

OCTOBER 11, 2025

# RAG- enhanced global assistant for agricultural pest and diseases management

## Group C6:

Gangula Venkata Raja Vineela – 245322733143

Gunti Sai Pranitha – 245322733148

Rishit Senapati – 245322733176

## Under the guidance of:

Mrs. KJ Archana

Assistant Professor, CSE

Neil Gogte Institute of Technology

# Abstract

- Agricultural productivity is often limited by crop pests and diseases that require timely, accurate identification.
- Recent large multimodal models (LMMs) like Agri-LLaVA show promise in automating pest and disease detection but face challenges such as hallucinated outputs and poor generalization due to limited datasets.
- To address this, we propose an enhanced agricultural assistant that integrates Retrieval-Augmented Generation (RAG) to ground responses in verified agricultural knowledge and reduce misinformation.
- We also expand and diversify the dataset with real-world crop, pest, and disease images from varied sources and conditions to improve accuracy and coverage.
- Experiments show that our approach outperforms Agri-LLaVA, demonstrating that combining RAG with data enrichment creates a more reliable, scalable, and effective AI system for crop disease diagnosis and management.



Literature Survey							
Title	Author(s)	Year	Objective	Methodology	Dataset	Findings	Relevance to the Project
Flamingo: A Visual Language Model for Few-Shot Learning	Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, et al.	2022	Develop a multimodal model capable of few-shot visual-language reasoning.	Combines a visual encoder with a frozen language model using cross-attention.	MultiModal MassiveWeb, ALIGN, LTIP, VTP	Demonstrated strong few-shot learning and reasoning across modalities.	Foundation for multimodal reasoning and image-text understanding in agriculture.
InstructBLIP: Towards General-Purpose Vision-Language Models with Instruction Tuning	Wenliang Dai, Junnan Li, Dongxu Li, et al.	2024	Enable instruction-following behavior in vision-language models.	Trains a Query Transformer on 26 multimodal datasets with instruction-tuning.	Multiple public VL datasets	Achieved strong instruction following and generalization.	Supports conversational and diagnostic tasks in crop disease identification.

Title	Author(s)	Year	Objective	Methodology	Dataset	Findings	Relevance to the Project
Allava: Harnessing GPT-4V-Synthesized Data for a Lite Vision-Language Model	Guiming Hardy Chen, Shunian Chen, Ruifei Zhang, et al.	2024	Build a lightweight multimodal model using synthetic GPT-4V data.	Uses GPT-4V-generated captions for efficient training.	LAION, Vision FLAN	Achieved high performance with small compute resources.	Ideal for low-resource agricultural deployments and mobile diagnostics.
LLaVA-Med: Training a Large Language-and-Vision Assistant for Biomedicine in One Day	Chunyuan Li, Cliff Wong, Sheng Zhang, et al.	2024	Adapt LLaVA for specialized biomedical reasoning tasks.	Fine-tunes LLaVA using GPT-4V-refined PubMedVision dataset.	PubMedVision	Rapid domain adaptation with improved visual reasoning.	Template for adapting LLaVA to agricultural disease domains.
PlantVillage: Open Access Repository of Plant Health Images	David Hughes, Marcel Salathé, et al.	2015	Create an open dataset for plant disease diagnosis using ML.	Curated large-scale image dataset of healthy and diseased leaves.	50,000+ plant leaf images	Enabled mobile plant disease detection tools.	Core dataset for training and evaluating agricultural AI systems.



Thank You!