

PaddyVLM: An Expert-tuned Vision-Language Model for Paddy Disease Diagnosis

Arun Kumar, Sangam Kumar Jena , Pandarasamy Arjunan

Indian Institute of Science (IISc)
Bengaluru, Karnataka, India

arunkumar12@iisc.ac.in, sangamjena@iisc.ac.in, samy@iisc.ac.in

Abstract

Large multimodal models (LMMs) excel at general vision-language reasoning but often underperform in agriculture, where disease and pest diagnosis demands fine-grained, domain-specific understanding.

We present **PaddyVLM**, a domain-adapted vision-language model for paddy crop analysis, capable of identifying diseases and pests, assessing severity, and providing actionable guidance. Built on `LLaVA-v1.5-7B-LoRA`, our model is trained using *PaddyInstruct*, a curated instruction-tuning dataset derived from the *Paddy Doctor* (10,407 images, 10 classes) and *Paddy Pest* (5,673 images, 20 classes) datasets, annotated and verified by agronomists.

PaddyInstruct combines LLaVA-13B-generated descriptions, Mistral-7B-generated Q&A and multi-turn dialogues, and expert knowledge refinement. Fine-tuning on this dataset equips PaddyVLM with robust fine-grained recognition and context-aware reasoning.

Experiments show that PaddyVLM substantially outperforms general-purpose LMMs in both disease and pest understanding, demonstrating its potential as a practical expert assistant for farmers and agricultural researchers. All code, datasets, and trained models are available at <https://anonymous.4open.science/r/paddy-vlm-7A67/>.

Introduction

Agriculture is central to global food security, with paddy (rice) sustaining over half of the world’s population. Farmers, however, face persistent threats from diseases, pests, nutrient disorders, and environmental stresses—pressures intensified by climate change. Rice health issues alone cause an estimated 10–30% yield loss annually, with severe outbreaks reaching up to 50% (Petchiammal et al. 2023). Early and accurate diagnosis is therefore critical for sustainable and climate-resilient farming.

Artificial intelligence (AI) has long supported agricultural tasks such as crop classification, disease detection, and yield estimation. While deep learning models like CNNs (Krizhevsky, Sutskever, and Hinton 2012) perform well on specific image-based tasks, they are narrow, dataset-dependent, and lack the flexibility needed for real-world decision support.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Dataset Name	No. of Images	No. of Classes
Paddy Doctor Dataset	10,407	10
Paddy Pest Dataset	5,673	20

Table 1: Overview of the Paddy Crop Datasets used for expert-tuning.

Recent advances in vision-language models (VLMs) (Li et al. 2022; Radford et al. 2021) and large multimodal models (LMMs) (Zhang et al. 2024) offer more generalizable capabilities. By combining visual perception with linguistic reasoning, these models can interpret crop symptoms, assess severity, recommend management practices, and answer domain-specific queries—making them promising tools for agricultural intelligence.

In this work, we introduce **PaddyVLM**, a vision-language model for comprehensive paddy health analysis, covering both disease and pest diagnosis along with farmer-oriented recommendations. Inspired by AgroGPT (Awais et al. 2024), PaddyVLM unifies detection, interpretation, and actionable advice within a single framework.

To specialize the model for rice agriculture, we construct *PaddyInstruct*, a multimodal instruction-tuning dataset built from the *Paddy Disease* and *Paddy Pest* datasets (Table 1). Expert annotations ensure strong coverage of diverse biotic stresses. Through multimodal learning and instruction tuning, PaddyVLM integrates visual cues with agricultural knowledge to produce accurate, interpretable outputs—positioning it as a step toward farmer-centric, climate-resilient decision-support systems.

PaddyVLM: Expert-Tuning and Model Training

PaddyVLM is built through a structured pipeline aimed at adapting general-purpose multimodal models to the specific needs of precision agriculture. Since agricultural datasets are predominantly image-only and lack textual annotations, we use a *self-instruct pipeline* inspired by AgroGPT (Awais et al. 2024) to generate high-quality descriptions, single-turn Q&A pairs, and multi-turn dialogues using image attributes and external agricultural knowledge (Figure 1).

We curate two complementary datasets for expert-tuning:

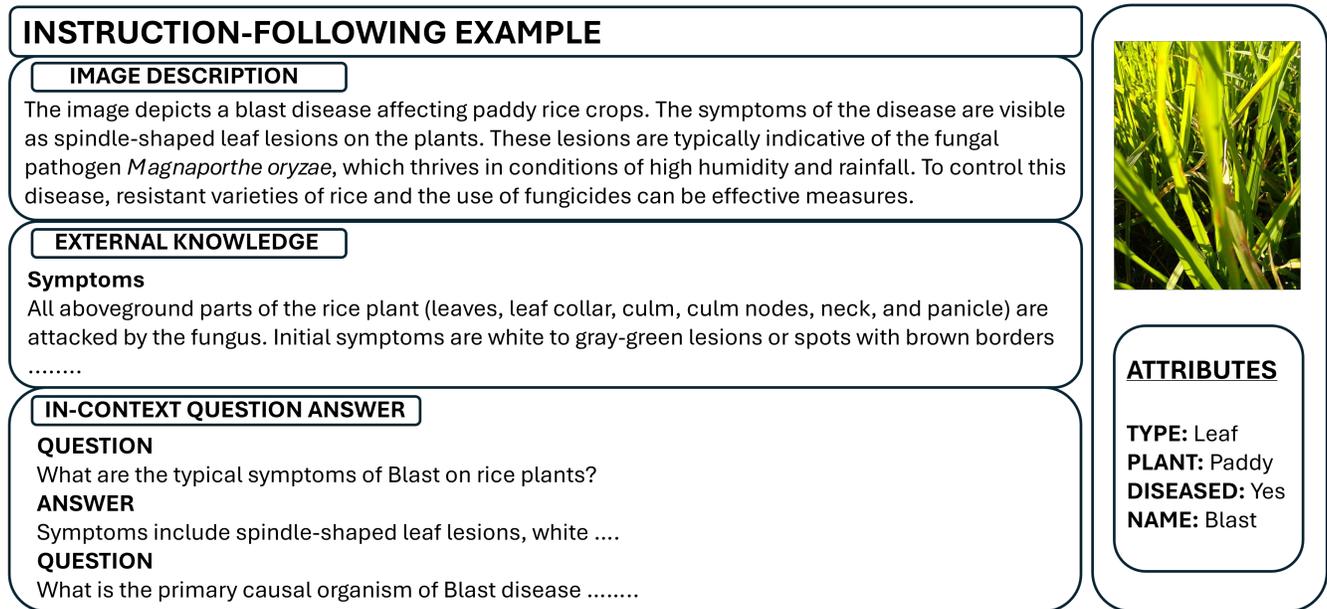


Figure 1: An illustration of instruction-following examples used to generate expert-tuning from vision-only agriculture data.

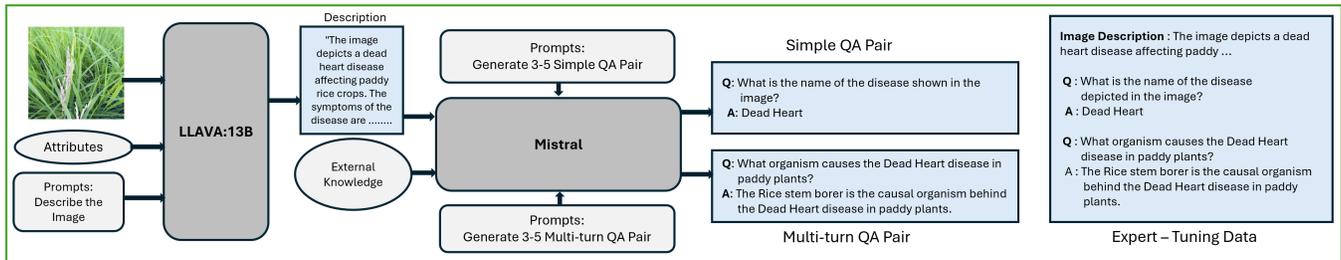


Figure 2: Pipeline used to generate the expert-tuning dataset for PaddyVLM, integrating both disease and pest datasets.

the **Paddy Doctor Dataset** (Petchiammal et al. 2023)¹ with 10,407 images across 10 disease categories, and the **Paddy Pest Dataset** with 5,673 images covering 20 major pest species. The disease dataset provides high-resolution images capturing subtle stress patterns, while the pest dataset includes diverse insect species such as *Brown Planthopper*, *Rice Stem Borer Larva*, and *Sogatella Furcifera*, ensuring broad coverage of real-world field conditions.

Textual descriptions are generated using **LLaVA-13B** (Liu et al. 2023b), and structured Q&A pairs are produced using **Mistral-7B** (Jiang et al. 2023). Prompts incorporate domain knowledge from TNAU (Tamil Nadu Agricultural University 2025) for greater agricultural specificity. Mistral-7B is also used to create multi-turn dialogues simulating farmer–agronomist interactions. Together, these outputs form the *PaddyInstruct* dataset, comprising detailed descriptions, Q&A pairs, and conversational samples (Figure 2).

For the model backbone, we adopt **LLaVA-v1.5-7B** (Liu et al. 2023b) and fine-tune it using LoRA for efficient adaptation. The combination of CLIP ViT-L/14 (Radford et al. 2021) and Vicuna-7B (Zheng et al. 2023) provides

strong multimodal reasoning while remaining computationally manageable. Training follows three stages: (1) vision–language alignment (Liu et al. 2023a), (2) generic instruction tuning, and (3) domain-specific expert-tuning with *PaddyInstruct*. Fine-tuning is performed on an NVIDIA RTX A6000 GPU, with preprocessing on an RTX 5090.

By integrating curated domain data, efficient adaptation, and staged training, **PaddyVLM** achieves strong performance in disease and pest recognition, symptom interpretation, and generation of actionable recommendations, making it a practical model for real-world precision agriculture.

Experiments And Results

The performance evaluation of **PaddyVLM** demonstrates the strong impact of domain-specific expert-tuning for agricultural vision–language tasks. This section provides two complementary analyses: a quantitative assessment through overall and class-wise accuracy comparisons on both the disease and pest datasets, and a qualitative evaluation focusing on interpretability and expert preference. Together, these results show how targeted fine-tuning can effectively adapt general-purpose multimodal models into specialized and reliable tools for precision agriculture.

¹<https://paddydoc.github.io/>



QUESTION : Can you identify the disease shown in the image ?

GROUND TRUTH: Dead heart

PaddyVLM :

The image shows a case of **Dead Heart** disease in paddy crop.

Gemini :

Based on the image, which shows a panicle of rice with many white, sterile, or 'blanked' ,florets (empty kernels), the disease is most likely **Bacterial Panicle Blight (BPB)** or possibly **Rice Blast (Neck Blast)**.

DeepSeek :

The image shows a plant with distinctively colored stem, which is white and brown, surrounded by green leaves. The plant appears to be a type of grass or a similar species, and the stem coloration suggests that it may be affected by a disease. The white and brown coloration of the stem is a visual clue that the plant may be suffering from a **fungal disease**.

Figure 3: Qualitative comparison for single-question diagnosis tasks.

Trained Model Performance Analysis

To comprehensively evaluate the fine-tuned **PaddyVLM**, we assess its per-class and overall performance on both the *Paddy Doctor* and *Paddy Pest* datasets after removing the “normal” class. The following tables present Precision, Recall, and F1-score for each category, along with macro and weighted averages that account for class imbalance.

Class	Precision	Recall	F1-score
tungro	0.91	0.93	0.92
hispa	1.00	0.58	0.74
downy mildew	0.67	0.82	0.74
dead heart	1.00	0.99	1.00
brown spot	0.94	0.95	0.94
blast	0.83	0.90	0.87
bacterial panicle blight	0.92	1.00	0.96
bacterial leaf streak	0.86	0.97	0.91
bacterial leaf blight	0.60	0.96	0.74
Accuracy	0.87	0.87	0.87
Macro Avg.	0.86	0.90	0.87
Weighted Avg.	0.89	0.87	0.87

Table 2: Classification performance across 9 Paddy disease categories

Analysis and Key Findings Across both datasets, **PaddyVLM** demonstrates strong and reliable performance.

Disease classification: The model achieves a weighted F1-score of **0.87**, with multiple classes such as *dead heart*, *brown spot*, and *bacterial panicle blight* achieving F1 above 0.90. The high macro-average Recall of **0.90** indicates consistent capability across all disease categories, including minority classes.

Pest classification: Despite higher inter-class similarity and a larger number of categories (20), the model maintains a weighted F1-score of **0.806**. Several pests—including *RiceYellowStemBorer*, *RiceSkipper*, and *BrownPlanthopper*—achieve near-perfect performance. Lower-performing classes such as *RiceGreenBug* and *RiceStemBorerLarvea* re-

Class	Precision	Recall	F1-score
ArgiopeSpider	0.875	0.875	0.875
BrownMarmoratedStinkBug	0.561	0.920	0.697
BrownPlanthopper	0.893	1.000	0.943
Damselfly	0.933	0.737	0.824
GreenGrasshopper	0.843	0.952	0.894
LadyBug	1.000	0.444	0.615
LeptcorisaAcuta	0.787	0.889	0.835
LongJawedOrbWeaver	1.000	0.552	0.711
RiceBlackBug	0.853	0.853	0.853
RiceBrownPlanthopper	0.400	1.000	0.571
RiceGrasshopper	0.730	0.794	0.761
RiceGreenBug	0.000	0.000	0.000
RiceGreenStinkBug	0.793	0.885	0.836
RiceShieldBugTrick	1.000	0.111	0.200
RiceSkipper	0.952	0.976	0.964
RiceStemBorerLarvea	0.000	0.000	0.000
RiceStinkBug	1.000	0.438	0.609
RiceWhiteStemBorer	0.563	0.900	0.692
RiceYellowStemBorer	0.948	0.982	0.965
SogatellaFurcifera	1.000	0.750	0.857
Accuracy	0.816	0.816	0.816
Macro Avg.	0.757	0.703	0.685
Weighted Avg.	0.849	0.816	0.806

Table 3: Classification performance across multiple Paddy pest categories.

flect limited sample diversity and close morphological resemblance to other species.

Overall: The results confirm that the LoRA fine-tuning effectively adapts the base vision-language model for domain-specific agricultural tasks. **PaddyVLM** achieves robust generalization across 29 total classes and demonstrates strong reliability for real-world deployment in automated crop disease and pest diagnosis.

Comparison with Gemini

We compare **PaddyVLM** with the general-purpose multimodal model **Gemini 2.5 Flash** and the open-source **LLaVA-7B** on two fine-grained tasks: *paddy disease recog-*



Question : see the image and tell me what type of disease, visible symptoms and possible solution?
Ground Truth : Dead Heart,
Symptoms : Central shoot of the tiller dries up and turns brown while outer leaves remain green.
Solution : Use resistant varieties, remove and destroy affected tillers, and apply recommended insecticides.

Paddy VLM :

The image depicts a paddy plant affected by the **Dead Heart** disease. **The central shoot of the plant has dried and browned**, indicating severe damage to the heart of the plant. However, the **outer leaves of the plant remain green**, which suggests that the disease is localized to the central shoot. The Dead Heart disease is caused by the Rice stem borer, a pest that thrives in warm climates. The presence of this pest in the field can be identified by the symptoms visible on the plant. To control the spread of this disease and prevent further damage to the crop, **resistant varieties of paddy plants can be used**. Additionally, insecticides can be applied to kill the Rice stem borer and protect the healthy parts of the plant from further damage. It is important to monitor the field regularly for signs of infestation and take timely action to prevent significant crop loss.

Gemini :

Disease Type:

The symptoms strongly suggest a **Panicle Disease** affecting grain development. The most probable specific diseases are:

Bacterial Panicle Blight (BPB): Caused by *Burkholderia glumae*. This is a very strong candidate given the white, sterile florets and the erect posture of the affected panicle (it doesn't bend over from the weight of filled grains).

Rice Blast (Neck Blast): Caused by *Magnaporthe oryzae*.....

DeepSeek :

The image shows a plant with a distinctively discolored stem and leaves. The stem is a **pale green color, while the leaves exhibit a mix of green a lighter, almost white or yellowish hue**. The discoloration is most pronounced on the stem, where it transitions from a healthy green to a pale green and then to a whitish color. The visible symptoms suggest that the plant may be affected by a **fungal disease**.

Figure 4: Qualitative comparison for multi turn-question diagnosis task.

tion and farm pest identification, both requiring attention to subtle cues such as lesion color, texture, and pest-induced damage.

As shown in Table 4, **PaddyVLM** surpasses both Gemini and LLaVA-7B by a significant margin, achieving more than twice Gemini’s accuracy on both benchmarks. While Gemini identifies general stress, it often misattributes symptoms, confusing pest damage with fungal or nutrient issues. In contrast, **PaddyVLM** delivers precise, agronomically consistent predictions—correctly recognizing cases such as *Hispa* and *Brown Spot* and grounding its outputs in visual evidence.

This improvement is driven by **PaddyVLM**’s domain-specific alignment and instruction tuning, which enable recognition of subtle inter-class differences that generic VLMs overlook. Its integration of agricultural descriptors and field-level context allows for more reliable interpretation of plant pathology and pest morphology.

Model	Paddy Diseases (%)	Paddy Pest (%)
LLaVA-7B (Baseline)	10.00	13.00
Gemini 2.5 Flash (Zero-shot)	36.63	23.02
PaddyVLM (Zero-shot)	87.00	81.83

Table 4: Zero-shot comparison of **PaddyVLM**, **Gemini 2.5 Flash**, and **LLaVA-7B**

Overall, **PaddyVLM** shows superior zero-shot reasoning for both disease and pest identification, demonstrating the value of domain-grounded visual–linguistic alignment for precision agriculture.

Comparative Inference Analysis

We conducted a zero-shot evaluation comparing **PaddyVLM** with the baseline multimodal models listed in Table 5, using the *Paddy Disease* and *Paddy Pest* datasets. Each test split contained balanced samples across all classes, and two tasks were evaluated: (1) *Identification* (healthy vs. affected) and (2) *Fine-grained Classification* (specific disease or pest category).

The results in Table 5 show that although baseline models perform reasonably well on the identification task, they struggle with fine-grained classification across both datasets. Some models achieve isolated gains, but overall accuracy remains low, indicating limited ability to capture the subtle visual and semantic cues present in agricultural imagery.

In contrast, **PaddyVLM** delivers consistently high performance on both tasks, outperforming all baselines by large margins—exceeding them by more than 40% in fine-grained pest classification. This improvement arises from its domain-specific tuning, which enables reliable differentiation between closely related symptoms such as fungal lesions and pest-induced damage.

Overall, these findings highlight the difficulty general-purpose VLMs face in modeling the nuanced visual semantics of plant pathology and entomology. **PaddyVLM**’s strong zero-shot results demonstrate the effectiveness of domain-grounded multimodal learning for practical crop intelligence and precision agriculture.

Qualitative Results

To complement the quantitative results, we performed a qualitative analysis evaluating interpretability, reasoning quality, contextual awareness, and the practical usefulness of model outputs. This evaluation includes three parts: (1)

You are an AI assistant specialized in agricultural topics. You are provided with the text description of an image of a plant, attributes of the plant (such as name, disease), and common information of the plant. Unfortunately, you don't have access to the actual image. You must generate exactly 3 to 5 pairs of question and answer (Q&A). Each question should begin with "Q:" and each answer with "A:". Do not include any narrative text outside the Q&A pairs.

Instructions:

- Focus on visual details that can be seen in the image (e.g., plant type, symptoms, disease, prevention).
- Do not refer to the 'text', 'context', or 'caption' — behave as if you are only seeing the image.
- Do not ask speculative or ambiguous questions.
- Avoid referencing numbers, datasets name or image name.
- Maintain consistent formatting as:

Q1: ...
A1: ...
Q2: ...
A2: ...
(and so on)

Context:
Image Description: {data["description"]}
Attributes: {data["attributes"]}
External Knowledge: {external_knowledge}

Figure 5: Prompt to generate the Question and Answer.

You are an agricultural assistant. Describe this image of a {class_label} from the {dataset_name} dataset. Use the following attributes for a more detailed and contextual description:\n{attributes}

Figure 6: Prompt to generate the Context Description.

In-Distribution Qualitative Analysis, (2) *Web-based Paddy Disease Analysis*, and (3) *Agronomist Preference Study*. Together, these analyses offer a comprehensive view of the model’s real-world performance and interpretability.

In-Distribution Qualitative Analysis In the single-question diagnostic task (Figure 3), **PaddyVLM** correctly identified *Dead Heart* and provided a biologically grounded explanation referencing key visual symptoms. In contrast, **Gemini** misclassified the case, and **DeepSeek** produced an unclear, non-causal interpretation.

In the multi-turn setting (Figure 4), **PaddyVLM** maintained contextual consistency, offering accurate diagnoses and relevant agronomic guidance, while baseline models generated generic or contradictory responses.

These findings demonstrate **PaddyVLM**’s superior reasoning, symptom interpretation, and domain-aware dialogue capabilities in in-distribution scenarios.

Qualitative Analysis on Paddy Disease and Pest Images from the Web To assess real-world generalization, we tested **PaddyVLM** on web-sourced paddy disease and pest images containing natural variations such as uneven lighting, cluttered backgrounds, occlusion, and inconsistent image quality.

As shown in Figure 7, **PaddyVLM** demonstrates stronger

Paddy Disease Dataset		
Model	ID (%)	Cls. (%)
LLaVA-7B (Liu et al. 2023a)	57.22	11.56
LLaVA-13B (Liu et al. 2023a)	56.78	10.11
LLaVA-34B (Liu et al. 2023a)	85.56	6.00
LLaVA-Next-8B (AI 2025)	51.00	9.33
Qwen3-VL-8B (Bai et al. 2024b)	42.44	0.00
Qwen3-VL-32B (Bai et al. 2024b)	61.89	0.78
Qwen2.5-VL-7B (Bai et al. 2024a)	15.44	0.33
Qwen2.5-VL-72B (Bai et al. 2024a)	20.67	0.00
Gemma3-4B (Research and DeepMind 2024)	95.78	10.78
Gemma3-12B (Research and DeepMind 2024)	79.67	11.89
Granite3.2-Vision (Research 2025)	17.67	59.11
LLaVA-Phi3-3.8B (Liu et al. 2024)	20.22	8.00
MiniCPM-V (Team 2024)	52.56	8.56
PaddyVLM (ours)	96.5	87.0
Paddy Pest Dataset		
Model	ID (%)	Cls. (%)
LLaVA-7B	27.07	4.46
LLaVA-13B	40.60	5.35
LLaVA-34B	51.13	4.10
LLaVA-Next-8B	10.90	3.57
Qwen3-VL-8B	54.14	12.12
Qwen3-VL-32B	48.50	12.48
Qwen2.5-VL-7B	12.78	8.02
Qwen2.5-VL-72B	28.20	14.44
Gemma3-4B	88.16	3.74
Gemma3-12B	70.86	9.98
Granite3.2-Vision	27.26	12.83
LLaVA-Phi3-3.8B	10.34	4.99
MiniCPM-V	29.51	8.73
PaddyVLM (ours)	95.2	83.07

Table 5: Performance comparison of models on the *Paddy Doctor* and *Paddy Pest* datasets.

contextual grounding and fine-grained recognition than general-purpose VLMs. For instance, it correctly identified the *Hispa* pest as “*Hispa (Di cladispa armigera)*,” whereas competing models produced vague or incorrect interpretations.

These results highlight **PaddyVLM**’s ability to distinguish pest-induced and pathological symptoms under noisy field conditions. Its domain-aligned reasoning and use of agricultural ontology enable more accurate and reliable diagnostics compared to generic VLMs.

Overall, the combined quantitative and qualitative results confirm that **PaddyVLM** effectively bridges the gap between multimodal reasoning and domain-specific expertise, offering a reliable, interpretable, and high-performing solution for intelligent paddy disease and pest diagnosis.

Conclusion and Future Work

In this work, we introduced **PaddyVLM**, a vision–language model tailored for paddy farming that integrates visual recognition with natural language reasoning. Beyond traditional classification, it provides diagnosis, explanation, and



QUESTION : Can you identify the disease shown in the image ?

GROUND TRUTH: Hispa.

PaddyVLM :

The disease shown in the image is a Hispa, scientifically known as *Dicladispa armigera*.

Gemini :

The image shows rice plants with several small, dark, oblong objects on the leaves, along with some white streaks or marks. Based on these visible symptoms, the most likely issue is:

Insect Pest Infestation, specifically from Planthoppers or Leafhoppers, which can also transmit viral diseases.

Copilot :

From the image you shared, the rice plant looks generally healthy, but I notice a few small dark spots on some of the leaves. These spots could be early signs of a **leaf disease**, though it's not possible to confirm the exact cause just from a photo alone.

Figure 7: Results on web image selected randomly for paddy disease.

actionable recommendations. Our results show that expert-tuning and multimodal learning effectively bridge the gap between advanced AI and field-level agricultural needs, supporting more sustainable and climate-resilient farming.

Looking ahead, this approach can be expanded to additional crops and agricultural tasks. Future efforts will focus on scaling expert-tuning to diverse datasets and improving generalizability across farming conditions. Incorporating modalities such as soil data, sensor readings, and climatic indicators will further address challenges posed by climate variability. Ultimately, our goal is to develop a comprehensive decision-support framework that enables farmers worldwide to make informed, sustainable, and climate-aware decisions for improved food security.

References

- AI, M. 2025. Mapler/LLaMA3-LLaVA-Next: Enhanced Visual Instruction Tuning. Hugging Face Model Card: mapler/llama3-llava-next-8b.
- Awais, M.; Alharthi, A. H. S. A.; Kumar, A.; Cholakkal, H.; and Anwer, R. M. 2024. AgroGPT: Efficient Agricultural Vision-Language Model with Expert Tuning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 5687–5696. IEEE.
- Bai, J.; et al. 2024a. Qwen2.5-VL: Large Vision-Language Models with Enhanced Visual Grounding. ArXiv preprint arXiv:2409.87654.
- Bai, J.; et al. 2024b. Qwen3-VL: Vision-Language Models from Alibaba Cloud. ArXiv preprint arXiv:2407.12345.
- Jiang, A. Q.; Sablayrolles, A.; Mensch, A.; Bamford, C.; Chaplot, D.; Lengyel, G.; Lengyel, G.; Saade, A.; Sayed, S.; Balestriero, R.; et al. 2023. Mistral 7B: A Compact and Powerful Language Model. *arXiv preprint arXiv:2310.06825*.
- Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25.
- Li, J.; Li, D.; Savarese, S.; and Hoi, S. C. 2022. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. In *International Conference on Machine Learning*, 12888–12900. PMLR.
- Liu, H.; Li, C.; Li, Y.; and Lee, Y. J. 2023a. Visual Instruction Tuning. *arXiv preprint arXiv:2304.08485*.
- Liu, H.; et al. 2024. LLaVA-Phi3: A Compact Multimodal Language Model. ArXiv preprint arXiv:2410.12345.
- Liu, W.; Zhang, P.; Li, D.; Li, J.; and Hoi, S. C. 2023b. InstructBLIP: Towards General-purpose Vision-Language Models with Instruction Tuning. *arXiv preprint arXiv:2305.06500*.
- Petchiammal, A.; Kiruba, B.; Murugan, D.; and Arjunan, P. 2023. Paddy Doctor: A Visual Image Dataset for Automated Paddy Disease Classification and Benchmarking. In *Proceedings of the 6th Joint International Conference on Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD)*, 203–207.
- Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; Krueger, G.; and Sutskever, I. 2021. Learning Transferable Visual Models from Natural Language Supervision. In *International Conference on Machine Learning*, 8748–8763. PMLR.
- Research, G.; and DeepMind. 2024. Gemma 3: Multimodal Foundation Models by Google DeepMind. ArXiv preprint arXiv:2406.11220.
- Research, I. 2025. Granite 3.2 Vision: IBM Foundation Model for Multimodal Reasoning.
- Tamil Nadu Agricultural University. 2025. Agritech Portal. <https://agritech.tnau.ac.in/>. Accessed: 2025-08-22.
- Team, O. 2024. MiniCPM-V: Lightweight Multimodal Foundation Model. GitHub repository: <https://github.com/OpenBMB/MiniCPM-V>.
- Zhang, J.; Huang, J.; Jin, S.; and Lu, S. 2024. Vision-Language Models for Vision Tasks: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8): 5625–5644.
- Zheng, L.; Chiang, W.-L.; Sheng, Y.; Zhuang, S.; Wu, Z.; Zhuang, Y.; Lin, Z.; Li, Z.; Zhang, H.; Xing, E. P.; Zhang, H.; Gonzalez, J. E.; and Stoica, I. 2023. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. arXiv:2306.05685.