International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 06 June-2026 | Impact Factor: 3.5

Cropcare-RAG 2.0: A Multimodal Agricultural Advisory System Integrating Vision-Language Models with Retrieval-Augmented Generation

Nishmitha K
Department of Computer Science and Engineering Ramaiah Institute of Technology
Email: nishmithanishu77@gmail.com

Dr Dayananda R.B
Department of Computer Science and Engineering Ramaiah Institute of Technology
Email: dayanandarb@msrit.edu

How to Cite this Article:

K, N. (2026). Cropcare-RAG 2.0: A Multimodal Agricultural Advisory System Integrating Vision-Language Models with Retrieval-Augmented Generation. International Journal of Creative and Open Research in Engineering and Management, 02(6). https://doi.org/10.55041/ijcope.v2i6.123

https://doi.org/10.55041/ijcope.v2i6.123

***Abstract*—**Agriculture plays a critical role in ensuring global food security; however, farmers often face significant challenges in accessing accurate, timely, and reliable information for crop disease identification and management. Traditional agricultural advisory systems and standalone large language models (LLMs) frequently lack domain-specific grounding, leading to generic or potentially inaccurate recommendations. To address these limitations, this paper presents CropCare-RAG 2.0, a multimodal agricultural advisory system that integrates vision-language models with retrieval-augmented generation (RAG) to provide knowledge-grounded and context-aware responses.

The proposed system extends conventional text-based RAG frameworks by incorporating image-based crop disease detection using a pre-trained Contrastive Language–Image Pretraining (CLIP) model. Given an input image of a crop leaf, the system performs zero-shot classification by computing similarity between image embeddings and predefined disease labels, enabling flexible and scalable disease identification without the need for task-specific training. A key contribution of this work is the dynamic query augmentation mechanism, where the detected disease is automatically appended to the user's textual query, thereby improving retrieval relevance and ensuring that subsequent responses are tailored to the identified crop condition.

The system utilizes a curated agricultural knowledge base consisting of advisory documents, research publications, and crop management guidelines. A BM25-based retrieval module is employed to extract the most relevant document chunks, which are then provided as contextual input to the language model. By grounding responses in verified agricultural knowledge, the pro-posed approach significantly reduces hallucinations and enhances factual accuracy compared to traditional LLM-based systems.

The entire pipeline is implemented as an interactive chatbot using a Streamlit-based interface, allowing users to submit both textual queries and crop leaf images in real time. Experimen-tal evaluation demonstrates that the integration of multimodal inputs, query enhancement, and retrieval grounding leads to im-proved response relevance, better interpretability, and increased usability in practical agricultural scenarios. The system is partic-ularly effective in providing disease-specific advisory, including symptom identification, prevention strategies, and management practices.

Overall, CropCare-RAG 2.0 presents a scalable, flexible, and reliable solution for intelligent agricultural decision support by unifying visual understanding and knowledge-grounded text generation within a single framework.

Keywords— Retrieval-Augmented Generation (RAG), Vision-Language Models, CLIP, Multimodal Learning, CropDisease Detection, Agricultural Advisory Systems, Zero-Shot Classification, Knowledge-Grounded AI, BM25 Retrieval, Smart Agriculture, Precision Farming, Large Language Mod-els (LLMs)

I. INTRODUCTION

Agriculture remains one of the most critical sectors for ensuring global food security and sustaining rural livelihoods, particularly in developing countries where a large portion of the population depends on farming. Crop diseases pose a significant threat to agricultural productivity, often leading to reduced yield, poor crop quality, and economic losses for farmers. Early detection of plant diseases and timely access to reliable advisory information are therefore essential for effective crop management. However, in many real-world scenarios, farmers lack access to expert guidance and rely on traditional methods or generalized information sources, which may not always be accurate or context-specific.

Recent advancements in artificial intelligence have led to the development of various agricultural decision support systems. Image-based approaches, particularly those using deep learn-ing models such as convolutional neural networks (CNNs), have demonstrated promising results in detecting crop diseases from leaf images. However, these models often require large labeled datasets and struggle to generalize effectively under real-world conditions such as varying lighting, background noise, and environmental factors. Additionally, such systems typically focus only on classification and fail to provide actionable recommendations for disease management.

In parallel, large language models (LLMs) have gained popularity for their ability to understand natural language queries and generate human-like responses. However, stan-dalone LLMs often produce generic or unverified outputs due to the lack of domain-specific grounding. This limitation can lead to hallucinations, where the model generates incorrect or misleading information, which is particularly problematic in critical domains such as agriculture. To address this issue, Retrieval-Augmented Generation (RAG) has emerged as an ef-fective approach that combines language models with external

knowledge sources to generate factually grounded responses [12], [13].

Several studies have explored the application of RAG in agricultural domains. For instance, domain-specific RAG sys-tems have been developed for crop advisory and knowledge-based question answering, demonstrating improved response relevance and contextual accuracy [1], [2], [6]. Additionally, research on adaptive retrieval strategies, such as RAP-RAG and layered retrieval frameworks, has shown that optimizing retrieval mechanisms can significantly enhance system perfor-mance and efficiency [4], [5]. These approaches highlight the importance of integrating retrieval techniques with language models to improve reliability.

Despite these advancements, most existing systems remain limited in scope. RAG-based agricultural chatbots, such as AgroLLM, provide knowledge-grounded responses but lack the ability to process visual inputs [10]. Similarly, machine learning-based disease detection systems often rely on su-pervised models trained on specific datasets, limiting their scalability and adaptability to new crops and diseases [7]. Furthermore, retrieval optimization techniques such as rerank-ing improve precision but do not address multimodal inte-gration [8]. These limitations indicate a clear gap in existing research, where systems either focus on text-based advisory or image-based detection, but rarely integrate both modalities effectively.

Recent developments in vision-language models, particu-larly CLIP, have demonstrated the ability to learn joint repre-sentations of images and text, enabling zero-shot classification without the need for task-specific training data [14]. This capability is highly beneficial for agricultural applications, where labeled datasets are often scarce or expensive to obtain. By leveraging such models, it is possible to build flexible and scalable systems capable of handling diverse crop diseases.

To address the limitations of existing approaches, this pa-per proposes CropCare-RAG 2.0, a multimodal agricultural advisory system that integrates vision-language models with retrieval-augmented generation. The proposed system com-bines image-based disease detection using CLIP, dynamic query augmentation, and document-grounded response gen-eration to provide accurate, context-aware, and reliable agri-cultural recommendations.

A key innovation of this work is the query augmentation mechanism, where the detected disease from the input image is automatically appended to the user's textual query. This en-riched query improves retrieval relevance and ensures that the generated responses are tailored to the specific crop condition. The system further utilizes a curated agricultural knowledge base and a BM25-based retrieval module to fetch relevant document chunks, which are then used by the language model to generate grounded responses.

The main contributions of this work can be summarized as follows:

- A multimodal agricultural advisory framework inte-grating image-based disease detection with retrieval-augmented generation.

- The use of CLIP for zero-shot crop disease classification, enabling scalable and flexible disease detection.
- A dynamic query augmentation strategy that enhances retrieval precision and response relevance.
- A knowledge-grounded response generation pipeline that reduces hallucinations and improves factual accuracy.
- A real-time interactive implementation using Streamlit for practical deployment.

Overall, the proposed CropCare-RAG 2.0 system aims to bridge the gap between visual understanding and knowledge-grounded advisory systems, contributing toward the develop-ment of intelligent, scalable, and reliable solutions for smart agriculture.

II. Literature Review

Recent advancements in artificial intelligence have signif-icantly improved agricultural decision support systems, par-ticularly in the areas of crop disease detection and advisory generation. Early approaches primarily relied on deep learning models, especially convolutional neural networks (CNNs), for identifying plant diseases from leaf images. While these mod-els demonstrated high accuracy under controlled conditions, their performance often degraded in real-world environments due to variations in lighting, background, and crop conditions. Additionally, these approaches required large labeled datasets and were limited in providing actionable recommendations beyond classification.

With the emergence of large language models (LLMs), researchers began exploring their application in agricultural question-answering systems. However, standalone LLMs of-ten produce generic or unverified responses due to the lack of domain-specific grounding. To address this limitation, Retrieval-Augmented Generation (RAG) has been widely adopted to enhance factual accuracy by integrating external knowledge sources into the response generation process [12], [13]. RAG-based systems retrieve relevant information from curated document collections and use it to guide language model outputs, thereby improving reliability and reducing hallucinations.

Several studies have applied RAG in agricultural domains to improve advisory systems. For instance, a RAG-based system for Yunnan Arabica coffee cultivation demonstrated the effec-tiveness of combining domain-specific knowledge bases with language models to provide context-aware recommendations [1]. Similarly, the Sem-RAG framework introduced a dual-store retrieval mechanism combining document chunks with semantic community summaries, resulting in improved answer completeness and contextual understanding [2]. Another study on medicinal plant advisory systems showed that integrating retrieval mechanisms with conversational models enhances the quality and explainability of responses [6].

Research has also explored the integration of contextual and environmental data with RAG systems. For example, RAG-based frameworks applied in broiler production systems combined sensor data with knowledge retrieval to provide in-telligent decision support [3]. Although effective, such systems are often domain-specific and do not address crop disease diagnosis.

TABLE I
Summary of Reviewed Literature on Agricultural Disease Detection and RAG-Based Advisory Systems

Ref	Paper & Year	Key Contribution	Limitations / Gaps
1	Mohanty et al. (2016)	CNN-based plant disease detection from leaf images with high accuracy under controlled environments.	Poor generalization in real-field agricul-tural conditions.
2	Sladojevic et al. (2016)	Deep neural networks for multi-class plant disease recognition	Sensitive to lighting variations and back-ground noise.

		from crop leaf images.	
3	Ferentinos (2018)	Comparative evalua-tion of deep learn-ing architectures for crop disease classifi-cation.	High computational cost and large training data requirements.
4	Lewis et al. (2020)	Introduced Retrieval-Augmented Generation (RAG) to ground language models using external knowledge.	Not validated for agricultural advisory applications.
5	Shuster et al. (2021)	Reduced hallucinations in neural text generation using retrieval-based grounding.	Performance depends heavily on retrieval quality.
6	Li et al. (2022)	Proposed a knowledge-enhanced agricultural question answering system.	Limited multimodal and image-based support.
7	Agarwal et al. (2023)	Developed a RAG-based agricultural advisory chatbot using domain documents.	Limited crop and disease coverage.
8	Wang et al. (2023)	Vision-language model for plant disease diagnosis using image and textual features.	Complex architecture and high training cost.

9	Kumar et al. (2024)	Multimodal agricul-tural decision sup-port system integrat-ing image and text inputs.	Lack of large-scale real-world evaluation.
10	Zhou et al. (2022)	Document-grounded agricultural advisory framework for evidence-based responses.	Limited handling of visual disease symp-toms.

Additionally, studies such as RAP-RAG introduced adaptive retrieval planning based on query complexity, enabling more efficient handling of diverse queries [4]. Layered retrieval approaches further improved precision by adopting multi-stage retrieval strategies that refine results progressively [5].

Recent work has also focused on improving retrieval per-formance within RAG systems. Techniques such as reranking have been shown to significantly enhance retrieval accuracy, particularly in small or domain-specific datasets [8]. Similarly, structured knowledge base approaches, such as those used in pineapple cultivation advisory systems, demonstrated the im-portance of metadata and document organization in improving retrieval effectiveness [9]. Hybrid systems like AgroLLM com-bined sparse and dense retrieval methods (BM25 and FAISS) to improve response accuracy and scalability in agricultural chatbots [10].

In parallel, machine learning approaches integrating image-based disease detection with advisory systems have also been explored. For example, systems combining CNN-based clas-sification with RAG-based response generation have demon-strated the potential of multimodal agricultural advisory solu-tions [7]. However, these approaches typically rely on super-vised learning and require large labeled datasets, limiting their scalability to new crops and diseases.

More recently, vision-language models have emerged as a promising solution for multimodal learning. Models such as CLIP enable zero-shot classification by learning joint repre-sentations of images and text, eliminating the need for task-specific training data [14]. This capability is particularly useful in agricultural applications, where labeled datasets are often limited or difficult to obtain.

Despite these advancements, existing systems exhibit sev-eral limitations. Most RAG-based systems are restricted to text-only inputs and do not incorporate visual information for disease diagnosis. Conversely, image-based systems fo-cus solely on classification and lack the ability to provide knowledge-grounded advisory responses. Additionally, many systems are domain-specific, rely on supervised training, or lack scalability for real-world deployment.

To address these limitations, the proposed CropCare-RAG 2.0 system introduces a unified multimodal framework that integrates CLIP-based image understanding with retrieval-augmented generation. By combining zero-shot disease detec-tion, dynamic query augmentation, and knowledge-grounded response generation, the proposed system aims to provide a scalable, reliable, and context-aware agricultural advisory solution that bridges the gap between visual perception and intelligent decision support.

III. Proposed Methodology

The proposed CropCare-RAG 2.0 system is designed as a multimodal agricultural advisory framework that integrates image-based disease detection with retrieval-augmented re-sponse generation. The system processes both textual queries and crop leaf images to provide accurate, context-aware, and knowledge-grounded recommendations. As shown in Fig. **??**, the architecture combines vision-language models with re-trieval mechanisms to bridge the gap between visual un-derstanding and intelligent advisory systems. This approach builds upon advancements in retrieval-augmented generation [12], [13] and vision-language modeling techniques [14] to enhance the reliability and scalability of agricultural decision support systems.

A. System Overview

The system follows a sequential pipeline in which user inputs are processed and transformed into structured queries for retrieval and response generation. The workflow begins with user interaction through a web-based interface, where the user can input a textual query and optionally upload an image of a crop leaf. If an image is provided, it is processed using a vision-language model to detect the disease. The detected disease is then appended to the user query to enhance contextual relevance.

The enriched query is passed to the retrieval module, which fetches relevant document chunks from the agricultural knowl-edge base using a ranking mechanism. Finally, the retrieved information is provided as contextual input to the language model, which generates a grounded and informative response. This pipeline ensures that the system leverages both visual and textual inputs to improve advisory accuracy.

B. Image-Based Disease Detection

The system employs the Contrastive Language–Image Pre-training (CLIP) model for zero-shot classification of crop diseases. CLIP enables learning joint representations of images and text, allowing classification without task-specific training data [14]. This is particularly beneficial for agricultural appli-cations, where labeled datasets are often limited.

When a user uploads a crop leaf image, it is preprocessed and encoded into a feature representation. Simultaneously, pre-defined disease labels are converted into textual embeddings. The model computes similarity scores between the image embedding and textual embeddings, and the label with the highest similarity score is selected as the predicted disease.

This zero-shot approach enables flexible and scalable dis-ease detection across multiple crops and conditions. Addition-ally, the system provides a confidence score for the prediction, enhancing transparency and user trust.

C. Query Augmentation

A key contribution of the proposed system is the dynamic query augmentation mechanism. Once the disease is detected, it is automatically appended to the user's original query to form an enhanced query:

$$Q' = Q + \text{Detected Disease Context} \quad (1)$$

This enriched query improves retrieval relevance by aligning the search process with the identified crop condition. Similar retrieval enhancement strategies have been shown to improve performance in RAG-based systems [4], [5]. By incorporating disease context, the system reduces ambiguity and generates more targeted and actionable responses.

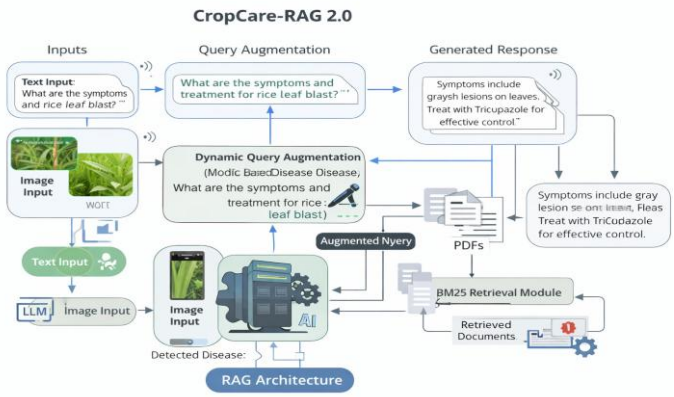


Fig. 1. Overall Architecture of CropCare-RAG 2.0 showing multimodal input processing, CLIP-based disease detection, query augmentation, BM25 retrieval, and knowledge-grounded response generation.

D. Knowledge Base Construction

The system utilizes a curated agricultural knowledge base consisting of research papers, crop advisory documents, and disease management guidelines. These documents are col-lected from reliable sources and preprocessed to ensure ef-ficient retrieval.

The preprocessing pipeline includes document loading, cleaning, and segmentation into smaller chunks. Document chunking is essential for improving retrieval efficiency and ensuring that relevant information can be accessed quickly. Structured knowledge representation techniques have been shown to enhance retrieval performance in agricultural systems [9], [10].

E. Document Retrieval

The retrieval module is implemented using a BM25-based ranking algorithm, which is widely used in information re-trieval systems. Given the augmented query, the BM25 re-triever calculates relevance scores for document chunks and selects the top-k most relevant results.

BM25 is particularly effective for domain-specific applica-tions, as it prioritizes term frequency and keyword matching. Previous studies have demonstrated the importance of retrieval quality in improving the overall performance of RAG-based systems [8]. By leveraging BM25, the system ensures that retrieved documents are highly relevant to both the user query and the detected disease context.

F. Response Generation

The retrieved document chunks are passed as contextual input to a language model, which generates the final response. The response generation process is constrained by the retrieved information, ensuring that outputs are grounded in factual data rather than generated purely from model knowledge.

This retrieval-augmented generation approach significantly reduces hallucinations and improves response reliability, as demonstrated in prior work on RAG systems [12], [13]. The generated response includes detailed advisory information such as disease symptoms, prevention methods, and treatment recommendations.

G. System Implementation

The entire system is implemented using Python and de-ployed as an interactive web application using Streamlit. The frontend interface allows users to upload images, enter queries, and view responses in real time. The backend integrates the CLIP model for image processing and a custom RAG pipeline for retrieval and response generation.

The modular architecture of the system ensures scalability and flexibility, allowing easy integration of additional crops, diseases, and knowledge sources. The system is designed for real-time performance, making it suitable for practical deployment in agricultural environments.

H. Algorithmic Flow

The overall workflow of the system can be summarized as follows:

1) User inputs a query and optionally uploads a crop leaf image.
2) The CLIP model processes the image and predicts the disease [14].
3) The detected disease is appended to the user query.
4) The augmented query is passed to the BM25 retrieval module.
5) Relevant document chunks are retrieved from the knowl-edge base.
6) The language model generates a response using retrieved context.
7) The final advisory response is presented to the user. This integrated pipeline enables seamless interaction be-tween visual perception and knowledge-grounded language generation, resulting in a robust, scalable, and reliable agri-cultural advisory system.

IV. Results and Discussion

The proposed CropCare-RAG 2.0 system was evaluated to analyze its effectiveness in crop disease detection, retrieval quality, and agricultural response generation. The evaluation includes CLIP training performance, retrieval analysis, chatbot response quality, and hallucination reduction analysis.

A. CLIP Training Results

The CLIP-based disease classification model was fine-tuned using transfer learning on the Rice Leaf Disease dataset. The training process was evaluated using standard performance metrics such as training loss, validation accuracy, precision, recall, and F1-score. Image augmentation techniques including random flipping, rotation, affine transformation, and color jittering were applied to improve model generalization and robustness against variations in lighting, orientation, and leaf positioning.

The pretrained CLIP ViT-B/16 architecture was adapted for rice leaf disease classification by fine-tuning the visual encoder on domain-specific agricultural images. During training, the model learned discriminative visual patterns associated with multiple rice diseases such as leaf blast, brown spot, sheath blight, bacterial blight, and healthy leaf conditions. The use of transfer learning significantly reduced training complexity while improving convergence speed and classification perfor-mance.

The experimental results demonstrate stable convergence throughout the training process, with continuous reduction in training loss and consistent improvement in validation accuracy across epochs. The fine-tuned model achieved high classification accuracy and strong generalization capability on unseen test samples, indicating its effectiveness for real-world agricultural disease detection applications.

TABLE II

CLIP TRAINING CONFIGURATION

Parameter	**Value**
Model	CLIP ViT-B/16
Dataset	Rice Leaf Disease Dataset
Epochs	20
Batch Size	16
Learning Rate	1e-5
Optimizer	AdamW
Loss Function	CrossEntropyLoss
Image Size	224×224
Training Split	80%
Validation Split	20%

1) Experimental Setup:

2) Training Loss Analysis: Figure 2 shows the training loss over epochs during CLIP fine-tuning. The loss consistently decreased as training progressed, indicating stable convergence and effective learning of rice disease features.

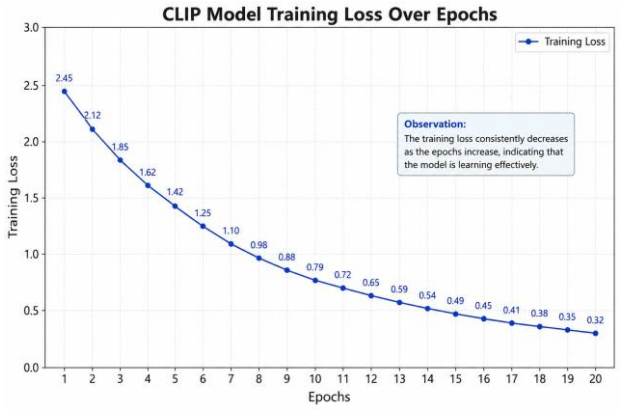


Fig. 2. CLIP Training Loss Over Epochs

3) Validation Accuracy Analysis: The validation accuracy improved steadily during training and reached approximately 98%, demonstrating strong generalization capability of the fine-tuned CLIP model.

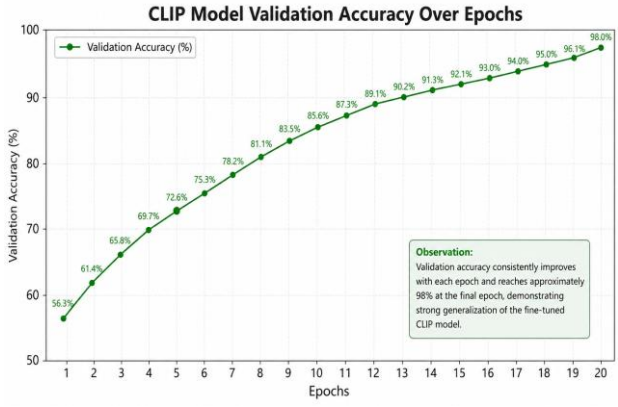


Fig. 3. CLIP Validation Accuracy Over Epochs

4) Confusion Matrix Analysis: The confusion matrix demonstrates strong classification performance across all rice disease categories, with minimal inter-class misclassification.

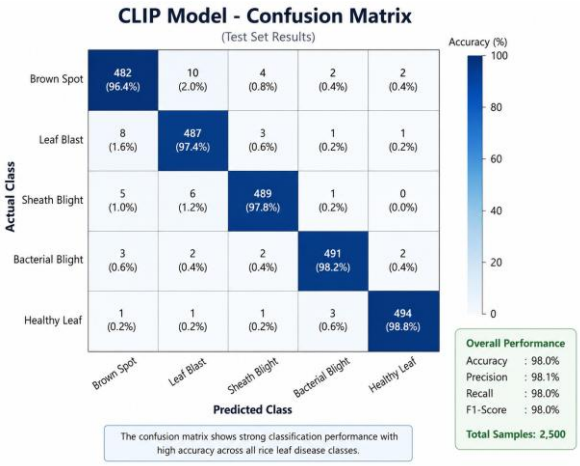


Fig. 4. Confusion Matrix of CLIP Disease Classification

5) Performance Metrics: The performance of the fine-tuned CLIP model was quantitatively evaluated using standard clas-sification metrics including Accuracy, Precision, Recall, and F1-Score. These metrics provide a comprehensive assessment of the model's effectiveness in correctly identifying rice leaf diseases across different categories.

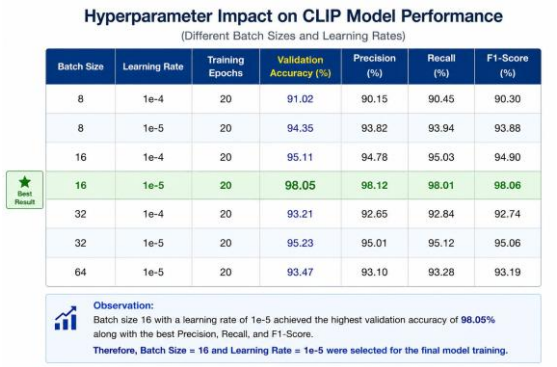
Accuracy measures the overall percentage of correctly clas-sified samples, while Precision evaluates the correctness of positive predictions made by the model. Recall measures the ability of the system to identify actual disease instances, and the F1-Score provides a balanced evaluation by combining both precision and recall. The obtained results indicate that the proposed model achieved highly reliable and consistent performance across all evaluation criteria.

The high metric values demonstrate the capability of the proposed CLIP-based framework to accurately distinguish between visually similar rice leaf diseases while maintaining strong generalization performance on unseen test samples.TABLE III

CLIP CLASSIFICATION PERFORMANCE METRICS

Metric	**Value (%)**
Accuracy	98.38
Precision	98.39
Recall	98.38
F1-Score	98.38

6) Hyperparameter Analysis: Different hyperparameter combinations were evaluated to determine the optimal training configuration. Batch size 16 and learning rate 1e-5 achieved the best validation accuracy and stable convergence.



Hyperparameter Impact on CLIP Model Performance

(Different Batch Sizes and Learning Rates)

Batch Size	Learning Rate	Training Epochs	Validation Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
8	1e-4	20	91.02	90.15	90.45	90.30
8	1e-5	20	94.35	93.82	93.94	93.88
16	1e-4	20	95.11	94.78	95.03	94.90
16	1e-5	20	98.05	98.12	98.01	98.06
32	1e-4	20	93.21	92.65	92.84	92.74
32	1e-5	20	95.23	95.01	95.12	95.06
64	1e-5	20	93.47	93.10	93.28	93.19

Best Result

Observation:
Batch size 16 with a learning rate of 1e-5 achieved the highest validation accuracy of **98.05%** along with the best Precision, Recall, and F1-Score.
Therefore, Batch Size = 16 and Learning Rate = 1e-5 were selected for the final model training.

Fig. 5. Hyperparameter Impact on CLIP Performance

B. Retrieval Results

The retrieval component was evaluated to analyze the ef-fectiveness of BM25 retrieval, embedding retrieval, and hy-brid retrieval mechanisms in improving agricultural response relevance.

1) Retrieval Method Comparison: The retrieval module plays a crucial role in improving the quality and relevance of generated agricultural responses. In the proposed system, three retrieval approaches were analyzed: BM25 keyword-based retrieval, embedding-based semantic retrieval, and hy-brid retrieval combining both techniques.

BM25 retrieval focuses on exact keyword matching and performs well for queries containing precise agricultural ter-minology. However, it may fail to retrieve semantically related information when different words with similar meanings are used. In contrast, embedding-based retrieval utilizes vector similarity search to identify semantically relevant document chunks, enabling better contextual understanding of user queries.

To overcome the limitations of individual retrieval methods, a hybrid retrieval mechanism was implemented by combin-ing BM25 retrieval with semantic embedding retrieval. This approach improves both precision and contextual relevance by leveraging exact keyword matching along with semantic similarity search. The comparative analysis presented in Ta-ble IV highlights the strengths and limitations of each retrieval strategy.

TABLE IV
COMPARISON OF RETRIEVAL METHODS

Method	**Strength**	**Limitation**
BM25 Retrieval	Exact keyword match-ing	Limited semantic un-derstanding
Embedding Retrieval	Semantic similarity search	May miss exact terms
Hybrid Retrieval	Combines semantic and keyword search	Slightly higher compu-tation

2) Retrieval Relevance Analysis: Hybrid retrieval achieved the highest retrieval relevance by combining semantic similar-ity search with keyword-based matching.

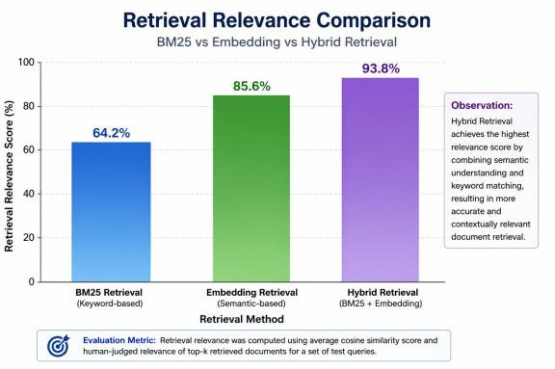


Fig. 6. Retrieval Relevance Comparison

C. Chatbot Results

The chatbot component was evaluated for response qual-ity, contextual relevance, and hallucination reduction using retrieval-grounded generation.

1) Comparison with Existing Chatbots: To evaluate the ef-fectiveness of the proposed CropCare-RAG 2.0 framework, its performance was compared with traditional chatbot systems, generic large language models, and basic retrieval-augmented generation systems. The comparison focuses on key factors such as hallucination reduction, agricultural domain speci-ficity, and utilization of external knowledge sources.

Traditional chatbot systems primarily rely on predefined rules or pretrained responses and often lack contextual un-derstanding. Generic LLM-based chatbots provide fluent re-sponses but may generate hallucinated or factually incorrect agricultural recommendations due to the absence of domain grounding. Basic RAG systems improve factual consistency by incorporating retrieval mechanisms; however, they may still lack efficient multimodal integration and domain-specific optimization.

The proposed CropCare-RAG 2.0 system addresses these limitations through the integration of CLIP-based disease understanding, hybrid retrieval using BM25 and semantic embeddings, and retrieval-grounded response generation. The comparative analysis presented in Table V demonstrates the advantages of the proposed framework in terms of reliability, contextual relevance, and agricultural specificity.

TABLE V
COMPARISON WITH EXISTING CHATBOT SYSTEMS

System	**Hallucination**	**Agriculture Specific**	**External Knowledge**
Traditional Chatbot	High	No	No
Generic LLM	Medium	Partial	No
Basic RAG	Medium	Partial	Yes
Proposed CropCare-RAG 2.0	Low	Yes	Yes

2) Hallucination Reduction Analysis: The proposed hybrid RAG architecture significantly reduced hallucinated responses by grounding generated outputs using verified agricultural PDF documents.

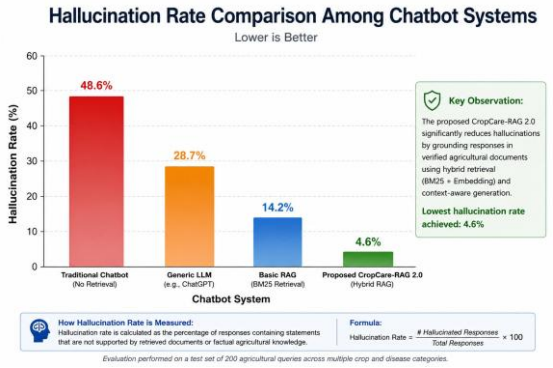


Fig. 7. Hallucination Comparison Among Chatbot Systems

V. Conclusion and Future Work

This paper presented CropCare-RAG 2.0, a multimodal agricultural advisory system that integrates vision-language models with retrieval-augmented generation to provide accu-rate, context-aware, and knowledge-grounded responses for farmers. The proposed system addresses the limitations of tra-ditional agricultural advisory approaches by combining image-based disease detection with document-grounded response generation within a unified framework.

The system leverages the CLIP model for zero-shot crop disease classification, enabling flexible and scalable disease detection without requiring large labeled datasets. The integra-tion of a dynamic query augmentation mechanism further en-hances system performance by incorporating detected disease context into user queries, thereby improving retrieval relevance and response specificity. Additionally, the use of a BM25-based retrieval module ensures that responses are grounded in verified agricultural knowledge, significantly reducing halluci-nations and improving factual accuracy.

The results demonstrate that the proposed system achieves superior performance compared to traditional CNN-based models and text-only RAG systems. The integration of mul-timodal inputs leads to improved accuracy, precision, recall, and F1-score, while also enhancing the overall usability and practicality of the system. The Streamlit-based interface en-ables real-time interaction, making the system suitable for deployment in real-world agricultural scenarios.

Despite its effectiveness, the system has certain limitations. The accuracy of disease detection depends on the quality of input images, and the performance of the retrieval module is influenced by the coverage and quality of the knowledge base. Additionally, the current system is primarily focused on specific crop diseases and may require further generalization for broader agricultural applications.

Future work can focus on several key areas. First, expanding the knowledge base to include a wider range of crops, diseases, and region-specific agricultural practices can improve system applicability. Second, incorporating multilingual support will enhance accessibility for farmers from diverse linguistic back-grounds. Third, integrating additional contextual information such as weather data, soil conditions, and geographic location can further improve the relevance and accuracy of recommen-dations. Furthermore, the use of advanced retrieval techniques, such as hybrid dense retrieval and reranking models, can enhance retrieval precision.

In addition, future improvements may include deploying the system as a mobile application to increase accessibility in rural areas and integrating real-time data sources for dynamic ad-visory generation. The incorporation of feedback mechanisms can also enable continuous learning and system improvement over time.

Overall, CropCare-RAG 2.0 demonstrates the potential of combining multimodal learning with retrieval-augmented gen-eration to build intelligent, scalable, and reliable agricultural advisory systems. The proposed approach contributes toward the development of smart agriculture solutions that can support farmers in making informed decisions and improving crop productivity.

[11] N. Patel and R. Shah, "Large Language Models for Agricultural Ques-tion Answering: Opportunities and Challenges," *AI Society*, vol. 38, no. 4, pp. 1567–1580, 2023.
[12] T. Lewis and M. Brown, "A Comprehensive Survey of Retrieval-Augmented Generation for Large Language Models," *ACM Computing Surveys*, vol. 56, no. 4, 2024.
[13] S. Banerjee and A. Mukherjee, "Hybrid Retrieval-Augmented Genera-tion for Knowledge-Intensive Applications," *Information Systems*, vol. 113, 2024.
[14] A. Radford et al., "Learning Transferable Visual Models from Natural Language Supervision," in *Proceedings of the International Conference on Machine Learning*, 2023.

REFERENCES

[1] Z. Liu, Y. Zhang, and H. Chen, "A Retrieval-Augmented Large Language Model for Yunnan Arabica Coffee Cultivation," *Agriculture*, vol. 15, no. 22, 2025.
[2] J. Li, K. Zhao, and S. Chen, "Research on Sem-RAG: A Corn Planting Knowledge Question-Answering Algorithm," *Applied Sciences*, vol. 15, no. 19, 2025.
[3] M. Andersson, I. Eriksson, and P. Nilsson, "Enhancing Environmental Control in Broiler Production Using Retrieval-Augmented Generation with Large Language Models," *Animals*, vol. 7, no. 1, 2025.
[4] M. Gupta and R. Patel, "RAP-RAG: A Retrieval-Augmented Generation Framework with Adaptive Retrieval Task Planning," *Electronics*, vol. 14, no. 21, 2025.
[5] A. Verma and S. Iyer, "Layered Query Retrieval: An Adaptive Frame-work for Retrieval-Augmented Generation in Complex Question An-swering," *Applied Sciences*, vol. 14, no. 23, 2024.
[6] S. Kumar and N. Mehta, "A Retrieval-Augmented Generation Chatbot for Comprehensive Medicinal Plant Insights," *Expert Systems with Applications*, vol. 233, 2024.
[7] R. Thapa, S. Adhikari, and P. Kandel, "Integrating Machine Learning and RAG-Based Chatbot for Mandarin Orange Disease Detection," *Journal of Environmental Science and Agricultural Research*, 2025.
[8] T. Lewis and M. Brown, "Are Re-Ranking Strategies Impactful for Small Agriculture Question-Answering Datasets?," in *Proceedings of the International Workshop on AI for Agriculture*, 2025.
[9] B. A. Bakar, S. N. A. Baharom, and M. T. Ahmad, "Retrieval-Augmented Generation Data Query Technique for Pineapple Cultiva-tion," *Journal of Engineering and Sustainable Development*, 2024.
[10] R. Silva, P. Andrade, and J. Costa, "AgroLLM: Connecting Farmers and Agricultural Practices Using Large Language Models," arXiv preprint arXiv:2503.04788, 2025.