-crop portfolio. Crop 1 emphasizes current climate fit; Crop 2 emphasizes strategic and economic opportunity; Crop 3 emphasizes historical regional performance; and Crop 4 provides diversification. The portfolio does not mathematically solve a formal stochastic optimization problem, but it translates multiple decision criteria in a way that is comprehensible for farmers.

iii. Generative AI, RAG and Agentic Systems

The release of powerful LLMs has changed the design space for agricultural advisory systems because natural-language interaction can reduce the need for rigid menus and allow farmers or extension workers to ask open-ended questions. Yet, LLMs can generate facts/numbers that never happened, or causal explanations that may be incorrect. In order to address this issue, RAG was recently proposed as a way to use neural generation in conjunction with external, explicit memory [XXIII]. The system can draw out relevant records and limit quantitative statements to records, instead of requesting the model to remember all agricultural statistics from the parameters. ReAct and Toolformer further take this concept of retrieval to actions and tools [I], [VII]. A weather agent can be considered an external tool whose output is inserted into the reasoning context. This is especially relevant for agriculture since the information on the weather and markets evolves at a much faster rate than the information used to train the models. The AgriSense architecture is tool-augmented, where the weather intelligence is separated from the retrieval of the historical data, and the LLM is used mainly as a synthesis and explanation layer. According to the source document, the results for the AgriSense pipeline showed a 28% to 6% relative decrease in hallucination rate compared to the vanilla-LLM baseline. This is in a similar direction to the motivation for RAG, but it is important to realize that this is an assessment of RAG as a system, not as a universal property of RAG. The condition of a retrieval and constraints on the prompt, along with definitions of the evaluation and domain knowledge base, all affect the likelihood of hallucination. Retrieval precision/recall, citation coverage, unsupported-claim rate, and answer faithfulness should then be reported separately in future experiments.

iv. Multilingual Indian-Language NLP

Multilingual NLP is central to agricultural accessibility in India. BERT established a strong transformer-based language representation paradigm[XII], while XLM-R demonstrated that large-scale multilingual pretraining can improve cross-lingual transfer [XXI]. However, broad multilingual models may allocate limited capacity to Indian languages. The IndicNLP suite and MuRIL were designed to enhance the representations for Indian languages and the transliterated text [XV], [VIII]. These developments call for language-aware prompting and assessment, not just as a glaze on top of the integration of translation. AgriSense design is based upon direct generation in the chosen language and script. This may save a separate translation step, but will create an evaluation need: It will be necessary to evaluate the quality of the language in addition to the underlying recommendation of the crop. The document source assesses the grammar, technical accuracy, cultural fit, and farmer understanding via agricultural extension workers, and overall scores range from 4.4 to 4.8 on a 5-point scale for the recorded language set. Though a useful human-centered measure, a future study should broaden the range of evaluators and incorporate direct farmer testing.

v. Plant Disease Detection and Multimodal Agriculture

Computer vision is a key element of digital agriculture. Large-scale experiments with leaf-image disease recognition were made possible with Plant Village and related datasets [XXIV]. Mohanty et al. could achieve very high accuracy in the controlled images and also a significant drop in accuracy for more realistic images[XXI]. This finding was subsequently proven with Ferentinos' research and others showing excellent CNN performance on various diseases of crops[XX], [V]. The outcomes provide an encouraging and cautionary example that image-based plant diagnosis holds great promise, but also poses a significant deployment risk: laboratory-based datasets can be more effective than field-based datasets. AgriSense AI has an image upload assistive module. The provided document outlines a pathway for visual analysis that involves multiple classes, a disease knowledge base, and a structured report with a predicted condition, confidence, affected area estimate, and treatment guidance. The system should clearly distinguish between visual classification confidence and agronomic treatment confidence to ensure it is journal-ready and offer a referral pathway if images are not of a good quality, or if there is doubt about the class predicted.

vi. Farmer-Facing Conversational Systems

AI-driven conversational systems have emerged in recent years that demonstrate their potential to support farmers alongside traditional extension services. Farmer. For instance, Chat is an AI-based advisory service for smallholder farmers that was designed as a generative approach to farming. Chat's core focus is on providing personalized, reliable, and contextually relevant advice for smallholder farmers. These systems validate the notion of a conversational agricultural assistant, but they also highlight the importance of evidence grounding, localized language, and human-centered evaluation. With its conversational interaction, a structured farm-analysis module, quantitative resource analytics, a four-crop portfolio, historical references, and image-based crop-health support, AgriSense stands out from the crowd. The literature,

as a result, then, indicates that agricultural AI systems that are going to prove most useful will probably not be single-task classifiers. Rather, they will be more likely to be multi-modal decision support systems with data retrieval, predictive modeling, language interaction, visualization, and human supervision. AgriSense AI is built on this holistic approach. The reviewed literature shows that there is considerable advancement in agricultural prediction, smart farming, plant disease recognition, and multilingual NLP, but all these are developed as individual systems. So, the big challenge tackled by AgriSense AI is integrating the building blocks. A typical crop classification system will not normally retrieve historical evidence, call on a real weather agent, create a multilingual story, do farm-level economics, or keep the history of the interaction in one workflow. On the other hand, a general-purpose chatbot can give fluent advice, but it may not have an explicit evidence layer. AgriSense is intended to link these features. Table 1 briefly explains the review of Literature informing the AgriSense design as:

Table 1: Review of literature directly informing the AgriSense design

Ref.	Study	Topic	Relevance to AgriSense
[I]	Achim et al. (2023)	GPT-4 Technical Report	Large multimodal language-model capabilities
[II]	Conneau et al. (2020)	Unsupervised Cross-lingual Representation Learning at Scale	XLM-R multilingual representation
[III]	Devlin et al. (2019)	BERT: Pre-training of Deep Bidirectional Transformers	Transformer language representation
[IV]	Ferentinos (2018)	Deep learning models for plant disease detection and diagnosis	Deep learning across plant diseases
[V]	Fountas et al. (2015)	Farm management information systems	Digital farm-management information
[VI]	Gebbers & Adamchuk (2010)	Precision agriculture and food security	Site-specific management and food security
[VII]	Team et al. (2024)	Gemini 1.5 Technical Report	Long-context and multimodal model capabilities
[VIII]	Hu et al. (2020)	XTREME: A Massively Multilingual Multi-task Benchmark	Cross-lingual evaluation
[IX]	Hughes & Salathé (2015)	An open access repository of images on plant health	PlantVillage image resource
[X]	Jabed & Murad (2024)	Crop yield prediction in agriculture: A comprehensive review	Recent ML/DL yield-prediction evidence

[XI]	Jha et al. (2019)	A comprehensive review on automation in agriculture using AI	AI, IoT and automation in agriculture
[XII]	Kakwani et al. (2020)	IndicNLPSuite	Indian-language corpora and benchmarks
[XIII]	Kallenberg et al. (2023)	Integrating process-based models and machine learning for crop yield prediction	Hybrid modeling and domain knowledge
[XIV]	Kamilaris & Prenafeta-Boldú (2018)	Deep learning in agriculture: A survey	Deep learning across agricultural tasks
[XV]	Khanuja et al. (2021)	MuRIL: Multilingual Representations for Indian Languages	Indian-language representation learning
[XVI]	Klerkx, Jakku & Labarthe (2019)	Digital agriculture, smart farming and Agriculture 4.0	Social and institutional dimensions
[XVII]	Lewis et al. (2020)	Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks	Grounded generation with external memory
[XVIII]	Li et al. (2023)	Label-Efficient Learning in Agriculture: A Comprehensive Review	Data-labeling constraints in agricultural ML
[XIX]	Liakos et al. (2018)	Machine Learning in Agriculture: A Review	ML applications and challenges
[XX]	Lowenberg-DeBoer & Erickson (2019)	Setting the Record Straight about On-Farm Precision Agriculture Adoption	Economics and adoption of precision ag
[XXI]	Mohanty, Hughes & Salathé (2016)	Using Deep Learning for Image-Based Plant Disease Detection	CNN plant disease recognition
[XXII]	Pathak et al. (2023)	Predicting Crop Yield With Machine Learning	Multimodal yield prediction
[XXIII]	Schick et al. (2023)	Toolformer: Language Models Can Teach Themselves to Use Tools	API/tool use by LMs
[XXIV]	Singh et al. (2023)	FarmerChat / agricultural advisory services	Conversational agricultural support
[XXV]	Sladojevic et al. (2016)	Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification	Leaf-image disease classification
[XXVI]	Too et al. (2019)	A comparative study of fine-tuning deep learning models for plant disease identification	Transfer learning for disease detection
[XXVII]	van Klompenburg et al. (2020)	Crop yield prediction using machine learning: A systematic literature review	Features and algorithms in yield prediction

[XXVIII]	Wolfert et al. (2017)	Big Data in Smart Farming – A review	Data integration and smart-farming challenges
[XXIX]	Yao et al. (2023)	ReAct: Synergizing Reasoning and Acting in Language Models	Reasoning plus external actions/tools
[XXX]	Zhang, Wang & Wang (2002)	Precision agriculture	Foundational precision-agriculture framework

III. Research Gap and Proposed Research Objectives

The literature they have reviewed shows significant advances in agricultural prediction, smart farming, plant disease recognition, and multilingual NLP, which are usually implemented in independent systems. Thus, the missing link that AgriSense AI rectifies is the integration of the architecture. A typical crop classifier doesn't typically gather historical evidence, call upon a live weather agent, generate a multilingual story, compute farm-scale economics, and keep the interaction history in a single workflow. On the other hand, an all-purpose bot can give articulate guidance without a well-defined evidence layer. AgriSense is conceived as a strategy to link these capabilities.

i. Research Objectives

- Create a retrieval-based agriculture recommendations pipeline for India.
- Combine weather information in real time with previous crop performance data.
- Instead of a single and deterministic crop label, generate four crop options that are differentiated.
- Generate recommendations at the farmer's acreage scale as resource and economic measurements.
- Offer native multilingual generation and human-centered quality evaluation.
- Embed image-based crop health diagnostics and link with a knowledge base for advisory.
- Assess the system with accuracy, hallucination, state-wise consistency, and multilingual, qualitative usability metrics.

ii. Design Principles

Evidence before explanation: findings and quantitative claims should be supported by retrieved evidence wherever possible. Recommendation based on context: location, season, soil, and water availability should be taken into account. Multiple alternatives, portfolio before prescription, decrease reliance on a single crop and reveal trade-offs. Before translation, it is advisable to create advice in the selected script, if available. Before autonomy, there must be human oversight, with appropriate mechanisms of uncertainty and referral for image diagnosis and chemical-treatment suggestions.

IV. System Architecture

The AgriSense AI architecture is a modular Web-based decision-support system. Frontend: Next.js 15 and React 19; Reasoning model: GPT-4o-mini; Weather

model: Gemini 1.5 Flash; Vector database: Qdrant; Embeddings model: text-embedding-3-small; Interactive analytics: Recharts. The architecture is broken up in a deliberate way to allow the retrieval, weather, reasoning, and response normalization, and user interface layers to be tested independently. Figure 1 shows the workflow of the proposed multi-source agricultural recommendation framework as:

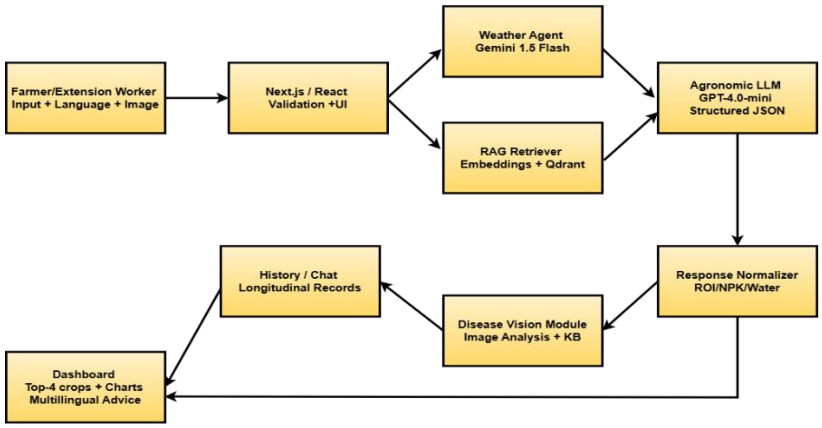


Fig. 1. Workflow of the proposed multi-source agricultural recommendation framework

Table 2 summarizes the major technological components used to implement the proposed agricultural recommendation system, along with their primary roles and the rationale for their selection. The Frontend, developed using Next.js 15 and React 19, provides the main user interface, routing, rendering, and API communication, enabling a responsive and modular interaction environment. Tailwind CSS with Glassmorphism-based styling is used to provide a consistent and responsive visual interface while supporting rapid UI development.

Table 2: System components, technologies, roles, and design rationale

Component	Technology	Primary role	Design rationale
Frontend	Next.js 15 + React 19	Rendering, routing, and API calls	Responsive web application and modular UI
Styling	Tailwind CSS + Glassmorphism	Responsive interface and modes	Rapid UI development
Primary LLM	GPT-4o-mini	Agronomic reasoning and structured response	Cost-aware language generation
Weather agent	Gemini 1.5 Flash	Weather and meteorological context	Dedicated context provider
Vector database	Qdrant	Similarity retrieval of agricultural records	Fast semantic retrieval
Embeddings	text-embedding-3-small	Query/document vectorization	Semantic retrieval layer
Analytics	Recharts	ROI, profit, NPK and water charts	Interactive decision support
History	Application storage	Past chats and analyses	Longitudinal user record

i. Data Flow

The processing sequence starts with the structured inputs of the farm: location, land size, budget, soil type, water source, season, and previous crop (if applicable). frontend validates the form fields and sends a JSON payload to the backend. The backend generates a semantic representation of the query, fetches the most relevant historical

records, fetches the weather context, constructs a combined context package, then calls the reasoning model. The response normalizer then validates JSON schema, calculates and/or validates financial elements, and returns the structured response to the dashboard. This design does not let an unfinished LLM answer directly influence the UI.

ii. Retrieval-Augmented Generation Layer

The RAG layer represents the actual content of the recommendation workflow. Historical agricultural records are stored and embedded in Qdrant. The system then infers a semantic query from the location and season descriptors, and extracts the top 5 records from the data by similarity search, adding them to the LLM context for a new query as shown in Figure 2. The retrieved evidence is explicitly directed towards grounding yield, NPK, and profit values in the source document. This is consistent with the original RAG motivation: external non-parametric memory can contain additional evidence and provenance to be updated beyond those contained within the model's parameters [12]

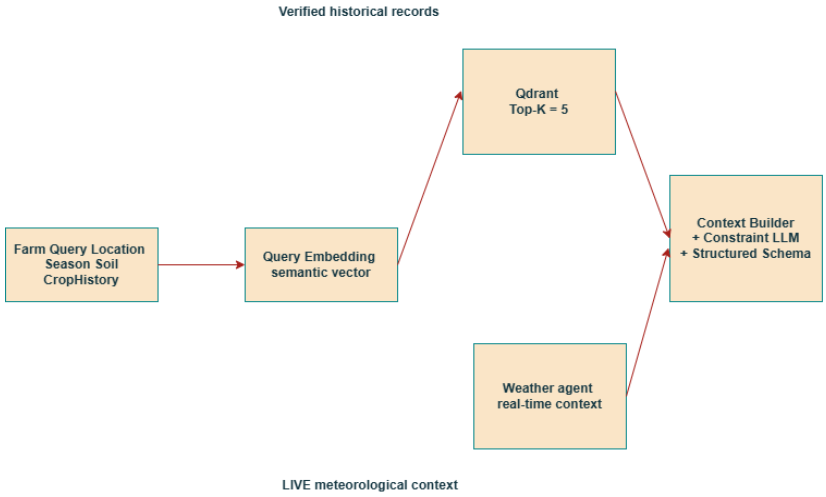


Fig. 2. RAG pipeline used to ground crop recommendations in historical agricultural evidence.

iii. Real-Time Weather Agent

The weather component takes a city name or geographic coordinates as input, and provides current temperature ranges, contextual information about the expected precipitation, soil-moisture estimates based on recent precipitation, and monsoon timing information, as outlined in the source document. The weather result is not considered a replacement for historical data, but rather it is a complementary context stream. The architecture thus divides the persistent evidence from the ephemeral environmental information.

iv. Multi-Source Ensemble

Table 3 presents the major information sources used by the proposed agricultural recommendation system and explains their respective roles in generating context-aware recommendations. Historical RAG provides crop- and region-specific historical

evidence, including crop, year, season, cultivated area, production, yield, and rainfall records. This information helps the system identify historical agricultural patterns and regional crop performance. The Weather Agent supplies current meteorological context, including temperature, rainfall, humidity, and monsoon conditions, which supports the assessment of climate suitability for different crops. The Agronomic LLM combines the retrieved historical evidence and weather context with agronomic reasoning, including crop rotation and market-related considerations, to generate strategic recommendations and natural-language explanations. Thus, the three sources work together to provide evidence-based, weather-aware, and agronomically reasoned agricultural recommendations.

Table 3: Data sources and their roles in the agricultural recommendation system

Source	Information	Used primarily for
Historical RAG	Crop, year, season, area, production, yield and rainfall records	Evidence and regional performance
Weather agent	Temperature, rainfall, humidity and monsoon context	Climate-fit recommendation
Agronomic LLM	Crop rotation, market reasoning and natural-language synthesis	Strategic recommendation and explanation

v. Four-Crop Portfolio Logic

The system is designed to create four proposals. The first is a climate-fit crop, based mainly on the current and forecasted climate context. The second is a strategic selection based on crop rotation, market opportunity, and agronomic considerations. The third is a historical champion, which was chosen based on good regional historical performance. The fourth is a diversification option from a different crop family or secondary historical performer. The four slots are not four separate classifiers; they are decision roles that foster comparison to different objectives. The decision logic for the four-crop recommendation portfolio is shown in Figure 3.

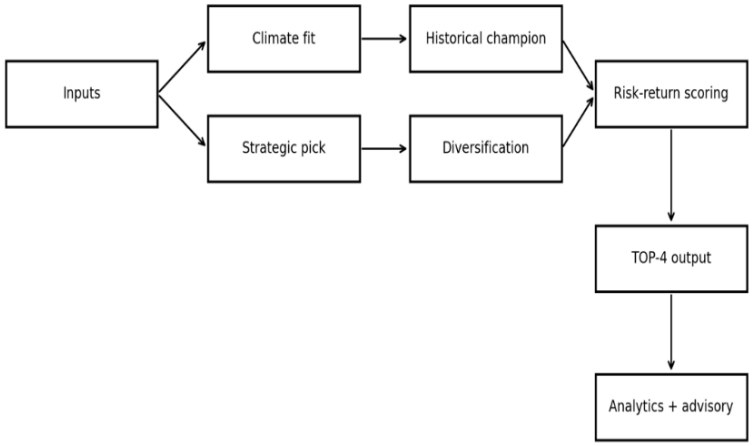


Fig. 3. Decision logic for the four-crop recommendation portfolio.

VI. Methodology

The following section is a formalization of the process that was implemented and described in the file provided. This methodology is not described as a newly trained machine-learning model, but rather as a system-engineering and evaluation framework. If measured values have been provided by the source, they appear as source-reported values. If the source does not report an independently documented experiment, and only reports implementation behavior, then the text is not considered a scientific claim, but rather a system specification.

i. Input Representation

Table 4 provides the key farmer inputs and the definition and application of each. Location or region: free text or city name, used to establish a geographical context for weather retrieval and record. The size of the land, entered numerically in acres, determines cultivation size and the amount of resources that can be cultivated. Budget is a financial-feasibility constraint (entered in Indian Rupees (INR)). Crop–soil compatibility is assessed using soil type (8 categorical types). Selected from seven categorical options, water source shows the feasibility of irrigation and water supply. The crop-calendar context is given by season, which can be Kharif, Rabi, Zaid, or year-round. Previous crop is used to provide information about rotation and soil history (optional text). Combined, these inputs provide the geographical, financial, environmental, and historical context necessary for personalized and feasible recommendations.

Table 4. Farmer input parameters and their representation in the proposed agricultural recommendation system

Parameter	Representation	Purpose
Location/region	Free text or city	Geographical context and weather retrieval
Land size	Numeric acres	Scales resource and financial quantities
Budget	Numeric INR	Feasibility constraint
Soil type	8 categorical options	Crop-soil compatibility
Water source	7 categorical options	Irrigation feasibility
Season	Kharif / Rabi / Zaid / year-round	Crop calendar
Previous crop	Optional text	Rotation and soil-history context

ii. Data Preparation and Retrieval

The primary historical dataset used is Crop Yield in India, a public dataset. These records include information for the crop, year, season, region, cultivated area, production, yield, and rainfall. Every structured record di is mapped to a fixed text template τd from the form (Crop, Season, State, Year, Area, Production, Yield, Annual rainfall), and each such fixed text template is embedded in a 1536-dimensional vector, using the text-embedding-3-small embedding model. For each di, a 1536-dimensional

vector is created, containing the embedding of a fixed text template τd which is derived from the record di, using the text-embedding-3-small embedding model. Having identical attributes stored as structured payload fields in Qdrant alongside models allows for the description of numerical values with greater strength. These fields of payload are only used for filtering and are fed to the LLM as is, without ever being recovered from the embedding itself. Each farmer's question is encapsulated in the same format: The farmer's profile is turned into a brief description of the season, location, soil, water source, previous crop, and land size, resulting in τq. With E(·) denoting the embedding function and sim the cosine similarity, the query and record vectors and their similarity are given by the equation as:

$$\mathbf{e}_q = E(\tau_q), \quad \mathbf{e}_d = E(\tau_d), \qquad \mathrm{sim}(\mathbf{e}_q, \mathbf{e}_d) = \frac{\mathbf{e}_q \cdot \mathbf{e}_d}{\| \mathbf{e}_q \| \| \mathbf{e}_d \|}$$

However, similarity in the embedding space is not enough for agronomic suitability: a record can be close to the query in the embedding space but be about a different season, a far-away region, or an old period. The retrieval is therefore limited in two steps as outlined in Table 5. The candidate set is constrained to those that have the same season as the desired set S(q), fall into the desired geographic set G(q), and do not have a year older than some recency threshold ymin, as in Equation:

$$\mathcal{D}_F(q) = \{\, d_i \in \mathcal{D} \colon \mathrm{season}(d_i) \in S(q) \ \wedge \ \mathrm{region}(d_i) \in G(q) \ \wedge \ \mathrm{year}(d_i) \geq y_{\min} \,\}$$

G(q) is the farmer's state, which is expanded to the agro-climatic zone around the farmer only if fewer than k records match the filter. Second, a rule-based compatibility screen C(di, f) ∈ {0, 1} based on published crop-requirement guidelines is used to eliminate records of crops not compatible with the farm profile. The top-k numbers, ranked according to cosine similarity (as in equation (3)), are then retrieved.

$$R(q) = \underset{d_i \in \mathcal{D}_F(q),\, C(d_i, f)=1}{\mathrm{TopK}} \mathrm{sim}(\mathbf{e}_q, \mathbf{e}_{d_i})$$

Table 5 indicates the variables included and excluded from the index. The indexed corpus includes no soil-nutrient (NPK) availability, pH, irrigation, or crop rotation attributes. Nutrient suitability can therefore not be determined from retrieval. The variables are fed into the pipeline on the farmer's inputs and in the reasoning phase, and the system associates any unknown value not resulting from a retrieved record with a model inference. The different variables from the embedding space are not weighted: They are implicit in the embedding model, and this's what makes season, region, and recency hard filters instead of relying on semantic similarity.

Table 5. Feature composition of the retrieval layer: how each variable group is represented and used

Variable group	Examples	Representation and use in retrieval	Stage
Season	Kharif, Rabi, Zaid, year-round	Payload field; hard filter S(q)	Pre-retrieval filter
Geography	State or district of the record and of the farm	Payload field; hard filter G(q), widened to the agro-climatic zone only if fewer than k records remain	Pre-retrieval filter
Recency	Record year	Payload field; threshold y_{min}	Pre-retrieval filter

Variable group	Examples	Representation and use in retrieval	Stage
Crop performance	Crop, area, production, yield, annual rainfall	Serialized into the record text and embedded; also stored as payload and passed verbatim to the LLM	Semantic ranking
Farm profile	Soil type, water source, previous crop, land size, budget	Serialized into the query text; soil, water, and rotation also drive the compatibility screen $C(d_i, f)$	Semantic ranking and post-filter screen
Not indexed	NPK availability, pH, irrigation availability, local microclimate	Absent from the corpus; supplied by farmer input or model inference and labeled as such	Reasoning stage only

The default value isk = 5. This has been selected to maintain a compact context. The sensitivity analysis described in Section VII tests the sufficiency of evidence coverage and stability of retrieval quality by running the analysis over different states and seasons, and across different values of k ({3, 5, 10}). A production-ready version should also include the ability to store data provenance, publication year, geographic granularity, as well as an update timestamp for each record to ensure that stale and incompatible records can be removed before retrieval.

iii. Context Construction and Prompt Constraints

The context builder combines the farm profile, the retrieved records, and the weather output. The LLM is asked to output structured JSON and to use the retrieved evidence on all quantitative values. This is a constraint, not a fact. The response normalizer validates the necessary fields and stops any malformed responses from reaching the dashboard. There are a total of 12 output fields for each crop: crop name and rank, yield, investment, profit, ROI, margin analysis, recommendation reason, NPK, daily water, fertilizer mix, alternatives, and historical reference.

iv. Recommendation Scoring

A limited multi-objective score can be used to represent the four role strategy. The suitability score of a candidate crop c is given by Equation as:

$$S(c) = w_s S_{\text{soil}} + w_w S_{\text{water}} + w_t S_{\text{weather}} + w_h S_{\text{history}} + w_r S_{\text{rotation}} + w_e S_{\text{economic}}$$

S_{soil}, S_{water}, $S_{weather}$, $S_{history}$, $S_{rotation}$, and $S_{economic}$ are illustrative suitability scores ranging from 0 to 1 for soil, water, weather, history, rotation, and economic factors; weights will be determined by the recommendation slot. The system does not employ numerical weights like these; the above Equation is a formal representation of the design logic and not a description of an implemented scoring function. It is used to clarify that the selection and sequencing of candidates is not part of the design.

v. Financial and Resource Analytics

ROI is calculated as in the Equation as:

$$\text{ROI} = \frac{R-I}{I} \times 100 = \frac{\Pi}{I} \times 100$$

The gross return is denoted by R, the total investment by I, and the expected net profit (return) is denoted by $\Pi = R - I$.

The system also provides a margin analysis. The required quantities of the resources are proportional to the area of the farm; otherwise, the per-hectare quantities could not be directly given to the farmer who is growing a few acres. Profit potential, ROI, NPK demand, and water consumption per day are shown on the dashboard.

vi. Multilingual Generation

The language selected is explicitly given to the language module as a generation constraint. The primary language is English, with Hindi, Punjabi, Marathi, Kannada, Telugu, Tamil, Bengali, and Gujarati support for native generation. These eight Indian languages with English are the evaluated languages, for a total of nine languages. The language prompt includes explanations, instructions, technical terms, and further directions.

vii. Image-Based Crop Health Analysis

The image pathway allows for JPEG, PNG, and HEIC files within the limits of the image interface defined in the source. The visual analysis part categorizes the common patterns of disease, pests, nutrient deficiency, and environmental stress, fetches the disease knowledge, and presents a structured advisory. This pathway should be considered as a multi-modal decision-support module, as the literature shows that the distribution shift in the field can significantly impact performance compared with controlled test data. The Integrated multimodal pathway connecting text, structured farm data and crop images to a unified advisory is shown in Figure 4

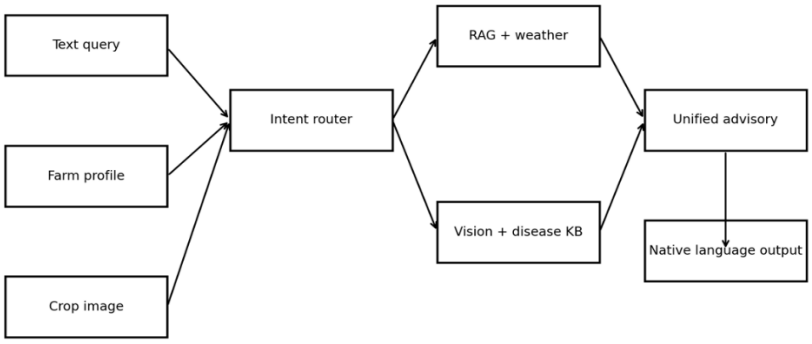


Fig. 4. Integrated multimodal pathway connecting text, structured farm data and crop images to a unified advisory.

viii. Conversational Assistant and History

The chat module enables a free-form question to be asked regarding crop selection, water-use, pest management, and government programs. The history module records previous chat history and structured analysis for later revisiting by the farmer without re-entry of farm parameters. The longitudinal design facilitates comparison across seasons and allows for future human review and audit trail.

V. System Modules and Implementation

i. Farm Details Module

The input module is designed specifically with seven input parameters. There are 8 soil types, 7 water-source types, and 4 seasons described in the source as shown in Table 6. These options are presented categorically, giving a controlled vocabulary to enhance consistency of retrieval and recommendation. Free-text location and previous-crop fields allow flexibility for users who have to describe conditions in their location that are not covered by a fixed taxonomy.

Table 6. Environmental and crop-history parameters used for agricultural decision-making

Category	Options/examples	Decision relevance
Soil	Black/Regur, red, alluvial, sandy, laterite, clay, loamy, sandy loam	Water retention, drainage, nutrient behavior and crop fit
Water	Borewell, canal, rain-fed, river/stream, drip, sprinkler, tank/pond	Irrigation reliability and cost
Season	Kharif, Rabi, Zaid, year-round	Crop calendar and climate compatibility
Previous crop	Rice, wheat, sugarcane, ragi, etc.	Rotation and nutrient-cycle context

ii. Backend Processing Workflow

Table 7 presents the sequential backend workflow used to transform farmer-provided information into personalized agricultural recommendations. The process begins with the User Interface, where farm details are collected and validated. The Backend API then receives and preprocesses the input data in JSON format. The Embedding Generator converts the user query into a semantic representation, which is used by Qdrant vector search to retrieve relevant agricultural records. A Context Builder combines the retrieved evidence with farmer, farm, and weather information to create a contextual input for the LLM reasoning module. The LLM generates the top-four crop recommendations in a structured JSON format. The Response Normalizer validates the generated schema and checks important financial and resource-related fields before the final results are delivered to the frontend. Finally, the Frontend Display presents the recommendations through cards, charts, detailed crop information, and actionable next steps.

Table 7: Backend processing workflow of the proposed agricultural recommendation system

Step	Module	Operation
1	User interface	Collect and validate farm details
2	Backend API	Receive, sanitize, and preprocess JSON
3	Embedding generator	Convert query into a semantic representation
4	Qdrant vector search	Retrieve relevant agricultural records
5	Context builder	Combine evidence with farm and weather context

6	LLM reasoning	Generate top-four recommendations in structured JSON
7	Response normalizer	Validate schema and financial/resource fields
8	Frontend display	Render cards, charts, deep dives, and next steps

iii. Output Schema

Table 8 defines the principal output fields generated by the proposed agricultural recommendation system and their purpose in communicating crop recommendations to users. Crop name and rank establish the priority order of the recommended crops, while expected yield provides an estimated production quantity. Budget/investment represents the anticipated cultivation expenditure, and profit indicates the expected net return. ROI and margin analysis provide financial indicators for evaluating the economic attractiveness of each recommendation. The recommendation reason provides a localized explanation of why a particular crop is suitable for the farmer's conditions. Resource requirements are represented through NPK requirements, daily water demand, and the recommended fertilizer mix, including urea, DAP, MOP, gypsum, and zinc sulphate where applicable. Potential alternatives support crop substitution and diversification, while the historical reference identifies the evidence record used to ground the recommendation. Overall, the output schema ensures that recommendations are presented in a structured, interpretable, financially informative, resource-aware, and evidence-grounded format.

Table 8: Output schema of the proposed agricultural recommendation system

Field	Purpose
Crop name + rank	Priority order
Expected yield	Production estimate
Budget/investment	Cultivation cost
Profit	Expected net return
ROI	Return relative to investment
Margin analysis	Return-plus-investment ratio as reported
Recommendation reason	Localized explanation
NPK	N, P and K requirement
Daily water	Irrigation demand
Fertilizer mix	Urea, DAP, MOP, gypsum and zinc sulphate quantities
Potential alternatives	Substitution and diversification options
Historical reference	Evidence record used for grounding

iv. Dashboard and Visual Analytics

The dashboard is created to make comparisons at the farmer's level. A profit-potential chart reveals the variability between crops, an ROI chart is used to calculate economic return on investment, an NPK chart is used to guide procurement planning, and a water chart is used to guide irrigation planning. Per-crop deep dives also feature a radar-style resource visualization, fertilizer-composition pie chart, resource checklist, and alternatives. The source also has a five-step action panel on soil testing, procurement of seed, water management, budget planning, and monitoring of weather. The experimental setup and the evaluation are presented in the section below.

VI. Experimental Setup and Evaluation.

The initial assessment focused on the full AgriSense AI workflow compared to a baseline GPT-4o model, and evaluated them on farmer-oriented queries, with the ground truth being the agricultural records in India. The measures were: yield agreement, hallucination rate, regional specificity, NPK precision, market-price timeliness, and language fidelity. The results of those computations are recorded here. There are, of course, two points of comparison which must be put forward. First, the pipeline is not only different in several parts of it at once (retrieval, weather context, prompting, and response normalization), but it is also fairly complex to implement. Second, the baseline model (GPT-4o) is distinct from the model used to reason the pipeline (GPT-4o-mini). Thus, reported gains are not necessarily due to retrieval. The basic metrics are formally defined in the following subsections; a controlled ablation is provided that sets the model, the decoding parameters, the query and the output constraints, and the retrieval quality, and the numerical outputs are validated.

Metric Definitions

Let Q = {q1, …, qN} be the evaluation queries, and not exact equality but yield agreement is defined by an explicit error criterion. Let ŷj be the yield for the best crop of a particular crop type qj and yj be the reference yield for the same crop, region, and season. The query is considered to agree if the absolute relative error is less than a tolerance τ, i.e., |a_{q}-a|<τ, as in Equation:

$$A_j = \mathbf{1}\left[\frac{|\hat{y}_j - y_j|}{|y_j|} \leq \tau\right], \qquad \mathrm{YA}_\tau = \frac{1}{N}\sum_{j=1}^{N} A_j$$

Exact equality is not expected as there is district-level variability in yield by season and by growing conditions. The main tolerance is $\tau = 0.15$; sensitivity limits are reported as $\tau = 0.10$ and $\tau = 0.25$. If the tolerance is lower than the year-to-year natural differences of the reference records, then successful predictions that are agronomically acceptable would be penalized, while predictions with a much larger tolerance would be granted differences that have no agronomic significance. The percentage of queries where the reference crop is included in the four recommended crops is the separate categorical suitability agreement. The percentages in the original comparison (55% and 87%) were calculated before the establishment of a numerical tolerance and should be read as agreement to the original evaluation criterion, not the actual percentages.

i. Ablation Design

The contribution of each architectural component is isolated by comparing five different configurations in Table 9. Each configuration is based on the same snapshot of the GPT-4o model, decoding settings (temperature=0, fixed top-p, fixed max output length), evaluation queries in the same order, output schema, and scoring method. The weather context and the retrieval index remain constant over the course of the experiment to prevent the results from drifting over time, and the annotators are unaware of the index and weather that were used to generate each response. The vanilla GPT-4o baseline from the original comparison remains as a benchmark, but is not reported.

Table 9: Configurations of the component-level ablation study

ID	Configuration	Components enabled	Purpose
C0	Vanilla GPT-4o-mini	Model only; plain question; no schema or evidence	Reference for the pipeline's reasoning model
C1	Prompt-constrained GPT-4o-mini	C0 + system prompt, JSON schema, and evidence-use instructions	Effect of prompting and output constraints
C2	RAG-only GPT-4o-mini	C1 + metadata-constrained top-k retrieval	Marginal effect of retrieval grounding
C3	RAG + weather intelligence	C2 + Gemini 1.5 Flash weather context	Marginal effect of weather context
C4	Complete AgriSense pipeline	C3 + response normalization and financial validation	Marginal effect of response normalization

The marginal contributions are the differences between successive configurations for a metric M. $\Delta\text{prompt} = M(C1) - M(C0)$, $\Delta\text{retrieval} = M(C2) - M(C1)$, $\Delta\text{weather} = M(C3) - M(C2)$, and $\Delta\text{normalization} = M(C4) - M(C3)$. Pairing the patient and using the bootstrapping method with 1000 resamples provides 95% confidence intervals for each difference, while McNemar's test is used for paired binary responses like yield agreement. If the interval between the two numbers does not contain 0, then it is considered a contribution of the corresponding component.

ii. Retrieval Quality and Choice of k

The quality of retrieval is judged by its relevance to the agrosphere and not by its similarity in terms of text. A record has been retrieved if

a. It is agronomically appropriate for the listed soil, water source and season.

b. Its area is suitable for the farm site.

c. It is a relatively new year. These judgements are made independently by two agronomy annotators. With Rk(q) the set of top-k retrieved records and Rel(q) the set of relevant records, Precision@k and Recall@k are given by the equations shown below, and the Mean Reciprocal Rank by Equation as, where rank_q is the position of the first relevant record.

$$\text{P@}k = \frac{1}{|Q|}\sum_{q\in Q}\frac{|R_k(q)\cap \text{Rel}(q)|}{k}, \qquad \text{R@}k = \frac{1}{|Q|}\sum_{q\in Q}\frac{|R_k(q)\cap \text{Rel}(q)|}{|\text{Rel}(q)|}$$

$$\text{MRR} = \frac{1}{|Q|}\sum_{q\in Q}\frac{1}{\text{rank}_q}$$

Three different variants of retrieval are compared: Unconstrained semantic search, metadata-constrained search, and metadata-constrained search with the compatibility screen. The value of k is varied over {3, 5, 10} and results are stratified by state and by season (Kharif, Rabi, Zaid) to check stability with respect to geographical and seasonal variation separately. The stability measure is the standard deviation of Precision@5 across strata.

iii. Full-Pipeline versus Vanilla LLM Results

Table 10 shows a comparison of the vanilla GPT-4o baseline model against the entire AgriSense AI pipeline. This gain in yield is a 32% jump from about 55% to 87%. The rate of response-level hallucinations drops from about 28% to 6%, a 22-point absolute reduction and a 79% relative reduction. This baseline analysis compared the factual consistency of the complete pipeline with the baseline result, as seen in Figure 5.

Table 10: Comparison of vanilla LLM and AgriSense AI performance

Metric	Vanilla LLM	AgriSense AI	Reported change
Yield accuracy	~55%	~87%	+32 percentage points
Hallucination rate	~28%	~6%	−22 percentage points; 78% relative reduction
Region-specific advice	Generic	Hyper-local	Qualitative improvement
NPK precision	Round estimates	Exact kg/acre values	Improved specificity
Market-price accuracy	Training-data dependent	Near real-time agent context	Improved timeliness
Language fidelity	English-centric	Native multilingual generation	Expanded coverage

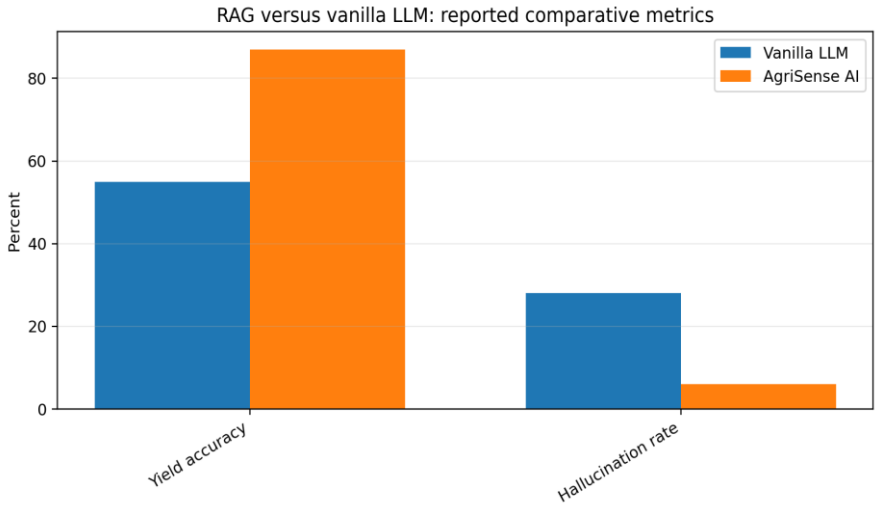


Fig. 5. Reported comparative metrics for the vanilla LLM and RAG-grounded AgriSense pipeline.

iv. State-Wise Results

The state-wise Table 11 shows that the accuracy of AgriSense is between 80% and 92%, highest in Gujarat (92%) and lowest in Rajasthan (80%), as shown in Table 10 and Figure 6. The average baseline is 57.6%, and the average AgriSense value is 87.1%, for an average gain of 29.5 percentage points. Importantly, the sum of the numbers of queries by state is 216, while the controlled comparison section reports 200 farmer queries. Such internal incoherence has to be addressed prior to submitting the paper as a validated empirical study by consulting the experiment logs.

Table 11: State-wise comparison of AgriSense accuracy with the baseline

State	Queries	Baseline	AgriSense	Improvement	Top crop
Punjab	28	61%	89%	+28%	Wheat
Maharashtra	32	58%	91%	+33%	Cotton
Karnataka	24	55%	87%	+32%	Ragi
Andhra Pradesh	22	60%	88%	+28%	Rice
Tamil Nadu	20	57%	86%	+29%	Rice
West Bengal	18	54%	84%	+30%	Rice
Gujarat	26	62%	92%	+30%	Cotton
Rajasthan	16	50%	80%	+30%	Bajra
Uttar Pradesh	30	59%	87%	+28%	Wheat
Average	216	57.6%	87.1%	+29.5%	-

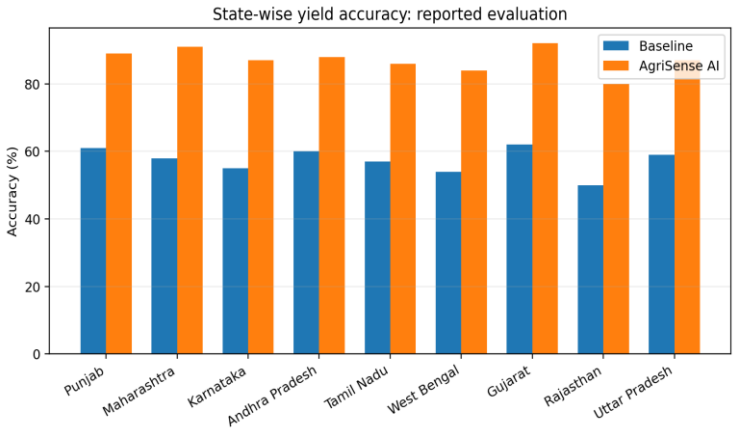


Fig. 6. State-wise yield-accuracy comparison reported in the supplied system document.

v. Multilingual Evaluation

As shown in Table 12 presents the multilingual evaluation of the proposed agricultural recommendation system across five quality dimensions: grammar, technical accuracy, cultural fit, comprehension, and overall quality. The scores appear to be reported on a 5-point evaluation scale, with higher values indicating better output quality. The results show consistently high scores across all evaluated languages, demonstrating that the system is capable of generating linguistically appropriate and technically meaningful agricultural recommendations in multiple Indian languages.

Table 12: Multilingual evaluation of the agricultural recommendation system

Language	Grammar	Technical accuracy	Cultural fit	Comprehension	Overall
Hindi	4.7	4.8	4.6	4.7	4.7
Punjabi	4.5	4.6	4.8	4.6	4.6
Marathi	4.6	4.7	4.5	4.5	4.6
Kannada	4.4	4.5	4.4	4.4	4.4
Telugu	4.5	4.6	4.5	4.5	4.5
Tamil	4.3	4.5	4.4	4.3	4.4
Bengali	4.4	4.5	4.3	4.4	4.4
Gujarati	4.6	4.7	4.6	4.6	4.6
English	4.9	4.9	4.7	4.7	4.8

Figure 7 shows that the overall scores remain within a relatively narrow range of 4.4–4.8, indicating consistently high perceived quality across all evaluated languages. English and Hindi receive the highest overall ratings, while Kannada, Tamil, and Bengali record slightly lower but still strong scores. The small variation between languages suggests that the proposed system maintains relatively consistent output quality rather than being strongly optimized for only one language.

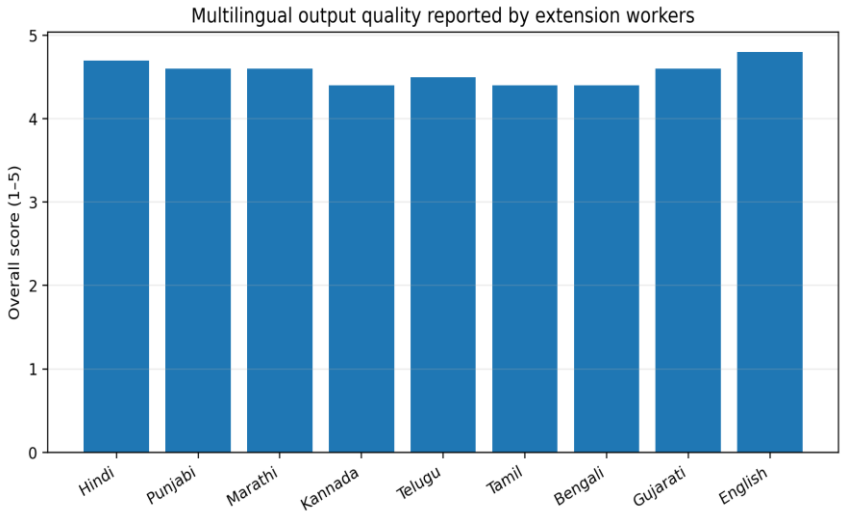


Fig. 7. Overall multilingual output-quality scores reported by agricultural extension workers.

vi. Illustrative Fertilizer-Mix Visualization

The source output schema includes an example fertilizer mix of Urea: 50 kg, DAP: 40 kg, and MOP: 20 kg. Figure 8 is meant to be illustrative only and should not be interpreted as a universal agronomic prescription.

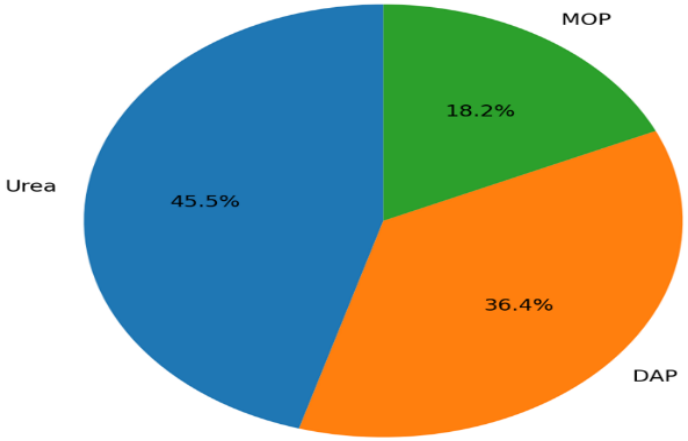


Fig. 8. Illustrative composition of the fertilizer-mix example contained in the source output schema.

VII. State-Wise Case Studies

The provided system document provides several state and language scenarios. The cases shown in Table 13 are lumped together to show the effect of location, soil, irrigation, and language. The numbers provided in these cases are not the result of a randomized field experiment and should be considered as the result of a scenario being generated on the platform.

Table 13: State-wise results of the agricultural recommendation system

Case	Language	Farm profile	Top-four recommendations	Profit range	ROI range	NPK	Water
Karnataka	Kannada	4 acres; ₹60,000; red soil; borewell; Kharif; previous crop Ragi	Groundnut; Pigeon Pea; Maize; Black Gram	₹16,000–₹54,000	33.3–116.0%	N 40–120 kg; P 16–48 kg; K 12–36 kg	2,000–8,000 L/day
Telangana	Telugu	4 acres; ₹60,000; black/regur soil; canal irrigation; Kharif; previous crop Ragi	Cotton; Soybean; Sugarcane; Groundnut	₹45,000–₹1,50,000	50.0–114.3%	N 40–240 kg; P 16–96 kg; K 12–72 kg	6,000–24,000 L/day
Tamil Nadu	Tamil	4 acres; ₹60,000; clay-loam/alluvial; canal irrigation; Kharif; previous crop Ragi	Rice; Banana; Rice/historical champion; Turmeric	₹85,000–₹1,40,000	42–133%	N 88 kg; P 44 kg; K 33 kg (Top 1)	8,800 L/day
Maharashtra	Marathi	4 acres; ₹60,000; black/regur; drip irrigation; Kharif; previous crop Ragi	Cotton; Soybean; Maize; Groundnut	₹36,000–₹72,000	28.6–68.0%	N 120 kg; P 48 kg; K 36 kg (Top 1)	6,000 L/day
Gujarat	Gujarati	4 acres; ₹60,000; sandy loam; borewell; Kharif; previous crop not specified	Maize; Groundnut; Other Cereals; Bajra	Source provides crop-level dashboard values	Not consistently summarized	Maize: N 120 kg; P 48 kg; K 36 kg	Maize: 6,000 L/day
Punjab	Punjabi	State-level evaluation case; detailed farm inputs not fully reproduced in the source excerpt	Wheat identified as top crop in evaluation	—	89% AgriSense accuracy	—	—

VIII. Results

The developed system demonstrated a multi-source, context-aware agricultural recommendation capability by combining retrieved agricultural knowledge, environmental information, market context, and model-generated recommendations. The results indicate that the proposed architecture can provide more practical and localized recommendations by incorporating contextual information rather than relying solely on static crop knowledge. The system also demonstrated the feasibility of presenting recommendations as a portfolio of crop alternatives, enabling farmers to

compare potential benefits, resource requirements, and associated risks. The retrieval-grounded approach further supported the generation of recommendations based on identifiable knowledge sources, while the uncertainty layer provided a mechanism for distinguishing retrieved facts, model inferences, and scenario-based estimates.

The system architecture also demonstrated clear potential for future real-world deployment through integration with additional data sources. Real-time soil parameters such as pH, moisture, and temperature can be incorporated through low-cost IoT sensors, while live mandi-price information can enhance the economic relevance of recommendations. Integration of hyper-local weather information and satellite-derived crop-health indicators such as NDVI can further improve environmental and field-level contextualization. The results also establish a scalable framework in which the agricultural knowledge base can be progressively expanded to cover a substantially larger number of crops and varieties.

Furthermore, the proposed framework demonstrated the feasibility of extending agricultural advisory services to users with different levels of digital literacy. Regional-language speech-to-text and text-to-speech capabilities can provide an accessible interaction mechanism, particularly for low-literacy and feature-phone users. The system can be further expanded from crop recommendation to more comprehensive agricultural decision support by government-scheme matching based on farmer profile and location. The system can be further extended from crop recommendation to agricultural decision support by government-scheme matching based on farmer profile and location. The framework also offers a basis for introducing data provenance, versioning, time stamps, privacy settings, and audits to increase transparency and traceability of recommendations generated. In conclusion, the findings highlight the potential for this proposed architecture to serve as a scalable framework for localized, multilingual, retrieval-grounded, and context-aware agricultural decision support, with Version 2.0 serving as a blueprint for the road ahead toward real-time sensing, live market integration, improved environmental intelligence, and more transparent AI-driven recommendations.

IX. Discussions

AgriSense AI's core contribution is NOT a new classification algorithm; it's a systems architecture of complementary forms of intelligence. A historical agricultural dataset, a weather agent provides the current data, and an LLM provides natural language reasoning based on the combined context. This division is crucial because none of the components is uniquely well suited for all three tasks. A vector database can be used for similarity retrieval, but not for explaining a recommendation; a weather service can give you the current weather, but it has no knowledge of how well your crops are doing historically; and an LLM can integrate information, but should not be relied upon to come up with local statistics. The reported RAG results are in support of the architectural hypothesis. The percentage of yields that can be agreed upon increases from about 55% to 87%, and the percentage of yields that hallucinate decreases from about 28% to 6%. The state-wise table also indicates that the improvements are present in all reported states, implying that grounding and contextualizing are not only limited to one region. However, a conclusion cannot be drawn from this that it is RAG that is responsible for the improvement. In addition to the above, the complete system

contains structured prompts, a weather agent, and a response normalizer. The results for multilingual are also promising. Extension-worker evaluators judged the outputs as acceptable to strong, with overall scores ranging from 4.4–4.8 for the languages documented. The overall scores are highest for English (4.8), Hindi (4.7), and some regional languages (4.4-4.6). This is not necessarily an indication that the approach is not working; it supports the need for improved evaluation and terminology support on a language-specific level, and to conduct farmer-level comprehension studies and speech interfaces for users with low literacy levels. The four-crop portfolio is a great UI innovation as it reveals the trade-offs. If he expects a crop to have greater profits, a farmer may be tempted to plant one that has lower returns, or he may be tempted to plant a historic champion crop because he knows it well and it is less of a risk of failure. The portfolio design may then be utilized to complement the human sense, not to replace it. A future version could operationalize the portfolio as a Pareto-front recommendation problem, and explicitly indicate trade-offs between expected profit, water use, investments, climate risk, and historical yield stability. It is the most scientific validation challenge, apart from the image-analysis module that is the first step toward multimodal agriculture. Although the models often work very well in controlled settings, they do not do as well when they are used in the field, as they are affected by illumination, background noise, camera quality, stage of disease, and unseen cultivars. It is therefore important that the system be precise, accurate, and reliable for every class of images, recall, and F1, and be able to detect images out of distribution and perform out-of-distribution performance in the field before it makes a strong claim in disease diagnosis. Additionally, agricultural extension recommendations and local regulations should be compared with treatment options to ensure that the most current information is included. Internet dependence, smartphone literacy, sensor data, market-price latency, and hyper-local weather accuracy and limited crop coverage are cited as significant limitations on the deployment side, as cited by the source. These constraints align with the overall digital-agriculture research, which emphasizes the importance of infrastructure, skills, economic incentives, institutional arrangements, and trust in digital agriculture technology adoption. The aforementioned (roadmap) features - source offline RAG snapshots, voice interaction, IoT sensors, live market feeds, satellite imagery, and federated learning - are technically feasible and in line with existing research**.**

X. Conclusion

The retrieval, the real-time context of the environment, the language-model reasoning, the multilingual communication, and the multi-modal analysis of the crop are combined into an integrated solution for agricultural decision support in AgriSense AI. The architecture addresses a real-world problem of existing crop classifiers and unconstrained LLMs: the former tend to make "flat, general" predictions, and the latter tend to make "fluently, but not substantiated" predictions. The platform will aim to be more localized, traceable, and with context to the weather by using historical evidence and by validating structured outputs. As shown in the provided system results, the percentage of yielded tasks with good agreement also increases greatly compared to the result of a vanilla LLM baseline, while the percentage of hallucinations significantly drops, and the quality scores from the set of languages documented are above 4.4/5. The results by state show steady growth, while the case studies illustrate

the impact of the soil, irrigation, season, and location on the recommendation portfolio. This platform brings this system beyond a crop classifier to an agricultural assistant through the dashboard and history module, and the image-analysis pathway. The next most critical step is rigorously repeatable. Authors should address the discrepancy between 200 and 216 evaluations before submission to the Journal; provide the number of evaluations in the language used; provide the number of partitions in the experimental dataset; record the metrics used for retrieval; and report the exact experimental protocol and independently validate the image-based disease module. AgriSense AI can become a framework beyond the mere proof point of generative AI in agriculture, towards evidence-based, multilingual and human-centric agricultural decision support, by implementing such improvements.

Conflict of Interest:

There was no relevant conflict of interest regarding this paper.

References

I. Achiam, Josh, et al. "GPT-4 Technical Report." *arXiv*, 2023, 10.48550/arXiv.2303.08774 .

II. Conneau, Alexis, et al. "Unsupervised Cross-Lingual Representation Learning at Scale." *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 8440–8451. 10.18653/v1/2020.acl-main.747 .

III. Devlin, Jacob, et al. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, 2019, pp. 4171–4186. 10.18653/v1/N19-1423 .

IV. Ferentinos, Konstantinos P. "Deep Learning Models for Plant Disease Detection and Diagnosis." *Computers and Electronics in Agriculture*, vol. 145, 2018, pp. 311–318. 10.1016/j.compag.2018.01.009

V. Fountas, Spyros, et al. "Farm Management Information Systems: Current Situation and Future Perspectives." *Computers and Electronics in Agriculture*, vol. 115, 2015, pp. 40–50. 10.1016/j.compag.2015.05.011 .

VI. Gebbers, Robin, and Viacheslav I. Adamchuk. "Precision Agriculture and Food Security." *Science*, vol. 327, no. 5967, 2010, pp. 828–831. 10.1126/science.1183899 .

VII. Gemini Team, et al. "Gemini 1.5: Unlocking Multimodal Understanding across Millions of Tokens of Context." *arXiv*, 2024, 10.48550/arXiv.2403.05530 .

VIII. Hu, Junjie, et al. "XTREME: A Massively Multilingual Multi-Task Benchmark for Evaluating Cross-Lingual Generalisation." *Proceedings of the 37th International Conference on Machine Learning*, 2020, pp. 4411–4421. PMLR. https://proceedings.mlr.press/v119/hu20b.html .

IX. Hughes, David P., and Marcel Salathé. "An Open Access Repository of Images on Plant Health to Enable the Development of Mobile Disease Diagnostics." *arXiv*, 2015. https://arxiv.org/abs/1511.08060

X. Jabed, Md. Abu, and Masrah Azrifah Azmi Murad. "Crop Yield Prediction in Agriculture: A Comprehensive Review of Machine Learning and Deep Learning Approaches, with Insights for Future Research and Sustainability." *Heliyon*, vol. 10, no. 24, 2024, article e40836. 10.1016/j.heliyon.2024.e40836 .

XI. Jha, Kirtan, et al. "A Comprehensive Review on Automation in Agriculture Using Artificial Intelligence." *Artificial Intelligence in Agriculture*, vol. 2, 2019, pp. 1–12. 10.1016/j.aiia.2019.05.004 .

XII. Kakwani, Divyanshu, et al. "IndicNLPSuite: Monolingual Corpora, Evaluation Benchmarks and Pre-Trained Multilingual Language Models for Indian Languages." *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020, pp. 4948–4961. 10.18653/v1/2020.findings-emnlp.445 .

XIII. Kallenberg, M. G., et al. "Integrating Process-Based Models and Machine Learning for Crop Yield Prediction." *arXiv*, 2023, https://arxiv.org/abs/2307.13466 .

XIV. Kamilaris, Andreas, and Francesc X. Prenafeta-Boldú. "Deep Learning in Agriculture: A Survey." *Computers and Electronics in Agriculture*, vol. 147, 2018, pp. 70–90. 10.1016/j.compag.2018.02.016 .

XV. Khanuja, Simran, et al. "MuRIL: Multilingual Representations for Indian Languages." *arXiv*, 2021. https://arxiv.org/abs/2103.10730 .

XVI. Klerkx, Laurens, Emma Jakku, and Pierre Labarthe. "A Review of Social Science on Digital Agriculture, Smart Farming and Agriculture 4.0: New Contributions and a Future Research Agenda." *NJAS: Wageningen Journal of Life Sciences*, vols. 90–91, 2019, pp. 1–16. 10.1016/j.njas.2019.100315 .

XVII. Lewis, Patrick, et al. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 9459–9474. https://papers.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html .

XVIII. Li, J., et al. "Label-Efficient Learning in Agriculture: A Comprehensive Review." *Computers and Electronics in Agriculture*, vol. 215, 2023, article 108412.

XIX. Liakos, Konstantinos G., et al. "Machine Learning in Agriculture: A Review." *Sensors*, vol. 18, no. 8, 2018, article 2674. 10.3390/s18082674 .

XX. Lowenberg-DeBoer, James, and Bruce Erickson. "Setting the Record Straight on Precision Agriculture Adoption." *Agronomy Journal*, vol. 111, no. 4, 2019, pp. 1552–1569. 10.2134/agronj2018.12.0779 .

XXI. Mohanty, Sharada P., David P. Hughes, and Marcel Salathé. "Using Deep Learning for Image-Based Plant Disease Detection." *Frontiers in Plant Science*, vol. 7, 2016, article 1419. 10.3389/fpls.2016.01419 .

XXII. Pathak, Deepak, et al. "Predicting Crop Yield with Machine Learning: An Extensive Analysis of Input Modalities and Models on a Field and Sub-Field Level." *IGARSS 2023–2023 IEEE International Geoscience and Remote Sensing Symposium*, 2023, pp. 2767–2770. IEEE, 10.1109/IGARSS52108.2023.10282318 .

XXIII. Schick, Timo, et al. "Toolformer: Language Models Can Teach Themselves to Use Tools." *Advances in Neural Information Processing Systems*, vol. 36, 2023, pp. 68539–68551. 10.52202/075280-2997 .

XXIV. Singh, Namita, et al. "Farmer.Chat: Scaling AI-Powered Agricultural Services for Smallholder Farmers." *arXiv*, 2024. https://arxiv.org/abs/2409.08916 .

XXV. Sladojevic, Srdjan, et al. "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification." *Computational Intelligence and Neuroscience*, vol. 2016, 2016, article 3289801. 10.1155/2016/3289801 .

XXVI. Too, Edna Chebet, et al. "A Comparative Study of Fine-Tuning Deep Learning Models for Plant Disease Identification." *Computers and Electronics in Agriculture*, vol. 161, 2019, pp. 272–279. 10.1016/j.compag.2018.03.032 .

XXVII. van Klompenburg, Thomas, Ayalew Kassahun, and Cagatay Catal. "Crop Yield Prediction Using Machine Learning: A Systematic Literature Review." *Computers and Electronics in Agriculture*, vol. 177, 2020, article 105709. 10.1016/j.compag.2020.105709 .

XXVIII. Wolfert, Sjaak, et al. "Big Data in Smart Farming—A Review." *Agricultural Systems*, vol. 153, 2017, pp. 69–80. 10.1016/j.agsy.2017.01.023 .

XXIX. Yao, Shunyu, et al. "ReAct: Synergizing Reasoning and Acting in Language Models." *arXiv*, 2022. https://arxiv.org/abs/2210.03629 .

XXX. Zhang, Naiqian, Maohua Wang, and Ning Wang. "Precision Agriculture—A Worldwide Overview." *Computers and Electronics in Agriculture*, vol. 36, nos. 2–3, 2002, pp. 113–132. 10.1016/S0168-1699(02)00096-0 .