BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Multimodal AI-Based Crop Health Monitoring Using Image, Weather and Soil Data for Precision Agriculture: A Review and Proposed Framework

[1]Pradnya Bhosale, [2]Prof. Ankita Pokale

[1]Student, [2]Assistant Professor – Seminar Guide
[1]Department of Artificial Intelligence and Data Science,
[1]Ajeenkya DY Patil School of Engineering, Pune, India

Abstract

Detecting disease symptoms in the field before they reach advanced stages gives farmers more options in how they respond. Classification of leaf images by deep learning models has already been shown to be possible. However, leaf images alone are not enough, because a leaf image does not capture the circumstances surrounding a crop's growth. These circumstances include temperature, humidity, rainfall, moisture, soil properties and the crop's stage of growth. This paper examines deep learning for plant disease detection, computer vision, multimodal language models, and privacy-aware and explainable AI. Using the existing literature, it proposes a framework for analysing leaf images together with structured weather, soil and crop-stage data. The framework is not complete and has not yet been tested. Its worth will have to be determined in a study that compares image-only and multimodal models on test datasets collected under actual field conditions.

Keywords

Crop disease detection, multimodal artificial intelligence, precision agriculture, deep learning, computer vision, feature fusion, crop health monitoring.

I. INTRODUCTION

Healthy crops are of great economic and social value to agriculture. Disease that is diagnosed too late can cause food loss and a loss of crop quality, and with them considerable economic cost. Farmers usually focus their attention on the visible parts of a crop, such as the leaves. This is a detection method that leaves a lot of ambiguity and room for error, because symptom identification is subjective and specialised consultation is not always available.

AI systems have made this process easier by combining computer vision with machine learning models and pattern recognition on crop photographs. The photographs are analysed to aid disease classification.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Convolutional Neural Networks (CNNs) and, more recently, Vision Transformers and vision-language models have gained much research attention [1]–[4].

Useful as these models are, they are prone to errors and leave ambiguity in a disease detection scenario. A leaf photograph does not give a direct interpretation of environmental conditions such as humidity, rainfall, moisture and the stage of plant growth, any of which might have been a factor in the plant's condition. The need for an integrated system that can handle all of this data is the key motivation of this work.

This paper introduces the proposed system and explains how it could be designed and evaluated, in order to assess whether it is indeed the answer to this need.

II. PROBLEM STATEMENT

Many plant disease classification models focus on leaf images. These models learn image patterns, but they do not perform well in the field because of poor-quality photographs and varying crop conditions and crop varieties. Many image-based models exist, but they lack environmental context.

The goal of this research is to see whether image-based plant disease detection can be improved by combining leaf image features with crop-stage and crop-environment features such as soil and weather. Any improvement must also be quantified. This requires comparing the proposed model against existing image-based models using consistent test methods.

III. OBJECTIVES

The objectives of this paper are as follows:

1. To review deep learning and computer vision for plant disease detection.
2. To look at the latest research directions, including Vision Transformers, multimodal models and explainable AI.
3. To look at the drawbacks of image-only models, such as bias and missing context.
4. To propose a framework that integrates image, soil, weather and crop growth stage data.
5. To list the requirements for evaluating whether the proposal adds benefit over image-only models.

IV. LITERATURE SURVEY

This review considers papers that present AI methods applied to agriculture and plant disease. Review articles explain common practices and the existing challenges in the field. Some articles present newly developed image-based AI methods for agriculture, whereas others consider issues surrounding the interpretation and implementation of AI in agriculture. These papers do not contain sufficient evidence to support every design decision in this research, but they provide relevant background.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

A. Deep Learning for Plant Disease Detection

A number of deep learning methods for plant disease detection and related agricultural tasks have been surveyed in [1]. That review helps clarify the need for deep learning and points out common research issues in the area. It does not, however, explain how adding weather and soil information would improve a particular system.

Pacal et al. also published a systematic review of deep learning for plant diseases [2]. Their work gives a better understanding of the methods and the evaluation techniques employed in the area. A significant issue is that the results of many papers are not robust enough, especially for models built on curated image datasets. Strong evaluation datasets with data that reflect a variety of field conditions are therefore important.

B. Computer Vision Technologies

Ding et al. summarised computer vision technologies used for monitoring plant diseases [3]. They described both traditional methods and the more recent transformer-based technologies for image processing. This survey provides a starting point for developing a computer vision system that can serve as the first building block of a multimodal system.

C. Multimodal Language Models, Multimodal Fusion and Explainable AI

Roumeliotis et al. compared a multimodal large language model, GPT-4o, with ResNet-50 for plant disease classification using leaf images [4]. The fine-tuned multimodal model slightly outperformed ResNet-50 and showed better generalisation, whereas zero-shot performance was relatively weak and still required training data.

A review by Xu et al. covers multimodal AI that integrates imaging, environmental and sensor data for disease and pest management [6]. It reports that multimodal systems achieved approximately 3–48.9% higher diagnostic accuracy than unimodal approaches, while noting that multimodal datasets are difficult to collect and require more computational resources. These figures are reported by the review and have not been verified in this work.

Shawon et al. reviewed federated learning and explainable AI (XAI) for plant disease detection [5]. Federated learning enables privacy-preserving model training and XAI improves interpretability, but federated learning faces communication overhead and heterogeneous-data challenges.

V. COMPARATIVE SUMMARY OF THE LITERATURE

Table I captures the relationship of the selected publications to the current study. The last column outlines considerations that the proposed framework should address; the scope and shortcomings of the cited papers may differ.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

TABLE I. SUMMARY OF SELECTED LITERATURE

Study	Main focus	Contribution relevant to this paper	Consideration for the proposed framework
Ngugi et al., 2024 [1]	Systematic review of deep learning for plant disease detection.	An overview of the deep learning techniques used for disease detection.	The review does not cover the effectiveness of combining data streams.
Pacal et al., 2024 [2]	Deep learning for plant disease recognition.	An overview of common research challenges and guidance on evaluation design in plant disease research.	Performance from controlled experiments and datasets may not be replicable in real-world field scenarios.
Ding et al., 2024 [3]	Computer vision techniques for plant disease recognition.	Summarises current and new computer vision technologies.	Feature extraction alone does not cover the fusion of image data with structured weather, soil and crop-stage data.
Roumeliotis et al., 2025 [4]	Multimodal LLMs and CNNs in plant disease recognition.	Compares GPT-4o with ResNet-50 on leaf images.	Combining language and vision should not be conflated with fusing specific data inputs collected from agricultural activities.
Shawon et al., 2026 [5]	Federated learning and XAI in plant disease recognition.	Privacy-conscious and interpretable models for plant disease recognition.	Explainable and federated methods do not by themselves justify using multi-source data in a fused model.
Xu et al., 2026 [6]	Multimodal AI for disease and pest management in precision agriculture.	Reviews systems that integrate imaging, environmental and sensor data.	Multimodal datasets are difficult to collect and need more computational resources.

VI. RESEARCH GAP

A large part of the literature on deep learning and computer vision in agriculture notes that foundation models, explainability and privacy-preserving machine learning are becoming increasingly important. However, in the literature reviewed there is still no established method for fusing leaf image, weather, soil and crop growth stage data for the same agricultural prediction task.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

The problem is therefore well defined and measurable: does integrating structured agricultural data with image features improve the performance or robustness of plant disease classification relative to an image-only baseline? Answering this question requires a more exhaustive literature search, a dataset in which each record represents the same crop at the same time and place, and well-controlled experiments. The identified gap applies only to the literature reviewed and is not a claim that no multimodal research has been conducted in agriculture.

VII. PROPOSED FRAMEWORK

The proposed framework is intended as a guideline for future work in this domain. It has not yet been designed in detail or implemented.

A. Input Modalities

Four types of information are proposed, as summarised in Table II.

TABLE II. PROPOSED INPUT MODALITIES

Modality	Typical content	Role in the framework
Leaf images	Visual symptoms of the crop, such as leaf or plant status.	Features extracted by a CNN or a Vision Transformer.
Weather data	Temperature, humidity and rainfall.	Crop-specific measures that depend on the disease in question.
Soil data	Soil moisture, pH and nutrient indicators (e.g., nitrogen).	Environmental context that an image does not capture.
Crop-stage information	Crop type and growth stage.	Used as features alongside the other inputs.

B. Image and Structured Data Processing

Branches dedicated to image and structured data would run in parallel. Visual data, such as a photograph of a leaf, would be processed by an image branch, whereas structured data, including weather, soil and crop stage, would be cleaned, encoded and scaled by a structured-data branch. After the branches are joined by a feature fusion technique, a model would make the prediction. Fig. 1 shows the proposed methodology.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

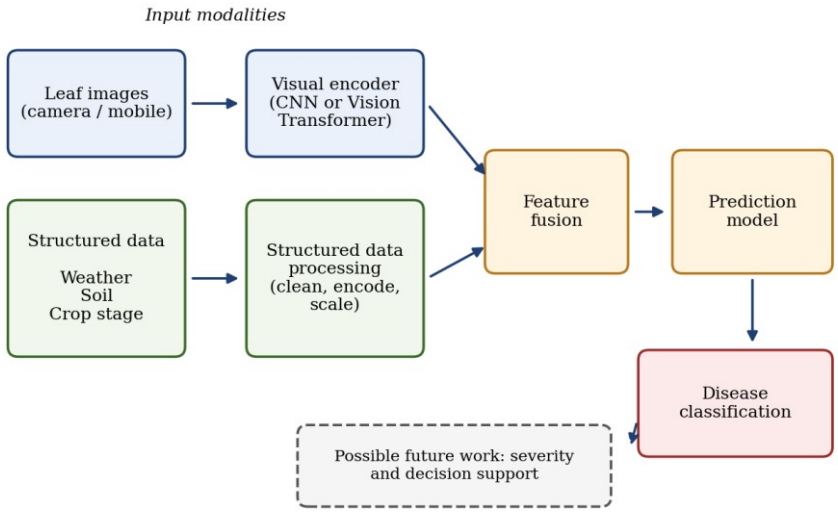


Fig. 1. Proposed multimodal framework for crop disease classification.

A key goal for future research is the development of models that make more sophisticated predictions, including disease severity, treatment suggestions and yield forecasting. Such models would need to be developed and evaluated separately. For now, the aim is to develop a model that classifies diseases. Its predictions will not replace a diagnosis and should not be used on their own to determine the course of treatment.

C. Evaluation Plan

Researchers should describe their methods in sufficient detail that comparisons can be made between studies. The baselines in Table III are proposed.

TABLE III. PROPOSED BASELINES FOR EVALUATION

Baseline	Inputs
1	Image only.
2	Image and weather.
3	Image and soil.
4	Combination of image, weather, soil and crop stage.

For each baseline, researchers can report accuracy, precision, recall and the macro F1 score, and a confusion matrix can also be provided. If the dataset is unbalanced, macro F1 and per-class recall can be reported instead of accuracy. Data splits should be designed so that no data leaks from the training split into the test split.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

VIII. STRENGTHS AND LIMITATIONS

A. Potential Strengths

Deep learning approaches allow patterns to be extracted from images, and this can be exploited to automate plant disease recognition. Using multiple inputs can supply context, such as environmental data, that an image otherwise lacks. A modular design would also allow each input to be analysed individually to evaluate its relative contribution to the system. These design proposals are hypothetical and have yet to be tested.

B. Limitations and Practical Challenges

Before real-world adoption, a number of important issues must be considered:

- **Data alignment:** image, meteorological, soil and crop growth stage data should be matched for time and location.
- **Sensor data noise:** sensor and meteorological data can be noisy, delayed or incomplete.
- **Field data generalisation:** images used for training may be of a more ideal nature than images captured on real farms.
- **Resources:** maintaining soil data sensors can be prohibitively expensive and logistically demanding for most farmers.
- **Trustworthy agronomic decisions:** without sufficient empirical evidence and oversight from an agronomist, disease predictions should not be treated as confirmed diagnoses or prescriptive interventions.
- **Model complexity:** including more data types and modalities may increase computational, maintenance and integration requirements.

The biggest current challenge is finding an aligned dataset of all inputs and modalities. If inputs come from different locations and times, the model is likely to learn wrong patterns and correlations. Ultimately, the effectiveness of the model will depend as much on data quality as on model design.

IX. ETHICAL, SOCIETAL AND SUSTAINABILITY CONSIDERATIONS

Crop health monitoring has potential as an early warning system, but it would take time to reach that potential. Incorrect outputs would cause loss of resources and money through misdiagnoses, overdiagnosis and unnecessary treatments. Systems intended for farmers should make uncertainty clear and should supplement, not replace, agricultural intuition. Farmers need to be informed how their data would be collected and managed and where it would go once collected.

Explainable models do not equate to correct predictions, but they can improve how predictions are interpreted. Federated learning could prove useful where data collection cannot be centralised, although other issues must then be considered. Better crop health management could improve the efficient use of

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

inputs and resources while decreasing food waste and loss. The architecture proposed here does not, on its own, address sustainability; this will be an objective of later field studies. There is also an opportunity for collaboration across multiple disciplines and domains.

X. CONCLUSION

This paper examined existing deep learning studies on crop disease detection, together with recent developments in computer vision and multimodal large language models. The surveyed studies are a starting point for assessing the contribution of AI to plant health, but the performance of AI in the field remains an open problem. The papers highlight the critical need for better dataset quality. Along with the survey, a framework for merging several key research components was put forward.

XI. DIRECTIONS FOR FUTURE WORK

The first steps should be to pick one crop and identify a manageable list of diseases. It must then be determined whether data on leaf images are available along with data on weather, soil and crop maturity stage. These data can be used to establish an image-only baseline model, which will serve as the foundation for analysing different ways of combining information.

The models should then be assessed using consistent data splits and metrics, with attention to missing-data imputation and field generalisation. Finally, after these initial steps and evaluations, the focus should turn to building an explainable AI component that assists the farmer in decision making [1][3], [5], [6].

REFERENCES

[1] H. N. Ngugi, A. E. Ezugwu, A. A. Akinyelu, et al., “Revolutionizing Crop Disease Detection with Computational Deep Learning: A Comprehensive Review,” Environmental Monitoring and Assessment, 2024, doi: 10.1007/s10661-024-12454-z.
[2] I. Pacal et al., “A Systematic Review of Deep Learning Techniques for Plant Diseases,” Artificial Intelligence Review, 2024, doi: 10.1007/s10462-024-10944-7.
[3] W. Ding, M. Abdel-Basset, I. Alrashdi, H. Hawash, et al., “Next Generation of Computer Vision for Plant Disease Monitoring in Precision Agriculture: A Contemporary Survey, Taxonomy, Experiments, and Future Direction,” Information Sciences, vol. 665, 2024, doi: 10.1016/j.ins.2024.120338.
[4] K. I. Roumeliotis, R. Sapkota, M. Karkee, N. D. Tselikas, and D. K. Nasiopoulos, “Plant Disease Detection through Multimodal Large Language Models and Convolutional Neural Networks,” arXiv preprint arXiv:2504.20419, 2025. [Online]. Available: https://arxiv.org/abs/2504.20419.
[5] S. M. Shawon, F. L. Niha, O. H. Sinan, et al., “Integrating Federated Learning and Explainable AI for Plant Disease Detection: A Systematic Review of Recent Advances in Precision Agriculture,” 2026.
[6] Xu et al., “Multimodal Artificial Intelligence for Disease and Pest Management in Precision Agriculture: A Review,” 2026.