

Smart Farm AI: A Hybrid Deep Learning System for Intelligent Plant Disease Diagnosis in Agriculture

Gaurav Sudhakar Bharambe¹, Dipak Raghunath Bharate², Shivamkumar Chandrakant Mantri³,
Dr. D. G. Agrawal⁴

^{1,2,3}*Student, Department of Artificial Intelligence and Data Science*

⁴*Head of Department, Department of Artificial Intelligence and Data Science*

Shri Sant Gadge Baba College of Engineering and Technology, Bhusawal, Maharashtra, India

Abstract—Agriculture remains the backbone of developing economies, yet plant diseases continue to cause significant crop losses due to delayed or inaccurate diagnosis. This paper presents Smart Farm AI, a hybrid plant disease diagnosis system that integrates a Convolutional Neural Network (CNN) for primary classification, Google Gemini Vision for low-confidence analysis, the Grok Large Language Model (LLM) for natural-language explanation generation, and an embedding-based JSON memory system for efficient result caching. The CNN model, trained on the Plant Village dataset (80/20 train-validation split), predicts disease labels and confidence scores for uploaded leaf images. When confidence is below 70%, Gemini Vision performs deeper multimodal reasoning. Grok subsequently converts predictions into structured, farmer-friendly explanations encompassing disease causes, visible symptoms, treatment recommendations, and preventive measures. The embedding-based cache eliminates redundant computation by retrieving stored diagnoses for similar inputs at $\geq 95\%$ embedding similarity. A Stream lit-based interface provides one-click image upload and result display without requiring technical expertise. Experimental evaluation demonstrates that the hybrid system reliably handles both clear and challenging real-world images, outperforming single-model CNN baselines in accuracy, interpretability, and efficiency.

Index Terms—Convolutional Neural Network, Deep Learning, Gemini Vision, Grok LLM, Hybrid AI, Intelligent Agriculture, Plant Disease Detection, Plant Village, Retrieval Cache, Stream lit.

I. INTRODUCTION

Agriculture is a foundational pillar of the global economy and holds exceptional importance in

developing nations like India, where a significant portion of the population depends directly or indirectly on farming. Among the many threats to agricultural productivity, plant leaf diseases stand out as a leading cause of crop loss. If not detected in early stages, these diseases spread rapidly across fields, destroy large portions of crops, and cause severe economic harm to farming communities [4, 5]. Traditional disease identification relies on manual visual inspection by farmers or agricultural extension workers an approach that is time-consuming, prone to human error, and dependent on expert availability. Many farmers lack awareness of specific leaf symptoms and frequently misidentify diseases, leading to improper or delayed treatment [22]. Advances in computer vision and deep learning have introduced CNN-based plant disease classifiers that achieve high accuracy on benchmark datasets such as Plant Village [5, 6]. However, these systems suffer from two critical limitations: (1) performance degrades on complex real-world images with poor lighting, occlusion, or mixed symptoms; and (2) they provide only predicted labels and confidence scores insufficient for farmers who require causes, treatment steps, and preventive guidance [28]. Smart Farm AI addresses both limitations through a four-component hybrid pipeline: a CNN primary classifier, Gemini Vision for low-confidence image reasoning, Grok LLM for farmer-friendly explanation generation, and an embedding-based JSON memory cache for efficient repeated-query handling. The system is delivered through a Stream lit web interface requiring no technical expertise to operate. This paper presents the system architecture, methodology, experimental analysis, and comparative evaluation of Smart Farm AI.

II. RELATED WORK

A. Evolution of AI in Agricultural Disease Detection
Agricultural disease detection has evolved from manual visual inspection through handcrafted digital image processing (color segmentation, texture analysis) in the early 2000s, to machine learning classifiers (SVM, Random Forest, KNN) using engineered features, and finally to deep learning with CNNs from around 2015 onward [8, 9, 21]. The Plant Village dataset, introduced around 2015, enabled training of CNNs that automatically learn hierarchical visual patterns, achieving high accuracy and robustness in plant disease identification [5, 6]. From 2020 onward, agriculture entered a new phase with AI-powered decision support systems combining image classification with treatment guidance. The emergence of LLMs in 2023–2024 enabled conversational reasoning, multimodal analysis, and context-aware explanations, laying the foundation for hybrid models like Smart Farm AI [1, 25, 28].

B. Key Related Works

Kuska, Wahabzada & Paulus [1] demonstrated that LLMs can interpret multimodal agricultural data, automate routine analysis, and generate context-aware recommendations directly validating Smart Farm AI's Grok explanation module. Deputy, Passi & Jain [2] confirmed CNNs as the industry-standard classifier for plant disease identification, outperforming traditional image processing and ML techniques. Sun, Zhang & Hu [25] showed that multimodal fusion of vision and text models improves crop disease analysis reliability consistent with Smart Farm AI's Gemini Vision integration. Wang, Chen & Zhao [28] presented explainable AI techniques for plant disease detection, supporting the LLM-based interpretability layer of the proposed system.

C. Research Gap

Existing systems predominantly rely on single-stage CNN classification without explainability, fail to handle low-confidence or poor-quality inputs, and recompute predictions for repeated similar queries. No prior published open-source system integrates CNN classification, vision-model fallback, LLM explanation, and embedding-based caching into a unified, deployable agricultural diagnostic platform. Smart Farm AI addresses this gap.

III. SYSTEM ARCHITECTURE

A. Overview

Smart Farm AI is designed as a sequential hybrid pipeline with four major layers:

1. Input Acquisition,
2. Prediction and Analysis,
3. Memory and Caching, and
4. Presentation.

The architecture is modular, enabling independent enhancement or replacement of each component.

A user uploads a leaf image via the Stream lit interface. The system first checks the embedding-based JSON memory cache for a similar previously processed image. If no match is found, the CNN predicts the disease and confidence score. If confidence $\geq 70\%$, the prediction is forwarded to Grok for explanation. If confidence $< 70\%$, Gemini Vision performs deeper multimodal analysis before Grok generates the final diagnosis. All new results are stored in the cache for future retrieval.

B. CNN Disease Classification Module

The CNN model follows a sequential architecture: an Input Layer accepting $128 \times 128 \times 3$ RGB images; three Conv2D layers with 32, 64, and 128 filters respectively for hierarchical feature extraction (spots, discolorations, texture irregularities); Max Pooling layers after each convolutional layer for spatial dimension reduction; a Flatten layer; a Dense layer with 128 units; a Dropout layer (rate = 0.3) for regularization; and a Soft max Output Layer for multi-class classification. The model is compiled with Adam optimizer, Categorical Cross entropy loss, and Accuracy metric. Training runs for 10 epochs with Image Data Generator augmentation (rotation, zoom, horizontal flip, brightness adjustment).

C. Confidence Evaluation and Routing

A threshold-based decision module evaluates CNN output confidence: predictions $\geq 70\%$ are routed to the High-Confidence Path (CNN \rightarrow Grok LLM); predictions $< 70\%$ are routed to the Low-Confidence Path (CNN \rightarrow Gemini Vision \rightarrow Grok LLM). This conditional routing prevents unreliable CNN outputs from reaching the user and reserves advanced vision analysis for genuinely uncertain cases, optimizing computational efficiency.

D. Gemini Vision Module

Gemini Vision is activated exclusively for low-confidence predictions. It performs multimodal image reasoning to produce: crop type identification, health status assessment, disease name, confidence estimation (high/medium/low), severity estimation (mild/moderate/severe), and symptom descriptions (yellowing, necrosis, circular spots, edge burns). This richer descriptive output is then passed to Grok for refinement into farmer-friendly text, combining visual understanding with natural language generation.

E. Grok LLM Explanation Module

Grok receives either the CNN label and confidence (high-confidence path) or the Gemini Vision descriptive output (low-confidence path) and generates a structured diagnosis containing: disease name, causes and infection source, visible symptoms, three actionable treatment steps (organic and chemical options), and preventive measures. Prompt engineering ensures simple language suitable for non-technical users. This transforms Smart Farm AI from a classifier into a complete agricultural diagnostic assistant.

F. Embedding-Based JSON Memory System

Before any prediction, the system extracts a numerical feature vector (embedding) from the uploaded image using the CNN model's intermediate layer. This embedding is compared against stored embeddings in the JSON memory file using cosine similarity. If similarity $\geq 95\%$, the stored diagnosis is returned instantly, skipping the CNN, Gemini Vision, and Grok steps entirely. After generating a new diagnosis, the system stores the embedding, predicted disease, confidence score, and timestamp for future retrieval. This mechanism significantly reduces response time and ensures consistency for repeated or similar inputs.

IV. DATASET AND METHODOLOGY

A. Dataset

The CNN is trained on the Plant Village dataset, one of the most widely accepted benchmarks for agricultural AI research. The dataset contains thousands of JPEG/PNG leaf images organized into folders by crop-disease class, including Tomato Bacterial Spot, Tomato Early Blight, Potato Late Blight, Potato Healthy, and Pepper Leaf Spot, among

others. Images are high-resolution RGB, photographed against mostly uniform backgrounds for consistent feature extraction.

B. Preprocessing Pipeline

All images are pre-processed through: RGB format conversion; resizing to 128×128 pixels; pixel normalization to the $[0, 1]$ range; and data augmentation via Image Data Generator (rotation, zoom, horizontal flip, brightness adjustment). The dataset is split 80% training and 20% validation, ensuring unbiased performance evaluation. Validation loss and accuracy are monitored at each epoch to detect overfitting or underfitting.

C. Model Training

The CNN is trained for 10 epochs using Adam optimizer and Categorical Cross entropy loss. GPU acceleration is used when available. Training graphs of accuracy and loss are monitored to ensure stable convergence without sudden fluctuations. The Dropout layer (rate = 0.3) and data augmentation together mitigate overfitting by preventing the model from memorizing training samples. The close alignment between training and validation accuracy in the resulting curves confirms good generalization.

D. System Integration

The complete pipeline is implemented in Python 3 using TensorFlow/Keras for the CNN, the Google Gemini Vision API for multimodal analysis, the Grok LLM API for explanation generation, and Streamlit for the user interface. NumPy, Pandas, OpenCV, and Matplotlib support data handling and visualization. The modular codebase allows each component to be upgraded independently.

V. RESULTS AND DISCUSSION

A. CNN Model Performance

The CNN model achieves high training accuracy and stable validation accuracy, with training and validation curves closely aligned throughout 10 epochs confirming effective generalization with minimal overfitting. The loss curves decrease gradually and smoothly for both training and validation sets, with no sudden spikes, indicating stable gradient convergence. The Dropout layer (0.3) and augmentation are effective in preventing memorization of training patterns.

On clear, well-lit leaf images closely matching Plant Village conditions, the CNN routinely produces confidence scores above 80%, routing directly to the Grok explanation path. On complex or ambiguous images (poor lighting, partial damage, mixed symptoms), confidence drops below 70%, triggering Gemini Vision analysis and demonstrating the necessity of the hybrid routing mechanism.

B. Hybrid System Evaluation

Testing across multiple leaf images under varied real-world conditions demonstrates that the hybrid system outperforms single-CNN baselines on three dimensions. First, accuracy: Gemini Vision successfully handles complex or unclear cases where CNN confidence is insufficient, reducing misclassification for challenging inputs. Second, interpretability: Grok LLM consistently generates structured diagnoses with disease causes, symptoms, treatment steps, and prevention guidelines, validated by agricultural domain review. Third, efficiency: the embedding-based cache retrieves stored diagnoses in under 1 second for similar images, eliminating full pipeline re-execution and demonstrating significant latency reduction compared to recomputing CNN + Gemini + Grok for every query.

C. Comparative Analysis

Traditional CNN-only systems provide class labels and confidence scores insufficient for non-technical farmers who need actionable guidance. They degrade on real-world images deviating from controlled datasets and recompute results for every input. Smart Farm AI addresses all three limitations: Gemini Vision handles challenging inputs; Grok LLM provides farmer-friendly explanations; and the JSON memory cache eliminates redundant computation. The hybrid architecture also improves robustness by not relying on any single model's performance ceiling, making the system suitable for deployment in unpredictable field conditions.

D. System Functionality

The Stream lit interface delivers five key functions:

1. Home interface with system overview;
2. image upload with drag-and-drop and optional crop name input;
3. CNN-based prediction display with confidence score;

4. Gemini Vision output for low-confidence cases;
5. cached result retrieval for similar images; and
6. Farm Diagnosis History page showing past predictions.

The sidebar provides memory reset functionality and developer credits. The interface is accessible to farmers without any technical background.

VI. DISCUSSION

A. Strengths

The primary strength of Smart Farm AI is its adaptive multi-layer decision pipeline, which intelligently allocates computational resources based on prediction confidence. The CNN handles the majority of clear images efficiently; Gemini Vision is reserved for genuinely uncertain cases; and the cache eliminates redundant processing entirely for repeated queries. This design minimizes average inference time while maximizing reliability across diverse input conditions. The LLM-based explanation layer fundamentally changes the system's utility profile. Rather than requiring farmers to consult external experts to interpret a predicted label, Smart Farm AI delivers actionable guidance directly at the point of diagnosis. This addresses the most critical barrier to adoption of AI tools in rural agricultural communities: the gap between technical output and practical usability.

B. Limitations

CNN performance is sensitive to input image quality; poor lighting, blur, or incorrect camera angles reduce confidence and increase reliance on the Gemini Vision fallback. The system requires internet connectivity for both the Gemini Vision API and the Grok LLM API, limiting offline use in remote areas. The JSON-based memory system does not scale efficiently beyond small to medium deployments; large-scale production use would require migration to a vector database. The CNN model is trained on Plant Village's controlled conditions, which may not fully represent the visual diversity of real-world field images, particularly for rare or region-specific diseases.

VII. FUTURE WORK

The JSON memory system will be replaced with a vector database (e.g., Chroma, Pinecone) to support large-scale deployment with efficient similarity

search. A mobile application will extend on-field image capture and real-time diagnosis to smartphones, improving access for farmers in rural areas. The training dataset will be expanded to include diverse crops, varied environmental conditions, and rare disease classes to improve model generalization.

IoT sensor integration (soil moisture, temperature, humidity) will enable continuous crop health monitoring and predictive disease alerts. Offline functionality will be implemented using on-device model inference to support low-connectivity environments. Multilingual output in regional Indian languages will improve adoption across linguistic communities. Future versions will integrate fertilizer and pesticide recommendation modules, weather-based disease outbreak prediction, and cloud deployment for multi-user scalability.

VIII. CONCLUSION

This paper presented Smart Farm AI, a hybrid plant disease diagnosis system that integrates CNN-based classification, Gemini Vision multimodal analysis, Grok LLM explanation generation, and embedding-based JSON memory caching into a unified, farmer-accessible platform. The system addresses the two critical limitations of existing CNN-only approaches: degraded performance on challenging real-world images, and the absence of actionable, farmer-friendly explanations.

The adaptive confidence-based routing mechanism ensures that computational resources are allocated efficiently fast CNN inference for clear images, deeper Gemini Vision reasoning for uncertain cases, and instant cache retrieval for repeated queries. The Grok LLM transforms raw predictions into structured diagnoses covering disease causes, symptoms, treatment steps, and preventive measures, eliminating the need for expert consultation at the point of diagnosis. Experimental evaluation confirms that the hybrid architecture outperforms single-model CNN baselines in accuracy, interpretability, and efficiency across varied real-world conditions. Smart Farm AI demonstrates strong potential for real-world agricultural deployment and contributes toward intelligent, sustainable farming solutions.

ACKNOWLEDGMENT

The authors thank Principal Dr. R.B. Barjibhe for providing the necessary facilities and infrastructure. Special gratitude is expressed to guide and Head of Department Dr. D.G. Agrawal for his invaluable guidance, support, and encouragement throughout this project. The authors also thank all faculty and staff members of the Department of Artificial Intelligence and Data Science, Shri Sant Gadge Baba College of Engineering and Technology, Bhusawal, for their continued support.

REFERENCES

- [1] M. T. Kuska, M. Wahabzada, and S. Paulus, "AI for Crop Production – Where Can Large Language Models (LLMs) Provide Substantial Value?" *Computers and Electronics in Agriculture*, 2024.
- [2] K. V. Deputy, K. Passi, and C. K. Jain, "Crop Disease Detection Using Deep Learning Techniques on Images," *Journal of Computer Science*, 2023.
- [3] K. Jhaharia, P. Mathur, S. Jain, and S. Nijhawana, "Crop Yield Prediction Using Machine Learning and Deep Learning Techniques," *Procedia Computer Science*, 2023.
- [4] S. Sladojevic et al., "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification," *Computational Intelligence and Neuroscience*, 2016.
- [5] P. Mohanty, D. P. Hughes, and M. Salathé, "Inference of Plant Diseases from Leaf Images through Deep Learning," *Frontiers in Plant Science*, 2016.
- [6] S. Ferentinos, "Deep Learning Models for Plant Disease Detection and Diagnosis," *Computers and Electronics in Agriculture*, 2018.
- [7] J. Too et al., "A Comparative Study of Fine-Tuning Deep Learning Models for Plant Disease Identification," *Computers and Electronics in Agriculture*, 2019.
- [8] N. Brahimi, K. Boukhalifa, and A. Moussaoui, "Deep Learning for Tomato Diseases: Classification and Symptoms Visualization," *Applied Sciences*, 2017.
- [9] J. Amara, B. Bouaziz, and A. Algergawy, "A Deep Learning-Based Approach for Banana Leaf

- Disease Classification,” in Data Analytics Symposium (DAS), 2017.
- [10] M. Rangarajan and R. Purushothaman, “Disease Classification in Potato Plants Using Convolutional Neural Networks,” Machine Learning for Applications, 2021.
- [11] X. Ma, Z. Li, and J. Li, “A Review of Deep Learning Applications in Agriculture,” Computers and Electronics in Agriculture, 2021.
- [12] Z. Zhang, F. Xing, and Y. Su, “Automatic Leaf Disease Detection Using Improved CNN,” Information Processing in Agriculture, 2022.
- [13] W. Fuentes and S. R. Y. Kusumaputri, “Tomato Leaf Disease Detection Using MobileNet,” International Journal of Advanced Computer Science and Applications, 2021.
- [14] P. Arsenovic et al., “Solving Current Limitations of Deep Learning-Based Plant Disease Recognition with Synthetic Data,” Scientific Reports, 2019.
- [15] T. Koirala, B. Dong, and D. Walsh, “Deep Learning-Based Severity Estimation of Plant Disease,” Biosystems Engineering, 2020.
- [16] P. V. Singh, A. Mathur, and R. Sharma, “Performance Assessment of Deep Learning Models for Crop Disease Detection,” Neural Computing and Applications. Springer, 2021.
- [17] N. Lakshmi and B. Priyanka, “A Review on Computer Vision-Based Systems for Smart Agriculture,” Information Processing in Agriculture, 2021.
- [18] C. Sun, B. Zhang, and S. Hu, “Multimodal Fusion for Crop Disease Analysis Using Vision and Text Models,” Computers and Electronics in Agriculture, 2023.
- [19] X. Wang, L. Chen, and Y. Zhao, “Explainable AI Techniques for Plant Diseases Detection,” Artificial Intelligence Review. Springer, 2023.