ISSN: 2320-5407

International Journal of Advanced Research

Publisher's Name: Jana Publication and Research LLP

www.journalijar.com

REVIEWER'S REPORT

Manuscript No.: IJAR-55327

Title: A Multimodal Framework for Crop Disease Diagnosis: Integrating Vision-Based Classification and Large Language Model Reasoning

Recommendation:
Accept as it is .....
Accept after minor revision.....
Accept after major revision
Do not accept (*Reasons below*)

Rating	Excel.	Good	Fair	Poor
Originality		✓		
Techn. Quality		✓		
Clarity			✓	
Significance			✓	

Reviewer Name: **ANAPANA GOPAL**

Reviewer's Comment for Publication.

General Comments

The manuscript presents a well-designed and timely study addressing an important challenge in precision agriculture: reliable, interpretable, and actionable crop disease diagnosis under real-world conditions. By integrating a vision-based detection model with a domain-adapted large language model through Structured Prompt Engineering (SPE), the authors propose a comprehensive multimodal framework that goes beyond conventional image classification. The paper is technically sound, clearly motivated, and supported by extensive experiments and user studies. Overall, the work represents a meaningful advancement in AI-driven agricultural decision support systems.

Content and Originality

The content is original and innovative. The proposed CropDiag-LLM framework distinguishes itself from prior work by enforcing **evidence-grounded reasoning** in LLM outputs using structured visual facts rather than free-text image descriptions. The introduction of Structured Prompt Engineering (SPE) as a formal mechanism to align perception and reasoning is novel and well justified. The integration of Expert Compliance Rate (ECR) as an evaluation metric further strengthens the originality by emphasizing practical agronomic reliability rather than only classification accuracy.

Technical Quality

The technical quality of the manuscript is high. The vision module (YOLOv11-Nano) is appropriately selected and well optimized for field deployment. The LLM fine-tuning strategy, dataset size, and LoRA-based adaptation are clearly described and technically appropriate. Experimental design includes strong baselines, ablation studies, latency analysis, and a real user study, which together provide convincing evidence of effectiveness. Minor clarification could be added regarding dataset

ISSN: 2320-5407
International Journal of Advanced Research
Publisher's Name: Jana Publication and Research LLP
www.journalijar.com

REVIEWER'S REPORT

diversity across seasons and geographic variability, but this does not detract from the core contribution.

Language and Presentation
The manuscript is written in clear, precise, and professional academic English. Technical terminology is used correctly, and the narrative flows logically across sections. Figures, tables, and structured lists enhance readability. Minor typographical issues (e.g., spacing in some references and symbols) can be corrected during copyediting but do not affect comprehension.

Structure and Organization
The paper is well structured and follows a logical progression from problem motivation to methodology, experiments, discussion, and conclusion. Each section is clearly labeled and sufficiently detailed. The methodology section, particularly the SPE description, is exemplary in clarity and reproducibility. The discussion appropriately reflects on limitations and future directions without overstating claims.

References and Citations
The reference list is comprehensive, relevant, and largely up to date, covering foundational works and recent advances in deep learning, multimodal fusion, and agricultural AI. Citations are appropriately placed to support claims. A minor suggestion would be to include one or two recent (2024–2025) peer-reviewed journal articles on multimodal LLM integration, if available, to further strengthen the literature grounding.

Overall Recommendation
This manuscript makes a significant contribution to the fields of precision agriculture, computer vision, and multimodal AI. The proposed framework is novel, technically robust, and validated through extensive quantitative and qualitative evaluation. The work demonstrates clear practical relevance and strong potential for real-world impact.

Final Decision
Accept with Minor Revisions
(Only minor editorial refinements and optional clarifications are suggested; no major methodological changes are required.)