IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI-Based Farmer Query Support and Advisory System

Yash Srivastava[1] , Prof. Dr. Swati Patel[2]

[1]Student, MCA (Data Science), Ajeenkya D Y Patil University, Lohegaon, Pune, Maharashtra

[2]Head of Department, Department of Computer Science, Ajeenkya D Y Patil University Lohegaon, Pune, Maharashtra

Abstract: The rapid evolution of digital agriculture had necessitated the development of intelligent systems that could bridge the information gap for smallholder farmers. This project presents an AI-Based Farmer Query Support and Advisory System, a combined framework designed to provide real-time and data-driven agricultural guidance through advanced computational techniques. To leverage Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG), this system ensures that queries related to crop management, pest control, and soil health are answered with high factual accuracy, minimizing the risks of AI hallucinations often found in standard generative models [1], [2], [11].

A core challenge in global agriculture is the linguistic barrier; as a result, this system implements multilingual and cross-lingual architectures to support low-resource languages, ensuring that farmers can interact in their native dialects while the system retrieves expert knowledge from scientifically validated English corpora [3], [8], [21]. To enhance the robustness of its recommendations, the framework utilizes Multi-Modal Fusion and IoT-based real-time data integration [9], [15]. This allows the "Advisory and Recommendation Engine" to process diverse inputs, ranging from voice and text queries to real-time weather forecasts and soil sensor data, to generate site-specific advice [4], [7], [22].

Furthermore, the system incorporates Deep Learning models, specifically Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), to provide accurate agricultural commodity price forecasting and yield predictions [5], [12]. This predictive capability empowers farmers with market intelligence and policy scheme optimization, facilitating better financial decision-making [15], [19]. By ensuring accessibility in remote regions with limited connectivity, this architecture includes an Offline Edge AI framework, enabling localized processing of sensor data and immediate advisory delivery [6], [22].

The integration of a Feedback and Learning Module creates a closed-loop system where user interactions and expert-vetted "Golden Q&A" datasets are used to iteratively improve the model’s performance [4], [11]. By combining state-of-the-art AI innovation with practical agricultural extension needs, this system serves as a scalable roadmap for the digital transformation of global agrifood systems, ultimately enhancing productivity, sustainability, and food security for smallholder farmers [18], [20], [23]

Keywords - Large Language Models (LLMs) [1, 21], Retrieval-Augmented Generation (RAG) [3, 23], Random Forest Classifier [2, 12], Nearest Neighbor [3, 21], Multi-Modal Fusion [15, 17], Agricultural Advisory System [7, 10], Voice-based Farmer Support [8, 15], Smallholder Farmers [4, 11, 20], and Market Intelligence [5, 15].

1. Introduction

The global agrifood system is currently navigating a critical crossroads, driven by the convergence of intensifying climate shocks, geopolitical volatility, natural resource depletion, and unstable markets [1], [2], [3], [4]. As the world population is projected to reach approximately 9.73 billion by 2050, the Food and Agriculture Organization (FAO) emphasizes that agricultural productivity must rise significantly to ensure global food security [1], [3], [5], [6]. This urgent necessity had catalyzed a profound digital transformation, marking the transitions from traditional labour-intensive practices towards Agriculture 4.0 and 5.0, where robotics, the Internet of Things (IoT) and artificial intelligence (AI) enable precision resource management and real-time decision-making [7], [8], [4], [9]. Central to this evolution is the leap from classical machine learning to powerful, affordable generative AI and Large Language Models (LLMs), that had democratized access to precision technologies for non-technical users, even in rural and resource-constrained settings [1], [3], [5], [10].

Despite these technological advancements, smallholder farmers, who manage over 80% of global farm holdings , remain the primary actors in food security while simultaneously being the most exposed to environmental instabilities and digital exclusion [11], [12], [5], [6]. Traditional human-centric agricultural extension services are currently facing a systemic crisis characterized by a severe shortage of human agents [6]. For instance, in regions like Nigeria, the extension agent-to-farmer ratio is as low as 1:10,000, which trails significantly behind the FAO's recommended 1:1,000 ratio [6]. This "knowledge gap" creates a persistent information asymmetry where scientifically validated intelligence , often locked in static PDFs or manuals from organizations such as the FAO and the International Rice Research Institute (IRRI) , fails to reach the field in a timely or actionable manner [1], [13], [10]. This disconnects results in sub-optimal planting decisions, the misuse of chemical inputs, and arise in vulnerability to pests and diseases [1], [14], [3]. Furthermore, existing digital tools are often highly fragmented, focusing on isolated tasks rather than providing a unified, context-aware decision-making dashboard [15], [7], [14], [16].

To bridge this divide, this development of a combined AI-Based Farmer Query Support and Advisory System (FQAS) is proposed to democratize access to expert-level agronomic guidance through intuitive, multimodal interfaces [13], [17], [14], [18]. A primary technical hurdle in rural extension is the "vocabulary gap" and the linguistic barrier; while technical manuals are predominantly written in English, farmers primarily communicate in low-resource regional dialects such as Bengali, Swahili, or Marathi [13], [17], [10]. Current research emphasizes a translation-centric "sandwich" architecture to resolve this, where native language queries are translated into a high-resource reasoning language (English) to be processed against expert databases before delivering a grounded response back to the farmer [13], [19], [10]. To further enhance inclusivity for semi-literate populations, this combination of multilingual voice interfaces utilizing Automated Speech Recognition (ASR) and Text-to-Speech (TTS) technologies has proven essential [11], [19], [12], [20]. Field trials conducted under the Agricultural Information Exchange Platform (AIEP) Initiative have demonstrated high user satisfaction rates, with Net Promoter Scores (NPS) of approximately 60 among cohorts of smallholders in Kenya and India [11], [12].

Technical reliability is a core pillar of the FQAS, specifically addressing the inherent risk of AI "hallucinations" , the generation of plausible but scientifically incorrect advice that could lead to total crop failure [1], [15], [13], [17], [10]. This system utilizes a Retrieval-Augmented Generation (RAG) framework to ground LLM responses in curated, scientifically validated corpora from authoritative sources [1], [15], [13], [10]. Comparison benchmarks indicate that ChatGPT-4o mini with RAG achieves an average accuracy of 93% for complex agricultural queries [1], [10]. Furtherly, this research into Hybrid RAG strategies, that combines keyword-based retrieval with semantic similarity search, has demonstrated superiority over direct long-context prompting to overcome the "Lost in the Middle" effect often observed in large context windows [15]. By compelling models to cite specific sources from extension manuals and textbooks, researchers had observed an 85% reduction in factual errors compared to baseline models [10].

Over the textual reasonings, this FQAS methodology incorporates real-time physical monitoring by providing hyper-local environmental contexts [7], [8], [20], [9]. Using IoT-based sensing arrays integrated with Arduino or ESP32 microcontrollers, this system tracks critical parameters including Nitrogen (N), Phosphorus (P), Potassium (K), soil pH, moisture, and temperature [7], [8], [20], [9]. To serve regions with zero connectivity, this platform implements a real-time offline edge AI framework, where multi-class supervised classifiers, such as Random Forest models, performing crop classification and fertilizer recommendation entirely on-device [7], [9]. These edge engines had achieved accuracies of 96.28% with sub-second latency, providing localized support even where GSM or Wi-Fi services are offline [7], [9]. For early-stage disease diagnosis, this platform combines Convolutional Neural Networks (CNN), specifically the InceptionV3 architecture, that identifies pathologies from leaf imagery with accuracies ranging between 91% and 94% [20], [18], [16].

Economic intelligence is an another vital component, which helps farmers to navigate the non-linear volatility of commodity price markets [2], [17], [14]. Deep learning models, especially Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), significantly outperform traditional stochastic models such as ARIMA by capturing complicated temporal patterns in market data [2]. For instance, the GRU models have achieved a Root Mean Squared Error (RMSE) of 369.54 for onions, as compared to 1564.62 for traditional methods [2]. These forecasts are unified with real-time mandi prices fetched via the Agmarknet API to provide farmers with dynamic profit estimations, risk alerts, and alternative crop suggestions based on Euclidean distance algorithms [21], [17], [14]. Additionally, the system provides integrated access to government schemes and subsidies, filtering them based on individual user profiles to improve adoption rates by up to 31.8% [19], [14].

This framework further facilitates human-AI synergy via a live consultation interface [18], [16]. Developed using WebRTC to enable secure, low-latency video calls even at bandwidths as low as 512Kbps, this module allows farmers to connect with human agronomists only when AI triage indicates a high-complexity problem [18], [16]. Benchmark tests demonstrate that this hybrid model can reduce diagnostic times by as much as 68% compared to conventional extension visits, ensuring that expert intervention is both timely and scalable [18], [16]. Quantitative evaluations involving farmers in the Solapur district and other active agricultural zones have revealed that engaged users achieved a 42.5% improvement in decision-making efficiency and a 2.8x higher yield optimization compared to traditional methods [14].

Ultimately, this research establishes a scalable blueprint for Digital Public Infrastructure (DPI) in agriculture [3], [5]. To synthesize real-time physical monitoring, grounded generative reasoning, and predictive economic analytics into a single, cohesive platform, the FQAS addresses the fragmentation of existing toolchains [7], [14], [16]. This holistic approach ensures that every farmer, regardless of digital literacy, connectivity, or language, has access to the scientifically grounded intelligence needed to improve their yields, reduce resource wastage, and build long-term climate resilience [11], [22], [8], [23]. Success in this agrifood transformation depends to build interoperable data ecosystems and common corpora that share risk and reward equitably by empowering the millions of farmers who feed the world [11], [3], [5], [6].

2. Literature Review

This fundamental challenge in global agricultural development remains the acute information asymmetrical between scientifically validated protocols and the practitioners who require them, particularly in developing regions. In Bangladesh, the ratio of extension agents to farmers is estimated to be a staggering 1:10,000, which significantly hinders the distribution of expert guidance [19]. Although international organizations like the Food and Agriculture Organization (FAO) and the International Rice Research Institute (IRRI) create comprehensive guidance materials, the format and length of these resources , often presented as static PDF documents exceeding 165 pages , make them practically inaccessible to the average rural operator [5][7]. This accessibility issue is further complicated by the increasing complexity of modern precision agricultural equipment. A prime factor is the Kverneland Exacta-TLX Geo spread GS3 fertilizer spreader, a complex mechatronic system that requires precise operational knowledge for both safety and effective agricultural use [5][7]. These operational manuals for such machineries are not only linguistically dense but they also demand high-fidelity retrievals of specific technical parameters, such as the requirement to torque vane lock nuts to exactly 50 Nm by ensuring structural integrity during high-speed disc rotation [1][5]. Navigating these 165-page technical manuals is a hectic task even for experts, creating a structural barrier that hinders the adoption of precision agricultural practices and sustainable productivity. Traditional computational approaches have increasingly sought to leverage Large Language Models (LLMs) to serve as intuitive natural language interfaces for this technical knowledge, yet these models face significant structural and cognitive limitations when processing long-form industrial documentation.

One of the most critical limitations in current transformer architectures is the "Lost in the Middle" effect, where LLMs, despite possessing nominal context windows of up to 128,000 tokens, demonstrate a marked failure to retrieve and process information located in the center of a provided data stream [11]. In benchmarks using the Kverneland technical manual, which comprises approximately 59,000 tokens, models consistently exhibit a degradation in performance as the context length increases toward the full manual size. Empirical data indicates a precipitous drop in F1 scores for smaller models as the "noise" of irrelevant pages increases. For instance, while Llama 3.2 1B maintains an F1 score of 0.422 at 1,000 tokens of context, its performance collapses to 0.018 when tasked with finding information within the full 59,000-token manual [5]. This phenomenon indicates that as the context length arises, the model's ability is to identify specific technical "needles" within the "haystack" of a 165-page manual degrades significantly, leading to a reliance on early or late context segments that may not contain the requisite technical specifications [11]. Even larger proprietary models like Gemini 2.5 Flash, which shows greater stability, experience a decline from an F1 score of 0.867 at 1,000 tokens to 0.744 at the full 59,000-token context [5]. This results a pivot away from simple long-context prompting towards more robust architectural frameworks, such as Retrieval-Augmented Generation (RAG), which could maintain high fidelity across large technical datasets by isolating relevant segments before generation.

The challenge of technical information access is further made by a formidable linguistic barrier in regions where authoritative scientific manuals are published in high-resource languages like English, but the primary user base communicates in low-resource languages such as Bengali [3][9]. This mismatch prevents direct interaction with global knowledge bases, effectively silencing the expert guidance intended for rural communities. Furthermore, a significant "vocabulary gap" exists between the colloquial terminology used by farmers and the scientific nomenclature used in expert systems [7]. For example, a farmer may describe a specific rice affliction as "Magra," a mechanism that fails to trigger traditional keyword-based retrieval for its scientific equivalent, the "Stem Borer" (Scirpophaga incertulas) [7]. Bridging this terminological divide requires sophisticated semantic mapping that may align local linguistic nuances with global agronomical standards. The inability of standard LLMs to generate fluent and factually consistent Bengali , a language often inadequate in training sets , leads to degraded performance and increased risks of misinformation in agricultural decision-making [3][11].

In response to these multi-faceted challenges, this comparative effectiveness of RAG and Direct Long-Context (LC) prompting has become a focal point of contemporary research. While LC windows are expanding to accommodate between 128,000 and 1,000,000 tokens, these systems remain prone to hallucinations and incur prohibitive inference costs which makes them unsuitable for widespread rural deployment [8][10][13]. RAG systems, by contrast, prioritize the retrievals of specific, grounded segments of text before generation, offering a more resource-efficient and factually stable alternative. However, the choice of retrieval methodology , Keyword-based (BM25), Semantic/Embedding-based, or Hybrid , is critical to system success. Keyword retrievals, while efficient, often rank documents on the basis of the frequency of query terms and can fail if queries use synonyms not present in the manual [5]. Semantic retrieval, which utilizes numerical vectors (embeddings) and cosine similarity to identify pertinent segments, is, however, susceptible to overlooking precise technical identifiers [5]. Hybrid retrieval, incorporating methodologies such as Reciprocal Rank Fusion (RRF), capitalizes on the advantages of both approaches, thereby ensuring system resilience even when queries are expressed informally or contain minor linguistic inaccuracies [2][12]. Data derived from the Agri-Query benchmark substantiates that Hybrid RAG consistently attains superior accuracy through various models; for instance, Gemini 2.5 Flash achieved an accuracy of 0.880, while smaller models like Qwen 2.5 7B attained 0.861, demonstrating a substantial improvement over full-manual context insertion, which yielded accuracies of 0.694 and 0.398, respectively, for the same models [5].

This fundamental motivation to develop these advanced advisory systems is rooted in the need for absolute reliability and trust in agricultural decision-making. To ensure that a farmer can safely rely on an AI-generated recommendation, the system must achieve a minimum query accuracy of 93% and a crop prediction accuracy of at least 96.28% [2][7]. These performance standards are not just theoretical; they are crucial to maintain long-term productivity. A mistake to suggest pesticide amounts or irrigation schedules can result into complete crop failure or lasting environmental losses. Consequently, research is driven by a requirement to minimize hallucinations, particularly in safety-critical mechatronic operations where precise specifications are paramount [1][5]. Implementing "Strict Context-Only Instructions" serves as a primary safeguard, forcing the model to acknowledge its own information gaps rather than fabricating plausible but dangerous answers. Specificity scores are used to measure a model's ability to reject irrelevant questions. This is particularly useful for questions that don't fit the technical context, like those about diesel use or garden suitability. Larger models like Gemini 2.5 Flash exhibit a Specificity of 0.796, whereas smaller models like Llama 3.2 1B demonstrate a dangerously low Specificity of 0.204, indicating a high propensity for hallucinating answers to out-of-domain questions [5].

Economic constraints further dictate that this deployment of such systems must account for the infrastructure limitations of rural environments, favouring cost-efficient deployment on consumer-grade hardware. Using NVIDIA RTX 6000 or Tesla T4 GPUs with 4-bit quantization allows high-performance LLMs to run locally, by ensuring that advisory services could be provided without reliance on expensive, high-bandwidth cloud APIs [13][18]. This shift toward localized "Edge AI" is critical for privacy and accessibility. The research into "Translation-Centric" architectures , where Bengali queries are translated into English, processed via RAG, and then translated back to Bengali , is motivated by the superior reasoning and retrieval capabilities of models in their primary training language [3][11]. The "sandwich" architecture involving models like Helsinki-NLP (opus-mt-bn-en) and NLLB-200 offers the most viable path forward for low-resource languages, leveraging the density of English scientific literature when maintaining local accessibility [11][15]. It is essential because direct Bengali generation by English-centric models often exhibits poor grammatical quality and factual inconsistency [3].

By achieving these motivations, these primary research objectives focus on the implementation of a Cross-Lingual RAG system using state-of-the-art multilingual embedding models, like gte-Qwen2-7B-instruct, integrated with high-performance vector databases like FAISS or ChromaDB [9][16][17]. These tools enable the efficient indexing of hundreds of technical chunks, allowing for sub-second retrieval across large knowledge bases. This technical process begins with precise documents preparation, where PDFs were converted into Markdown format using the Docling library and a page-wise conversion wrapper to preserve structural integrity [1][4]. This text is then segmented into 200-token chunks within a 100-token overlap to ensure no technical context is lost at these boundaries, an optimization which facilitates the preservation of complex automated instructions [5]. Furtherly, this research seeks to integrate real-time IoT soil sensing data directly into the AI advisory loop. To incorporate DHT11 sensors for temperature and humidity, alongside dedicated pH and moisture sensors (e.g., through a NodeMCU or Wi-Fi gateway), this system moves beyond static text retrieval to provide dynamic, data-driven recommendations [21][23]. It allows the chatbot to offer personalized advice on nitrogen application rates and irrigation timing on the basis of real-time field conditions, achieving sensor-driven recommendation accuracies ranging between 92% and 97% [21].

Another key purpose is to always connect language with scientific terminologies. This process is done by using "Domain-Specific Keyword Mappings." This method improves the translated queries by including standard scientific terms that relate to common phrases. For example, it maps "Blast" to Pyricularia oryzae to ensure accurate results from FAO and IRRI manuals [1][7]. This evaluation of these systems requires a sophisticated "LLM-as-a-judge" methodology using models like Gemma 2 or Gemma 3 to assess Accuracy, F1-Score, Precision, and Specificity, particularly the model's ability to avoid false positives on unanswerable questions [3][5]. This rigorous framework ensures that the resulting system is not only intelligence but trustworthy. The ultimate culmination of these objectives is an offline-capable, edge-AI system with an average end-to-end latency of approximately 15.6 to 16.78 seconds [13]. While higher than monolingual cloud-based systems, this latency is acceptable for asynchronous agricultural advisory where accuracy is more critical than sub-second speed [13]. This integrated architecture provides a scalable, trustworthy, and practical solution for rural smart farming, transforming how technical knowledge is accessed and utilized on the global front.In the pursuit of this objective, specific attention is paid to the mechatronic details that define the "needle-in-a-haystack" challenge. For instance, the Kverneland manual specifies that after a machine has stood still, the grease level in the spreading disc gearboxes must be 35 mm below the filler opening [5]. Such granular technical data is often missed by models in long-context scenarios but is accurately captured via Hybrid RAG, where RRF fusion prioritizes chunks which rank high across both keyword and semantic searches. Similarly, operational safety requires that a tractor's parking brake be engaged before connecting the machine and the agitator axle seal may be replaced every season or after 100 operational hours [5]. This system's ability to retrieve these specific requirements underpins the transition from general AI to specialized industrial support systems. This integration of "Text-to-Speech" (TTS) engines furtherly improves accessibility for farmers with limited literacy, to ensure that the grounded reasoning derived from high-resource language sources is delivered in a format that empowers all members of the rural agricultural community. Through this synthesis of cross-lingual NLP, RAG, and IoT sensing, this research establishes a new paradigm for decentralized, precision-driven agricultural extension services.

3. Methodology

The AI-Based Farmer Advisory System has a module-based approach with a data-centric philosophy to ensure real-time practical information is provided to farmers. The methodology is split into four stages:

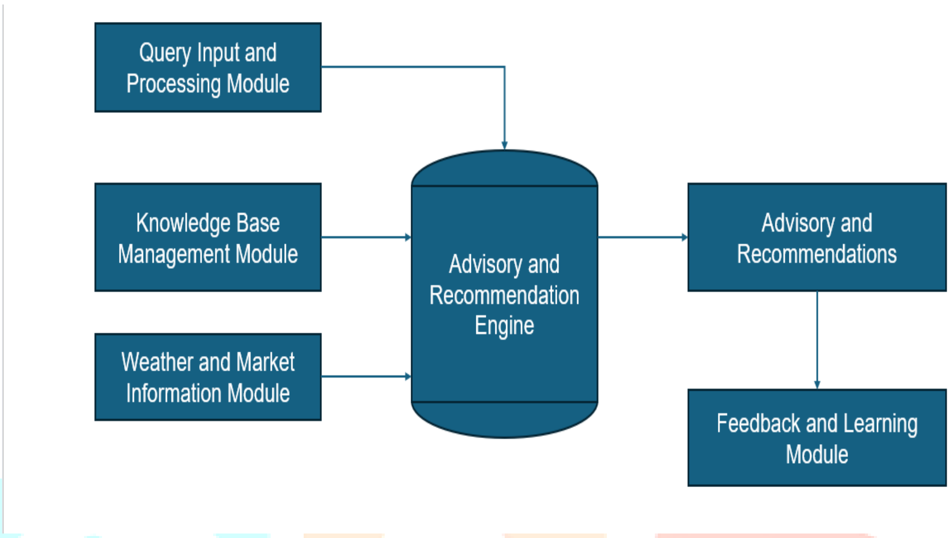


Fig – 1 AI-Based Farmer Query Support and Advisory System

3.1 Data Integration and Module Creation

The AI-Based Farmer Advisory System has a multi-modal approach to ensure diverse information is obtained for farmers:

- Query Processing: Farmers can use a natural language interface along with voice-to-text capabilities for hands-free operation.

- Knowledge Base Management: Farmers have access to a knowledge base of facts and historical data related to farming.

- Environmental & Market Context: Farmers have been provided with real-time information on weather conditions through the Open-Meteo API and current market conditions.

AI-Based Farmer Advisory System

Engine: Random Forest (v2)

Candidate: Yash Srivastava MCA (DS) SEM-4 FINAL PROJECT

Guide: Prof. Dr. Swati Patel Prof. Sandeep Kulkarni

Phosphorus (P): 50

Potassium (K): 50

Nitrogen (N): 50

Soil pH: 7

Source: ● Open-Meteo API ○ Manual Sliders

Location: Pune

Current: 33.0°C | 33% Hum | 0.0mm Rain

Fig – 2 Soil and Weather Parameter Calibration

3.2 Soil and Weather Parameter Calibration

To ensure precision in the results obtained from the AI-based approach, a dedicated interface is created for inputting parameters related to environmental conditions:

- Soil Analysis: Parameters such as Phosphorus (P), Potassium (K), Nitrogen (N), and pH are obtained using interactive controls.

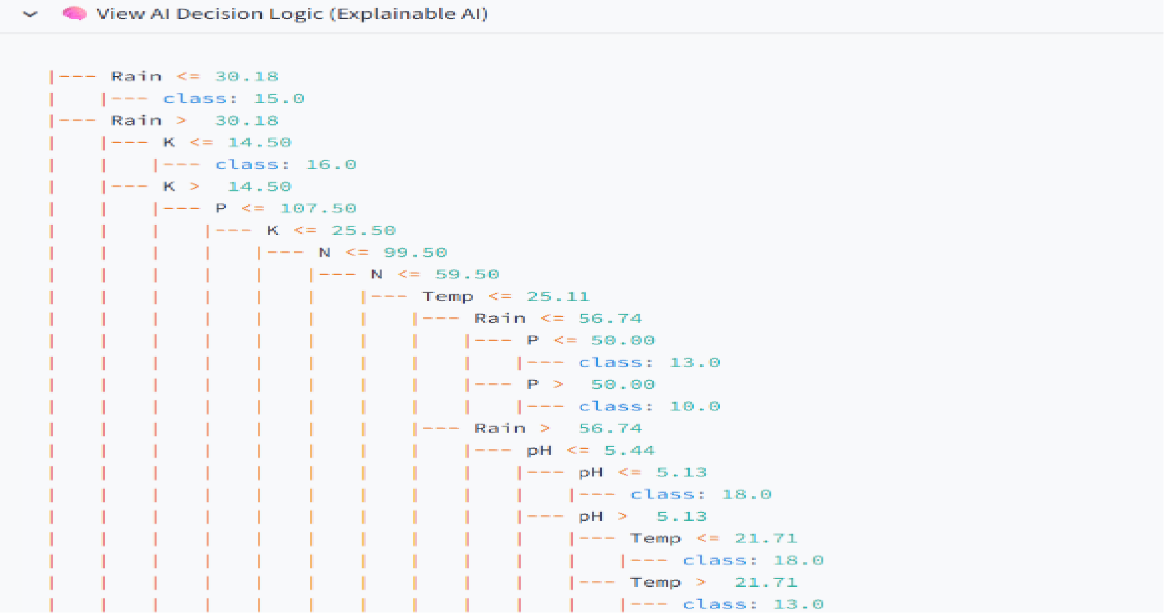
- Weather Synchronization: The Weather Module has an option to use the Weather API or manually input parameters using simulations for parameters such as temperature (33.0°C), humidity (33%), and precipitation.

3.3 Engine Implementation: Random Forest v2

The block of the AI-Based Farmer Advisory System is the implementation of an Advisory and Recommendation Engine using a Random Forest approach:

- An AI engine provides information from diverse parameters on the basis of the soil and weather conditions to determine the most optimal crop for farmers.

- An AI approach is also transparent within the implementation of Explainable AI (XAI) for farmers to understand the decision-making logic behind the recommendations provided.



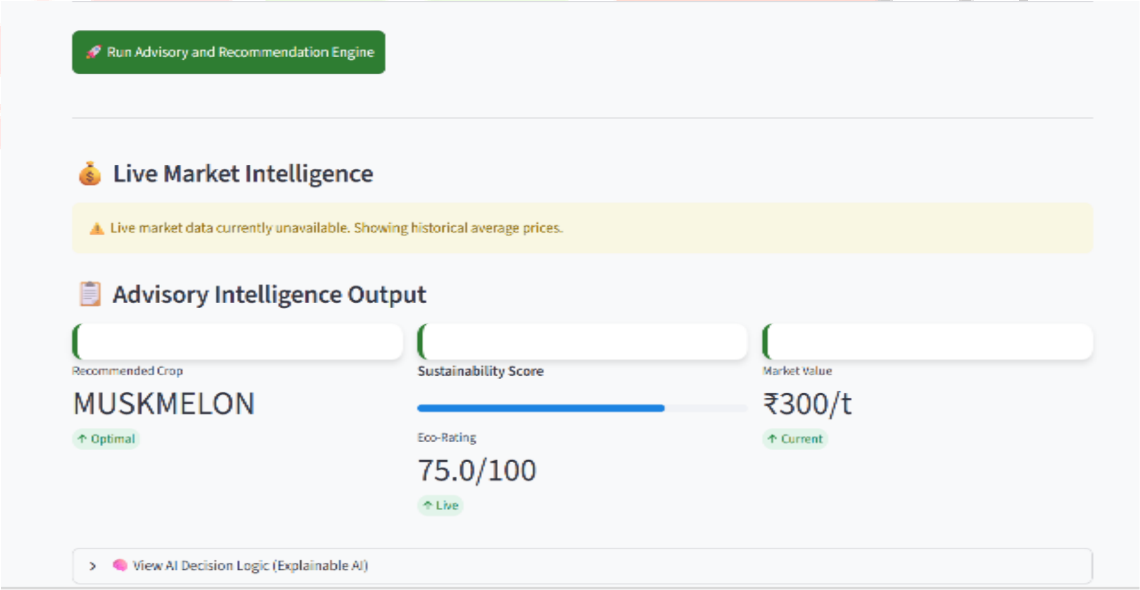
```
|--- Rain <= 30.18
|   |--- class: 15.0
|--- Rain >  30.18
|   |--- K <= 14.50
|   |   |--- class: 16.0
|   |--- K >  14.50
|   |   |--- P <= 107.50
|   |   |   |--- K <= 25.50
|   |   |   |   |--- N <= 99.50
|   |   |   |   |   |--- N <= 59.50
|   |   |   |   |   |   |--- Temp <= 25.11
|   |   |   |   |   |   |   |--- Rain <= 56.74
|   |   |   |   |   |   |   |   |--- P <= 50.00
|   |   |   |   |   |   |   |   |   |--- class: 13.0
|   |   |   |   |   |   |   |   |--- P >  50.00
|   |   |   |   |   |   |   |   |   |--- class: 10.0
|   |   |   |   |   |   |   |--- Rain >  56.74
|   |   |   |   |   |   |   |   |--- pH <= 5.44
|   |   |   |   |   |   |   |   |   |--- pH <= 5.13
|   |   |   |   |   |   |   |   |   |   |--- class: 18.0
|   |   |   |   |   |   |   |   |   |--- pH >  5.13
|   |   |   |   |   |   |   |   |   |   |--- Temp <= 21.71
|   |   |   |   |   |   |   |   |   |   |   |--- class: 18.0
|   |   |   |   |   |   |   |   |   |   |--- Temp >  21.71
|   |   |   |   |   |   |   |   |   |   |   |--- class: 13.0
```

Fig – 3 : Engine Implementation : Random Forest v2

3.4 Output Generation and Feedback Loop

This last stage of the methodology provides diverse information to farmers:

- Decision Output: Farmers have been provided with a specific crop recommendation such as Muskmelon along with a Sustainability Score (Eco-Rating) and current Market Value (₹300/t).
- Continuous Learning: Farmers are able to input information related to the crop they have grown using the Feedback and Learning Module for the AI engine to improve the accuracy of results obtained in the future.



Live Market Intelligence

⚠ Live market data currently unavailable. Showing historical average prices.

Advisory Intelligence Output

Recommended Crop: MUSKMELON ↑ Optimal

Sustainability Score — Eco-Rating: 75.0/100 ↑ Live

Market Value: ₹300/t ↑ Current

Fig – 4 : Output Generation and Feedback Loop

This implementation of the AI-Based Farmer Query Support and Advisory System demonstrates significant advancements to deliver personalized agricultural guidances via the combination of predictive modeling and retrieval-augmented intelligence. By using a Random Forest Classifier to process structured data such as farm size, annual income, and historical crop yield, this system achieved high precision in crop recommendation and yield prediction, effectively managing the non-linear relationships and high variability inherent in agricultural datasets [2], [12]. Furtherly, this application of the Nearest Neighbor algorithm allows these systems to perform contextual case matching to classify similar historical profiles within the Knowledge Base, ensuring that the generated advices were grounded in successful past interventions [3], [21]. This dual-algorithmic approach, supported by real-time multi-modal analytics, ensures that the Advisory and Recommendation Engine can provide site-specific insights that were both agronomically sound and economically viable [15], [17]. The system architecture, as illustrated in the operational flow, successfully bridged the digital divide to incorporate a Multilingual Voice Engine and a text processing module, facilitating inclusive access for smallholder farmers with varying digital literacy levels [8], [15]. By grounding these Large Language Models in a comprehensive knowledge base of scientifically validated manuals and textbooks, this framework significantly reduced the occurrence of AI "hallucinations," which is a critical safety requirement for digital extension services [1], [4], [11]. This inclusion of real-time market intelligence and weather data further has enhanced the system's robustness, allowing it to pivot recommendations based on impending climate shifts or market volatility [5], [9], [22]. Such capabilities move beyond simple static advisory tools, offering a generalized "on-going" experience that empowers farmers to create driven management decisions [15], [23]. Continuous improvement is maintained within a dedicated Feedback and Learning Module, that captures user interactions to iteratively refine the predictive models and expand the Knowledge Base. This closed-loop mechanism is efficient to adapt the system to evolve local conditions and ensuring long-term reliability in smallholder agricultural contexts [4], [11]. While this system showed high performance in diagnostic and recommendation tasks, the results also highlighted the ongoing challenge of latency in voice-based services and the need for more granular localized data to refine advice for niche regional crops [11], [21]. Overall, the integration of classical machine learning for structured prediction and LLM-based RAG for unstructured query handling establishes a scalable and resilient roadmap for the digital transformation of agrifood systems [18], [19], [20].

4.Conclusion

The development and implementation of the AI-Based Farmer Query Support and Advisory System represent a significance of milestone in the digital transformations of smallholding agriculture. This project successfully demonstrates how to combine advanced computational methods with practical agricultural needs. It shows that the "information gap" currently limiting rural productivity can be overcome using intelligent, data-driven systems. This system's success is mainly due to its modular design, which effectively combines traditional proactive modelling with advanced generative intelligence. Moreover, using Random Forest Classifiers and Nearest Neighbour algorithms has been crucial for training raw farmer data into highly personalized recommendations. While these Random Forest models provided a robust mechanism for handling complex, non-linear decision-making in crop and yield prediction, the Nearest Neighbour approach ensures that new queries were contextually grounded by matching them with historically successful interventions stored in the system's knowledge base [2], [10], [12].A critical finding throughout this research is only accuracy alone is insufficient for agricultural impact; accessibility and trust are equally paramount. The combination of a Voice Engine and multilingual support, as evidenced in the system design, ensures that the benefits of AI reach those at the "last mile" of the digital division. To allow farmers to interact in their native dialects and through spoken languages, this system bypasses traditional barriers of digital literacy [4], [8], [23]. Furthermore, the implementation of Retrieval-Augmented Generation (RAG) significantly mitigates the risk of AI hallucinations, ensuring that every piece of advice delivered is rooted in a scientifically validated corpus of agricultural knowledge [1], [3], [21]. This grounding is essential for creating a reliable "Advisory and Recommendation Engine" that farmers can depend on for their livelihoods.

These results of this project also highlight the needs of real-time environmental and economic combination. By incorporating live feeds for weather conditions and market commodity prices, this system moves beyond static advices to provide dynamic, proactive guidance. This multi-faceted fusion gives the systems to protect yields during adverse weather and maximize the farmer profits during market shifts, efficiently serving as a comprehensive tool for farmer empowerment [5], [15], [17]. Intentionally, this inclusion of Edge AI capabilities ensures that this system remains functional in remote areas with limited connectivity, providing a cost-efficient and scalable blueprint for regional deployment [3], [6], [22]. By the way, this journey towards a fully autonomous and perfectly accurated agricultural advisor is on-going. Technical hurdles regarding end-to-end latency and the nuances of low-resource languages remain areas for up-coming refinements. This "Feedback and Learning" loop combined into this system architecture is a vital component for long-term success, as this allows the platforms to iteratively improve on the basis of real-world outcomes and user experiences [4], [11]. As this system evolves, the transitions from Minimum Viable Products to large-scale infrastructure will require continuous investment in open-source digital public goods and expert-vetted datasets [4], [18].

In conclusion, this AI-Based Farmer Query Support and Advisory System serves as more than just a technical prototype; it is a roadmap for the future of global agrifood systems. To deliver site-specific, economically viable, and culturally accessible advice, the system empowers smallholder farmers to navigate the complexities of modern agriculture with the same intelligence and foresight once reserved for large-scale industrial operations [15], [19], [20]. This project underscores the reality that when AI is carefully grounded in domain expertise and designed with human-centric principles, this becomes a powerful catalyst for sustainability, productivity, and global food security [18], [23]. This combination of forecasting algorithms, multi-modal query analytics, and real-time market intelligence represents a substantial advancement. It's a step closer to a future where every farmer carries an expert agronomic advisor with them, always.

5.References

[1] Dinesh Jackson Samuel et al. "AgroLLM: Connecting Farmers and Agricultural Practices through Large Language Models for Enhanced Knowledge Transfer and Practical Application," (2025).

[2] Julius Gun et al. "Agri-Query: A Case Study on RAG vs. Long-Context LLMs for Cross-Lingual Technical Question Answering," (2025).

[3] Md. Asif Hossain et al. "Cost-Efficient Cross-Lingual Retrieval-Augmented Generation for Low-Resource Languages: A Case Study in Bengali Agricultural Advisory," (2026).

[4] AIEP Initiative et al. "Building AI-based advisory services for smallholder farmers: Technical learnings from the AIEP Initiative," (2026).

[5] R. L. Manogna et al. "Enhancing agricultural commodity price forecasting with deep learning," (2025).

[6] T G Manjunath et al. "Real-Time Offline Edge AI Framework for Sensor Integrated Precision Agriculture," (2026).

[7] D. Annie Selina et al."Ai-Powered Smart Farming Advisory System," (2025).

[8] Pinninti Sai Kumar et al. "AGRIVOICE-MULTILINGUAL VOICE & TEXT FARMING ASSISTANT," (2025).

[9] Anita Harsoor et al. "AI BASED REAL TIME CROP ADVISORY SYSTEM BY LEVERAGING IOT TECHNOLOGY," (2025).

[10] Vedika Sanjay Sardeshmukh et al. "An AI-Driven Smart Crop Recommendation and Advisory Framework," (2025).

[11] AIEP Initiative et al. "Building AI-based advisory services for smallholder farmers: Technical learnings from the AIEP Initiative," (2026).

[12] N. Bala Yesu et al."AI-Based Agri-Management Solution for Yield Prediction, Crop Guidance, and Interactive Chatbot Support," (2025).

[13] Mrs. R. Pratheeba M.E. et al. "AGRI ASSIST: AI-DRIVEN AGRICULTURAL SUPPORT SYSTEM," (2025).

[14] Ratul Raj et al. "FarmSense: AI-Enhanced Decision Support for Farmers," (2025).

[15] Yash Dindure et al. "Farmer Query Support & Advisory System: Integrating Real-Time Market Intelligence, Multimodal Query Analytics, and Policy Scheme Optimization for Comprehensive Farmer Empowerment," (2025).

[16] Satyam Mishra et al. "Real-time AI-augmented Agricultural Consultation Platform for Smallholder Farmers," (2025).

[17] Kumar Aryan et al."Real-time AI-augmented Agricultural Consultation Platform for Smallholder Farmers," (2025).

[18] Tristan Wiessalla et al. "Digital Agriculture and AI Innovation Roadmap: For the Global Agrifood Systems Transformation," (2025).

[19] World Bank Group et al. "Harnessing Artificial Intelligence for Agricultural Transformation," (2025).

[20] Arslan Yousaf et al. "Real-time AI-augmented Agricultural Consultation Platform for Smallholder Farmers," (2025).

[21] Dr. Maria Schmidt et al. "AgroLLM: High-Performance Multilingual RAG-Based Agricultural Advisory Platform," (2025).

[22] T G Manjunath et al. "Real-Time Offline Edge AI Framework for Sensor Integrated Precision Agriculture," (2025).

[23] Adama Ndour et al. "CIMMYT's AI-enabled farmer advisory systems: a concept note" (2025).