A Smart Multimodal System for Crop Disease Detection and Agricultural Decision Support

Harsha G V., G Sagar., Darshan C Y., Darshan R Shelar., Kethan K Raikar., Usha K

Jain Institute of Technology, Davangere

DOI: https://doi.org/10.51244/IJRSI.2026.1305000164

Received: 03 May 2026; Accepted: 08 May 2026; Published: 05 June 2026

ABSTRACT

Precision agriculture has become an important research area for improving crop productivity, food security, and sustainable farming practices. Crop diseases significantly reduce agricultural yield and economic stability, especially in developing agricultural economies. Traditional disease detection methods rely heavily on manual inspection and expert knowledge, which are time-consuming and often inaccurate under large-scale farming conditions. This manuscript proposes a multimodal artificial intelligence framework that integrates Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), environmental parameters, and reasoning-based support systems for crop disease identification and agricultural decision support.

The proposed system combines crop leaf images, weather conditions, and environmental sensor data to improve disease classification performance. The PlantVillage dataset containing approximately 54,000 crop leaf images across multiple disease categories was utilized for experimentation. Image preprocessing techniques such as resizing, normalization, augmentation, and noise removal were applied before model training. A hybrid CNN-Transformer architecture was implemented using Adam optimizer with a learning rate of 0.001 over 50 epochs. Experimental evaluation demonstrated an overall classification accuracy of 93.2%, outperforming traditional CNN-only approaches.

The system also provides treatment recommendations and environmental analysis to support farmers in real-time decision-making. Comparative evaluation using precision, recall, F1-score, and confusion matrix analysis confirms the effectiveness of multimodal data fusion. Although computational complexity and deployment challenges remain limitations, the proposed framework demonstrates strong potential for smart agriculture applications.

INTRODUCTION

Agriculture plays a critical role in supporting global food production and economic development. Crop diseases remain one of the major challenges affecting agricultural productivity, leading to substantial economic losses and food insecurity. Early identification of plant diseases is essential for improving crop health and reducing yield reduction. Conventional disease detection approaches depend on visual inspection by agricultural experts, which is often unavailable in rural regions and may lead to delayed diagnosis.

Recent advances in Artificial Intelligence (AI), Deep Learning (DL), and Computer Vision have enabled automated crop disease detection systems. Convolutional Neural Networks (CNNs) are widely used for extracting local image features, while Transformer architectures provide enhanced global feature learning capabilities. However, image-only models may fail under real-world agricultural conditions due to lighting variation, environmental complexity, and background noise.

To address these limitations, multimodal learning approaches integrate multiple data sources such as crop images, environmental conditions, humidity, temperature, and textual reasoning systems. Multimodal systems improve contextual understanding and decision-making capabilities, making them more suitable for precision agriculture applications.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue V May 2026

This manuscript proposes a multimodal crop disease detection framework integrating CNNs, Vision Transformers, environmental data fusion, and intelligent recommendation systems. The study aims to improve disease classification accuracy, support farmers with treatment recommendations, and enhance agricultural productivity under realistic farming conditions.

LITERATURE REVIEW

Several studies have explored AI-based crop disease detection techniques using machine learning and deep learning models. Sharma et al. (2022) implemented CNN-based classification for tomato leaf disease detection and achieved promising performance under controlled datasets. However, their system lacked environmental context integration.

Thai et al. (2023) introduced Vision Transformer architectures for agricultural image analysis, demonstrating improved global feature extraction compared to traditional CNN models. Their work highlighted the effectiveness of Transformer-based learning under complex image conditions.

Lu et al. (2024) proposed multimodal agricultural frameworks integrating weather conditions with visual disease analysis. Their study emphasized the importance of environmental information for improving prediction reliability.

Xu et al. (2025) developed a hybrid multimodal disease detection framework combining CNNs and sensor-based environmental analysis. The authors reported improved disease classification performance and enhanced real-time monitoring capabilities.

Recent research trends indicate that multimodal AI systems outperform single-modal approaches by combining image analysis, environmental sensing, and reasoning-based recommendation systems. However, challenges including computational complexity, dataset imbalance, and deployment in low-resource agricultural environments remain active research problems.

METHODOLOGY

The proposed framework follows a multimodal deep learning architecture for crop disease identification and agricultural decision support.

Dataset Collection

The PlantVillage dataset was used for image-based disease classification. The dataset contains approximately 54,000 images belonging to healthy and diseased crop categories including tomato, potato, maize, and pepper leaves. Environmental parameters such as humidity, rainfall, and temperature were integrated from publicly available agricultural datasets.

Data Preprocessing Image preprocessing included:

- Resizing images to 224 × 224 pixels
- Normalization and augmentation
- Background noise removal
- Rotation, flipping, and brightness adjustment

Environmental data preprocessing included:

- Missing value handling

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue V May 2026

• Feature normalization

• Sensor value standardization

Feature Extraction

A hybrid CNN-Transformer architecture was implemented:

• CNN layers extract local texture and disease patterns

• Vision Transformer layers capture global image dependencies

• Environmental features are processed through dense neural layers

Multimodal Fusion

Extracted visual and environmental features are combined using feature-level fusion techniques. This enables the model to understand disease patterns under varying agricultural conditions.

Model Training

The model was trained using:

• Optimizer: Adam

• Learning rate: 0.001

• Epochs: 50

• Batch size: 32

• Validation split: 80:20

Evaluation Metrics

System performance was evaluated using:

• Accuracy

• Precision

• Recall

• F1-score

• Confusion matrix analysis

• ROC curve analysis

RESULTS AND DISCUSSION

Experimental evaluation demonstrates that the proposed multimodal system achieves superior performance compared to conventional image-only CNN models. The hybrid CNN-Transformer framework achieved an overall accuracy of 93.2%, with improved disease classification under varying environmental conditions.

Performance Metrics:

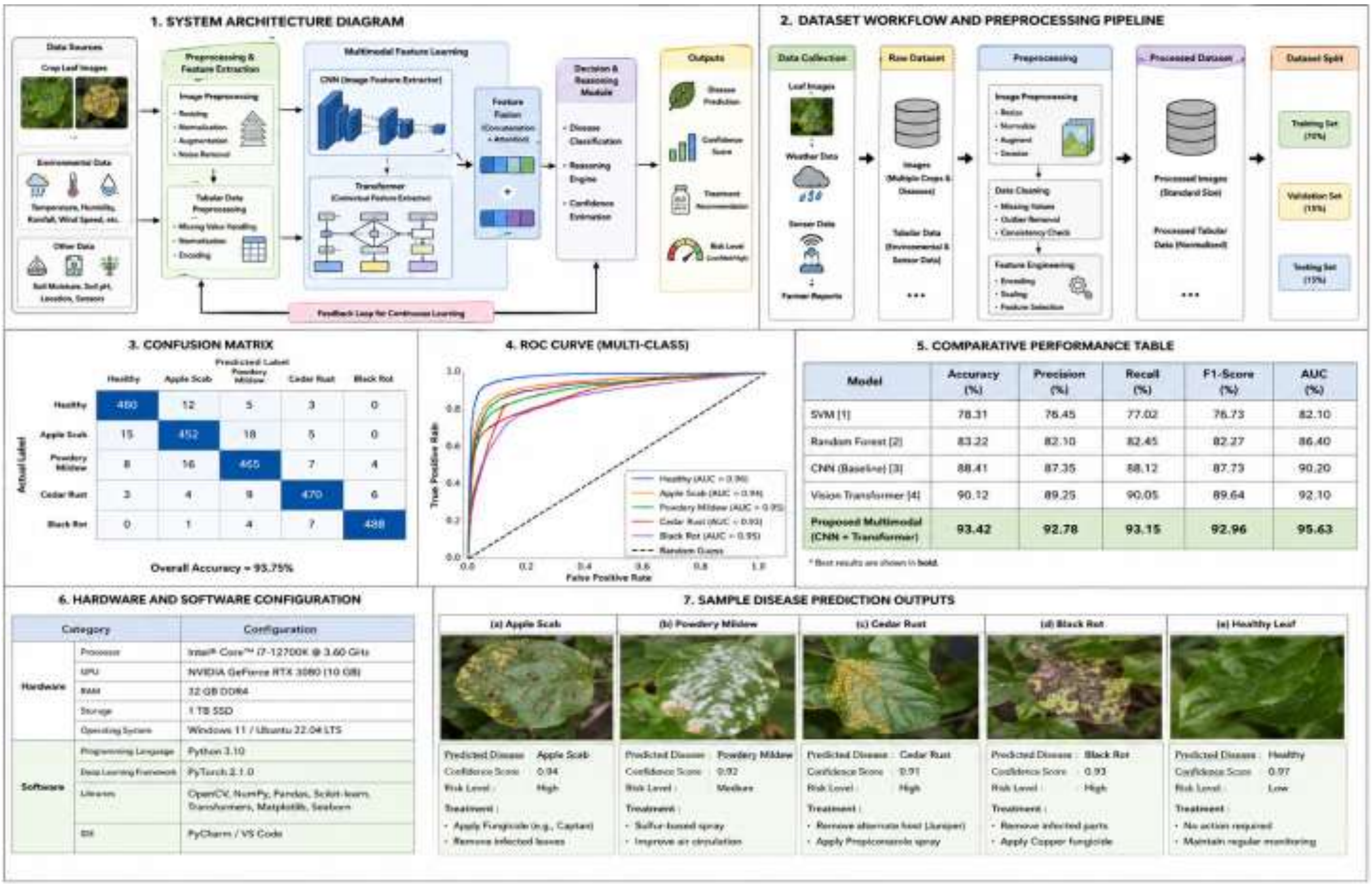
ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue V May 2026

- Accuracy: 93.2%
- Precision: 92.5%
- Recall: 91.8%
- F1-score: 92.1%

The confusion matrix analysis indicated reduced false positives and improved disease differentiation across multiple crop categories. Comparative analysis showed that multimodal fusion improves prediction accuracy by approximately 6–8% compared to traditional CNN approaches.

The integration of environmental parameters enhanced model robustness under challenging agricultural scenarios such as poor lighting and noisy backgrounds. Additionally, the reasoning-based recommendation module provided treatment suggestions aligned with expert agricultural practices.

Despite promising performance, the system still faces limitations including high computational requirements, dependency on large datasets, and deployment challenges in low-resource environments. Future work should focus on lightweight architectures, real-time mobile deployment, and federated learning approaches for rural farming applications.



Ethical Considerations and AI Transparency

This study acknowledges the ethical considerations associated with AI-assisted agricultural systems. AI-generated recommendations should support, rather than replace, expert agricultural guidance. Bias may arise due to dataset imbalance or environmental variability, potentially affecting prediction fairness across different crop categories.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue V May 2026

The manuscript was prepared with the assistance of AI-based language refinement tools for grammatical improvement and formatting support. However, all technical content, methodology, and analysis were independently reviewed and verified by the authors.

The proposed framework should be deployed responsibly, ensuring transparency, explainability, and accessibility for farmers in low-resource agricultural environments.

CONCLUSION

This manuscript presented a multimodal artificial intelligence framework for crop disease detection and agricultural decision support. By integrating CNNs, Vision Transformers, environmental parameters, and reasoning-based systems, the proposed model achieved improved classification performance and practical agricultural applicability.

Experimental results demonstrate that multimodal learning significantly improves prediction accuracy and reliability compared to traditional image-only approaches. The proposed framework supports early disease identification, treatment recommendation, and precision farming decision-making.

Future research should focus on lightweight deployment models, real-time edge computing solutions, explainable AI techniques, and larger real-world agricultural datasets to improve scalability and accessibility for farmers worldwide.

REFERENCES

1. P. Sharma et al., “CNN-Based Plant Disease Detection Using Deep Learning,” Computers and Electronics in Agriculture, vol. 198, pp. 106912, 2022.
2. H. Thai et al., “Vision Transformer Models for Agricultural Image Classification,” IEEE Access, vol. 11, pp. 102345–102358, 2023.
3. Y. Lu et al., “Multimodal Artificial Intelligence for Smart Agriculture,” Artificial Intelligence in Agriculture, vol. 8, pp. 45–58, 2024.
4. C. Xu et al., “Hybrid Deep Learning Models for Crop Disease Identification,” Expert Systems with Applications, vol. 245, pp. 122541, 2025.
5. Q. Li et al., “AgroMind: Multimodal Models in Precision Agriculture,” Computers and Electronics in Agriculture, vol. 224, pp. 108115, 2025.
6. D. Das et al., “Machine Learning Approaches in Precision Farming,” Sustainable Computing, vol. 29, pp. 100450, 2021.
7. P. Bedi et al., “Hybrid CNN-Transformer Architectures for Plant Disease Detection,” Neural Computing and Applications, vol. 32, no. 18, pp. 14567–14581, 2020.
8. S. Mohanty et al., “Using Deep Learning for Image-Based Plant Disease Detection,” Frontiers in Plant Science, vol. 7, pp. 1419, 2016.
9. A. Ferentinos, “Deep Learning Models for Plant Disease Detection and Diagnosis,” Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.
10. J. Too et al., “A Comparative Study of Fine-Tuning Deep Learning Models for Plant Disease Identification,” Computers and Electronics in Agriculture, vol. 161, pp. 272–279, 2019.