ISSN(O): 2320-5407 | ISSN(P): 3107-4928

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			x	
Techn. Quality				x
Clarity			x	
Significance			x	

Reviewer Name: Dr. Hari Prashad Joshi

Detailed Reviewer's Report

Decision: Major Revision

This paper presents a well-motivated and timely contribution by introducing CropDiag-LLM, a multimodal framework that integrates a YOLOv11-based vision module with a domain-adapted LLM for crop disease diagnosis. The proposed Structured Prompt Engineering (SPE) strategy is a notable innovation that effectively bridges visual detection and causal reasoning, leading to impressive quantitative gains (93.1% accuracy, 97.2% ECR) and strong user preference in field trials. The work addresses significant practical limitations of vision-only systems, such as symptom ambiguity and lack of actionable advice, and demonstrates a meaningful step toward interpretable and trustworthy AI in agriculture.

However, several major revisions are required before publication. First, the dataset description is insufficient; details regarding image collection protocols, geographic and seasonal variability, and class imbalance are absent, which impacts reproducibility and generalizability claims. Second, the choice of YOLOv11 is not adequately justified compared to other recent detectors, and its cited reference (arXiv: 2407.xxxxx) appears placeholder—a proper citation or ablation against other YOLO variants is needed. Third, baseline comparisons are limited; the "vision-only" baseline should include modern CNN or ViT-based classifiers beyond YOLO, and the "LLM-only" baseline should incorporate stronger vision-language models (e.g., GPT-4V) for fairness. Additionally, the hardware specifications (Intel i7, RTX 4070) seem unnecessarily detailed and distract from the methodological contributions.

Minor issues include clarifying the exact fine-tuning dataset size post-processing, discussing potential biases in the expert compliance evaluation, and expanding the limitations section to address ethical considerations regarding farmer dependency on AI. The writing is generally clear, but the abstract could more succinctly highlight the core innovation (SPE) and its impact.

With these revisions, the paper has the potential to be a strong contribution to the field of multimodal AI in agriculture.