

Hybrid CNN and LLM for Image-Based Classification of Plant Leaf Diseases

Frenky Riski Gilang Pratama¹, Sugiarto Surono², Aris Thobirin³
^{1,2,3}Department of Mathematics, Universitas Ahmad Dahlan, Indonesia

Article Info

Article history:

Received Mar 04, 2026

Revised Mar 24, 2026

Accepted Apr 17, 2026

Keywords:

EfficientNetV2-M

GPT-4o

PlantVillage Dataset

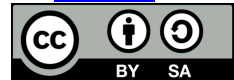
Hybrid CNN and LLM

Plant Disease Classification

ABSTRACT

Plant diseases have continued to threaten agricultural productivity, while manual inspection methods have remained inefficient and prone to subjectivity. This study proposed and assessed a hybrid framework integrating a Convolutional Neural Network (CNN) with a Large Language Model (LLM) to perform image-based plant leaf disease classification accompanied by interpretable diagnostic explanations. EfficientNetV2-M was employed as the visual backbone and trained on 11 selected classes of apple, grape, and potato leaf images derived from the PlantVillage dataset. A structured data splitting strategy was applied to ensure reliable model validation and unbiased testing. The classification capability of the CNN component was examined through standard multi-class evaluation indicators, including class-wise predictive consistency and error distribution analysis. Experimental results indicated that the model delivered highly consistent predictions, reaching a peak test accuracy of 99.79%, reflecting its robustness in distinguishing visually similar disease patterns. To overcome the black-box limitation, prediction outputs were transformed into structured prompts and processed by GPT-4o to generate contextual explanations. The generated narratives systematically described observable symptoms, highlighted distinguishing characteristics, and suggested initial management actions. Overall, the proposed hybrid system demonstrated that combining high-performance visual recognition with language-based reasoning enhanced both diagnostic reliability and interpretability in digital agriculture applications.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Frenky Riski Gilang Pratama,
Department of Mathematics,
Universitas Ahmad Dahlan,

Jl. Ringroad Selatan, Kragilan, Tamanan, Banguntapan, Bantul, Special Region of Yogyakarta.

Email: 2200015030@webmail.uad.ac.id

1. INTRODUCTION

A serious threat to agricultural productivity is posed by plant diseases, with significant crop losses being one of the consequences and global food security being another [1], [2]. The rapid spread of plant pathogens, coupled with environmental factors such as humidity and temperature, makes early detection crucial to prevent further economic losses [3]. Conventional approaches to the identification of leaf diseases, including visual examination by experts, are inherently limited in their effectiveness. These methods are characterised by their slow pace, the substantial labour intensity required, and a susceptibility to human error. Consequently, their application to vast agricultural areas is rendered ineffective [4]. Therefore, there is an urgent need for an automated and effective system to detect plant diseases [5]. Recent advances in deep learning, particularly convolutional neural networks (CNNs), have demonstrated strong capability in extracting hierarchical visual features directly from leaf images, enabling highly accurate classification without handcrafted feature engineering [6], [7]. The integration of artificial intelligence into

precision agriculture frameworks further supports proactive monitoring and data-driven disease management [8].

In recent years, substantial methodological improvements have been reported. Transfer learning strategies have significantly enhanced classification performance by leveraging knowledge from large-scale pre-trained models. [4]. Among recent works, Shafik et al. [9], proposed a transfer learning-based framework combined with fusion voting across multiple CNN architectures, achieving an accuracy of 97.79% while improving model robustness for sustainable agriculture applications. Beyond purely visual models, emerging research has begun incorporating large language models (LLMs) into agricultural diagnostics. Pan et al. [10], introduced ChatLeafDisease (ChatLD), a GPT-4o-based framework utilizing chain-of-thought prompting and a disease knowledge base to assist both classification and explanation generation. These developments indicate a shift from purely predictive systems toward more reasoning-oriented and multimodal approaches in plant disease diagnosis.

However, at the same time, there are still significant gaps in research. Most CNN models today operate exclusively in the visual domain, without leveraging text information or in-depth domain knowledge [11], [12]. Furthermore, the inherent opacity of CNN and transformer architectures limits their transparency, making it challenging for end-users to clearly understand the rationale behind the generated predictions [9], [13]. For example, Pal et al. [1], and many previous studies have highlighted that LLMs are optimized for specific datasets and little attention is paid to model interpretability, resulting in low user confidence. In the agricultural sector, studies on LLMs are still rare [14]. This combination not only reduces accuracy but also highlights the urgent need for interpretable methods, which have not been widely discussed in the current literature.

Based on this gaps, this study proposes a hybrid CNN and LLM architecture that integrates CNN-based visual feature extraction with LLM-driven textual explanation and contextual interpretation [15], [16]. The CNN component functions as a robust classifier that identifies disease categories from leaf images, while the LLM module transforms prediction outputs into structured, symptom-based explanations. This integration is designed not only to maintain high classification accuracy but also to enhance interpretability and practical usability. Unlike conventional CNN-only systems, the proposed approach enables the generation of diagnostic narratives and recommendation-oriented outputs, aligning with the broader development of explainable AI in smart agriculture [17], [18].

The specific objective of this study is to develop and evaluate a hybrid model of CNN and LLM to classify plant diseases based on images of three main types of crops, namely apples, grapes, and potatoes. The model is designed not only to achieve high classification efficiency, but also to generate specific textual diagnostic explanations that correspond to visual symptoms and are easy to understand even by non-expert users, in line with the development of diagnostic systems powered by explainable AI in digital agriculture [19]. From a conceptual point of view, this study tests the hypothesis that a close integration between the CNN module (as a visual feature extractor and disease labeler) and the LLM module (as a generator of explanations and resulting recommendations) will result in a more accurate, informative, and transparent plant disease diagnosis system compared to an approach that relies solely on CNN, and will contribute to faster and more accurate decision-making in the context agriculture.

2. RESEARCH METHOD

This section describes in a structured manner the methods used to construct and evaluate the proposed CNN and LLM hybrid model. The research flow began with the collection and pre-processing of PlantVillage data, followed by training and evaluation of the EfficientNetV2-M architecture as a leaf image classifier. Once the CNN model is trained, new leaf images from users are processed at the inference stage to produce predictions of disease classes. These predictions are then converted into structured prompts that are fed to the GPT-4o model, so that the LLM can generate textual diagnostic explanations that are informative to the user. The series of stages is summarized in the flowchart in Figure 1.

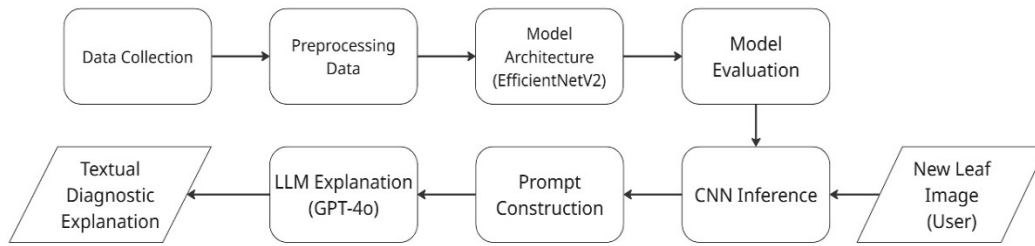


Figure 1. Research Flowchart

2.1. Dataset and Preprocessing

The dataset used in this study is PlantVillage [20], a benchmark resource for plant disease classification research. Although the original dataset comprises 38 categories, this study focuses on 11 selected classes representing three types of plants, namely apple, grape, and potato. In total, 9,385 RGB leaf images with an approximate resolution of 500×500 pixels were utilized. The detailed distribution of images across all selected classes is summarized in Table 1.

Table 1. PlantVillage Dataset

Class	Total
Apple Scab	630
Apple Black Rot	621
Apple Cedar Rust	275
Apple Healthy	1645
Grape Black Rot	1180
Grape Esca/Blac Measles	1383
Grape Leaf Blight/Isariopsis Leaf Spot	1076
Grape Healthy	423
Potato Early Blight	1000
Potato Late Blight	1000
Healthy Potatoes	152
Total	9385

As presented in Table 1, the number of samples varies across classes, reflecting a degree of class imbalance. Representative examples of leaf images from each disease category are displayed in Figure 2.

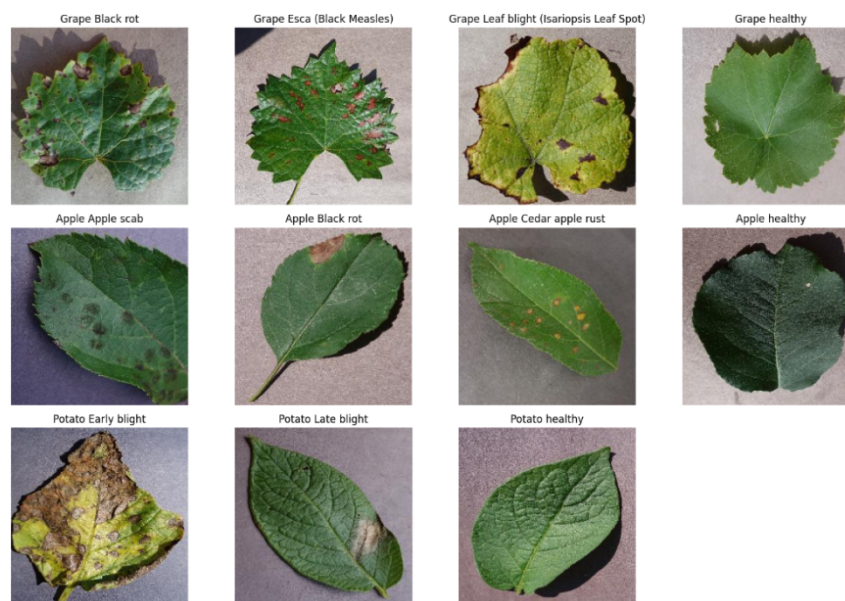


Figure 2. Example dataset per class

Each image was rescaled to 320×320 pixels to match the input resolution of EfficientNetV2-M using bilinear interpolation. The images were then converted into tensor format and normalized per RGB channel using ImageNet statistics ($\mu = 0,485, 0,456, 0,406$ and $\sigma = 0,229, 0,224, 0,225$ for R, G, B channels). Normalization is carried out by linear transformations such as equation 1.

$$I'_c(x, y) = \frac{I_c(x, y) - \mu_c}{\sigma_c} \quad (1)$$

Where:

$I'_c(x, y)$ = standardized pixel intensity after channel-wise normalization,

$I_c(x, y)$ = raw pixel intensity at coordinate (x, y) in channel c ,

μ_c = channel-specific mean derived from ImageNet statistics,

σ_c = channel-specific standard deviation derived from ImageNet statistics,

$c \in \{R, G, B\}$

This transformation adjusts the input distribution to be consistent with the statistical properties expected by ImageNet-pretrained CNN models [21], [22]. The dataset was partitioned into three subsets: 80% for training, 10% for testing, and 10% for validation. Augmentation is applied only to training data, including vertical ($p = 0.2$) and horizontal ($p = 0.5$) flips, $\pm 15^\circ$ rotation, and affine transformations (shift $\leq 10\%$, zoom 0.9–1.1). Photometric adjustments involve brightness, contrast, and saturation variations ($\pm 20\%$) and hue shifts (± 0.1). Images are then center-cropped to 320×320 for model input. This approach enhances robustness to spatial and illumination variability while maintaining unbiased evaluation on validation and test sets [23], [24].

2.2. Model Architecture and Training

To achieve a balance between computational efficiency and classification performance, this study employs EfficientNetV2-M as the CNN backbone for feature extraction and preliminary leaf disease classification. EfficientNetV2-M represents a medium-scale variant of the EfficientNetV2 family that integrates MBConv and Fused-MBConv blocks and contains approximately 51 million parameters [25], [26]. The efficiency of this architecture has enabled strong results across a range of image classification problems, including applications related to plant disease and pest identification, as reported in recent studies using EfficientNetV2-based transfer learning. [27], [28]. The core 2D convolution operation in this network is formulated as:

$$Y_{(i,j)} = \sum_{m=1}^k \sum_{n=1}^k I_{(i-1) \cdot s + m, (j-1) \cdot s + n} * K_{(m,n)} \quad (2)$$

Where:

$Y_{(i,j)}$ = output convolution at position (i, j) ,

I = input image,

k = kernel dimension (kernel size),

K = convolution kernel of size $k \times k$,

s = stride

(m, n) = kernel indices,

(i, j) = spatial position in the output feature map.

The spatial size of the resulting feature map produced by a convolution layer can be determined using Equation 3:

$$O = \left\lceil \frac{N - F + 2P}{s} \right\rceil + 1 \quad (3)$$

Where:

N = spatial dimension of the input feature map (height or width),

F = kernel size,

P = amount of zero-padding applied to the input,

s = stride,

O = output dimension.

In EfficientNetV2-M, most convolution operations use depthwise convolution followed by pointwise (1×1) convolution to improve efficiency. Each feature map is then passed through the Swish activation function:

$$\text{swish}(z) = z \cdot \sigma(z) = \frac{z}{1 + e^{-z}} \quad (4)$$

Where:

z = input to the activation function,
 $\sigma(z)$ = sigmoid function,
 e = Exponential value.

Swish is smooth and non-monotonic, offering improved representational capability compared to ReLU with minimal additional computational overhead [29], [30]. In the final stage of the network, Global Average Pooling (GAP) compresses each feature map into a single scalar, resulting in a feature vector of dimension 1280, which corresponds to the final representation of EfficientNetV2-M. The pooled feature value for channel k is defined as:

$$g_k = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W f_k(u, v) \quad (5)$$

Where:

$f_k(u, v)$ = activation value at spatial coordinate (u, v) in channel k ,
 H = height of the feature map,
 W = width of the feature map,
 g_k = averaged feature value for channel k .

Where that is then passed to the layer is fully connected to the 11 neurons and the activation of the softmax to produce a class probability distribution of class $\hat{y} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{11}]$.

Training was conducted in two stages using ImageNet-21k pretrained weights [31]. Initially, convolutional layers were frozen and only the final classifier was optimized using Adam with a learning rate of 10^{-3} . After convergence, all layers were unfrozen for end-to-end fine-tuning with a reduced learning rate of 10^{-5} , following a common two-stage fine-tuning strategy [32]. The categorical cross-entropy loss function is defined as:

$$\ell = - \sum_{i=1}^{11} t_i \log(p_i) \quad (6)$$

Where:

p_i = predicted softmax probability for class i ,
 t_i = binary indicator representing the true label of class i in one-hot encoded form,
 i = class index (from 1 to 11),
 ℓ = total classification loss.

The loss reaches its minimum when the predicted probability of the correct class approaches 1. During backpropagation, gradients are efficiently computed across all layers, while Adam adaptively updates parameters using first moment (β_1) and second moment (β_2) estimates of the gradients, resulting in stable optimization for deep neural networks [33], [34].

2.3. Model Performance Evaluation

Model performance was evaluated after the training process using a separate test dataset consisting of 939 leaf images, with the results then used to inform future model development. This evaluation uses a standard multi-class classification framework and key metrics of accuracy, precision, recall, and F1 scores for each class. Mathematically, it can be defined as follows:

$$\text{Accu} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$\text{Prec}_i = \frac{TP_i}{TP_i + FP_i} \quad (8)$$

$$\text{Rec}_i = \frac{TP_i}{TP_i + FN_i} \quad (9)$$

$$F1_i = \frac{2 \times Prec_i \times Rec_i}{Prec_i + Rec_i} \quad (10)$$

Where: TP_i, TN_i, FP_i, FN_i , and represent true positives, true negatives, false positives, and false negatives for class i , respectively. An 11×11 confusion matrix was constructed to analyze misclassification patterns. A classification report summarizing precision, recall, F1-score, and support (including macro and weighted averages) was generated automatically using the `sklearn.metrics` module from the `scikit-learn` library.

2.4. GPT-4o LLM Integration

In the final stage, the hybrid system integrates GPT-4o to generate text-based diagnostic explanations. The CNN outputs the predicted disease labels k and probabilities p_k , which are combined with characteristic symptom lists and structured task instructions (e.g., “agronomic assistant”). These elements are compiled into a prompt that guides the LLM to produce an explanatory description sequence y . The probability of the generated text is formulated as:

$$P(y|\text{prompt}) = \prod_{t=1}^T P(y_t|y_{<t}, \text{prompt}) \quad (11)$$

Where:

P = probability of the generated text sequence,

y_t = token at position t ,

$y_{<t}$ = all previous tokens before t ,

prompt = structured input containing CNN predictions and task instructions,

T = total number of tokens.

This probabilistic formulation aligns with modern autoregressive language models for sequential generation tasks [35]. By integrating GPT-4o, the system supports explainable AI (XAI) principles: the CNN performs accurate visual recognition, while the LLM transforms predictions into clear diagnostic explanations, highlights key symptoms, differentiates similar diseases, and provides practical early recommendations [36]. This approach, while still relying on general knowledge in GPT-4o (not local agronomic adjustments), provides a clear path to reducing the “black box” in the classification model and increasing user confidence in CNN–LLM-based automated diagnostic systems.

3. RESULTS AND DISCUSSION

This section summarizes the experimental findings of the proposed CNN–LLM hybrid framework, highlighting the technical implementation, model performance, and readability of the generated diagnostic explanations.

3.1. Technical Implementation and Reproducibility

All experiments were conducted in a controlled and well-documented environment to ensure reproducibility. EfficientNetV2-M was implemented in PyTorch and trained on a T4 GPU in Google Colab using ImageNet pre-trained weights. Training was conducted in two stages. First, all feature extraction layers were frozen and only classification was trained with a learning rate of 1×10^{-3} . Second, the entire network was fine-tuned with a lower learning rate of 1×10^{-5} . The Adam optimizer and ReduceLROnPlateau scheduler were applied in both phases, with key hyperparameters including a batch size of 8, 20 training epochs, stochastic regularization, and stochastic augmentation. The experiment uses the PlantVillage color dataset, limited to 11 classes of apple, grape, and potato leaves (including healthy samples). A publicly accessible Streamlit-based system demonstration is available at <https://llm-explain.streamlit.app/>.

3.2. CNN Model Performance

The performance of the EfficientNetV2-M model was evaluated based on training, validation, and testing results on 11 leaf disease classes. This section summarizes the data distribution and key evaluation metrics, including learning curves, confusion matrices, and classification reports, to assess generalization capabilities and potential error patterns.

Table 2. Splitting Data

Class	Train	Validation	Test	Total
Apple Scab	504	63	63	630
Apple Black Rot	497	62	62	621
Apple Cedar Rust	220	28	27	275
Apple Healthy	1316	164	165	1645
Grape Black Rot	944	118	118	1180
Grape Esca/Blac Measles	1106	138	139	1383
Grape Leaf Blight/Isariopsis Leaf Spot	861	108	107	1076
Grape Healthy	338	42	43	423
Potato Early Blight	800	100	100	1000
Potato Late Blight	800	100	100	1000
Healthy Potatoes	122	15	15	152
Total	7508	938	939	9385

Table 2 details the distribution of training, validation, and testing samples for each class. Although most categories contain a substantial number of observations, class imbalance is evident, particularly in the Healthy Potatoes category, which may affect metric stability due to limited sample support.

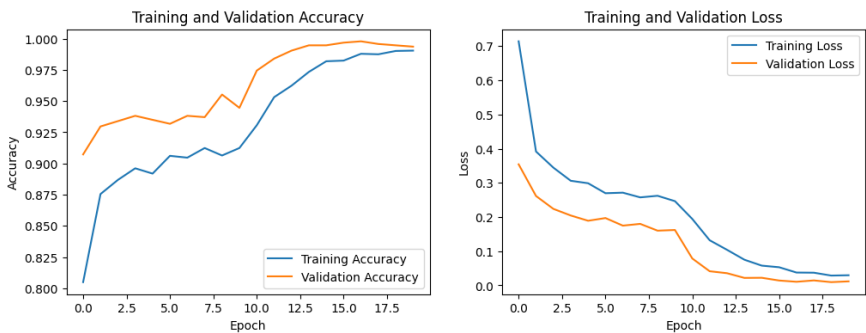


Figure 3. History accuracy and loss

Figure 3 illustrates the training and validation accuracy and loss trajectories over 20 epochs, comprising 10 epochs of head training followed by 10 epochs of fine-tuning. Accuracy increased progressively from approximately 0.80–0.95 in the initial epochs to nearly 0.99, while loss values declined sharply and stabilized at minimal levels. The marginal gap between training and validation curves indicates effective transfer learning without observable overfitting. The final model achieved an overall accuracy of 0.9979 with a loss of 0.0103, demonstrating high discriminative capability on the PlantVillage dataset.

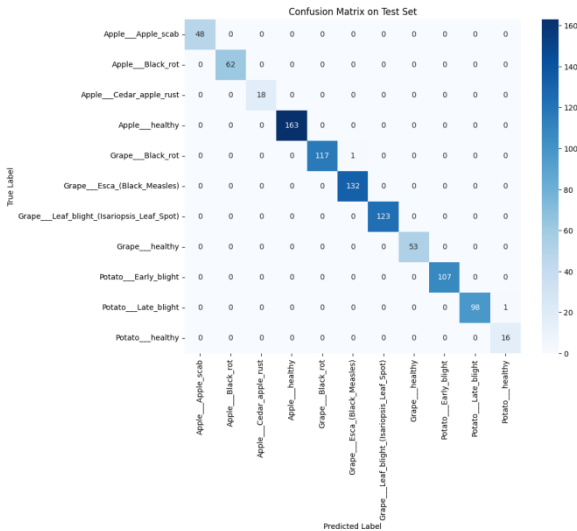


Figure 4. Confusion Matrix

Figure 4 presents the confusion matrix for the test set, where non-zero elements are predominantly concentrated along the principal diagonal, indicating correct classification for the vast majority of samples. Only minor misclassifications were observed, such as a Potato Late Blight instance predicted as Healthy Potatoes.



Figure 2. Classification report

Figure 5 displays precision, recall, and F1-score values for each class, all approaching 1.00. Slightly lower scores in the Healthy Potatoes class are attributable to its limited test samples, where a single misclassification significantly influences the metric. Overall, these results demonstrate strong generalization performance within a controlled dataset, however further validation on real-world field data remains essential.

3.3. LLM-Based Diagnostic Explanations

To enhance interpretability, the CNN classifier was integrated with GPT-4o, enabling each prediction to be accompanied by a structured textual explanation. While the CNN provides disease labels and associated probabilities, the LLM translates these outputs into agronomically meaningful narratives by identifying key visual symptoms and relating them to established plant pathology knowledge. This framework alleviates the “black-box” characteristic of CNN models by improving the clarity and traceability of the resulting decisions.

Table 3. Explanation for each plant

Class	Explainability
Predicted: Apple___Apple_scab (94.70%) 	The CNN inference results indicate a high probability (0.947) that the image depicts Apple Scab, a common disease affecting apple trees. This classification is supported by the presence of characteristic symptoms such as olive-gray to brown scabby lesions on the leaves and fruit, which are rough to the touch. These lesions often coalesce into larger patches and can lead to deformation of the fruit, a distinctive sign of Apple Scab. What sets Apple Scab apart from other diseases, like Early Blight or Late Blight, is the specific texture and color of the lesions. Unlike Early Blight, which features concentric rings, or Late Blight, known for its water-soaked appearance, Apple Scab lesions are hard and rough with an olive-gray to brown hue, often appearing on young leaves and fruit. For managing Apple Scab, it's advisable to sanitize fallen leaves to reduce the spread of the pathogen. Additionally, planting resistant apple cultivars and applying appropriate fungicides can be beneficial preventive measures. However, it is important to consult with local agricultural extension services for specific fungicide recommendations, as this guidance may vary based on regional practices and regulations.

Class	Explainability
Predicted: Potato __Early_blight (99.92%) 	The image has been classified with a high confidence level as 'Potato __Early_blight', primarily due to the presence of characteristic symptoms associated with this disease. Early blight in potatoes is identified by brown leaf spots that exhibit a clear concentric ring pattern, resembling a target. These spots often develop on older leaves first, which aligns with the typical progression of this fungal disease. Additionally, the presence of a chlorotic halo surrounding the lesions is a key symptom that distinguishes early blight from other potato diseases. A distinguishing feature of early blight is the well-defined concentric rings within the brown spots, a pattern not typically observed in other potato diseases such as late blight. Late blight, for instance, tends to manifest as water-soaked lesions without concentric rings, often accompanied by a more rapid and destructive spread under wet conditions. To manage early blight effectively, it is recommended to promptly remove any infected leaves to reduce the spread of the disease. Ensuring adequate plant spacing can improve air circulation and reduce humidity around the foliage, thereby limiting fungal growth. Additionally, maintaining balanced plant nutrition can help minimize plant stress, making the crop less susceptible to disease. These actions should be part of an integrated disease management strategy tailored to local conditions.
Predicted: Grape __Black_rot (96.32%) 	The image is most likely classified under the label 'Grape __Black_rot' with a high confidence level of 96.3%, primarily due to the presence of characteristic symptoms observed on the grapevine. Grape Black Rot is typically identified by large brown spots with distinct concentric rings on the leaves, which resemble a target pattern. This visual pattern is a critical indicator, as it is unique and aids in distinguishing Black Rot from other grape diseases. Additionally, the disease affects the berries, causing them to turn black and shrivel, often developing firm brown lesions. Another key symptom is the presence of elongated brown lesions on the canes. These features, particularly the target-like spots on the leaves and mummified berries, are not commonly seen in other grape diseases, such as Downy Mildew or Powdery Mildew, which have different symptomatology. For early management of Grape Black Rot, it is recommended to practice good sanitation by removing and destroying any infected plant material to prevent further spread. Applying a preventive fungicide may be considered if the disease pressure is high, though specific product recommendations should be obtained from a local agricultural extension service. These actions can help manage the disease effectively and protect the healthy parts of the grapevine.

As illustrated in Table 3, the generated explanations summarize discriminative visual characteristics, differentiate similar diseases, and suggest general management strategies. Beyond presenting class labels, the system contextualizes predictions with symptom descriptions and practical recommendations, thereby strengthening user trust and supporting informed agricultural decision-making. Compared with a standalone CNN, this hybrid framework better aligns with explainable artificial intelligence (XAI) principles by integrating predictive accuracy with domain-relevant interpretability.

4. CONCLUSION

This study demonstrates that integrating CNN and LLM models significantly enhances plant disease classification while improving interpretability. By integrating the visual feature learning capability of EfficientNetV2-M with the generative reasoning capacity of GPT-4o, the proposed hybrid framework achieved an accuracy of 99.79% on the PlantVillage dataset, demonstrating strong effectiveness in detecting leaf diseases across three major crop types.. Beyond producing highly accurate labels, the system generates contextual text-based explanations aligned with observed visual symptoms and agronomic knowledge. The LLM outputs also meet the defined diagnostic criteria, as they consistently present general symptoms, distinctive characteristics, and recommended initial treatment actions. This approach addresses the need for diagnostic systems that are not only precise but also transparent and practically applicable in digital agriculture, thereby supporting informed decision-making in crop management.

Future research may explore more diverse hybrid architectures and evaluate performance on additional plant species and diseases under real-world environmental conditions to assess robustness against variations in lighting, background complexity, and field noise. Developing models capable of detecting multiple disease classes within a single image is also an important direction, particularly for cases involving overlapping leaves or co-occurring infections. Furthermore, integrating visualization techniques such as Grad-CAM can enhance interpretability by highlighting image regions contributing to classification, enabling cross-validation between CNN visual attention and LLM explanations. Improvements in computational efficiency and large-

scale generalization will further strengthen the applicability of CNN and LLM hybrid systems in intelligent and explainable agricultural diagnostics.

ACKNOWLEDGEMENTS

The authors gratefully acknowledged the Department of Mathematics, Universitas Ahmad Dahlan, for institutional support and research facilities. The authors also expressed sincere appreciation to all colleagues who provided valuable feedback during the development of this study.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper. The research was conducted for academic purposes, and no financial, commercial, or personal relationships influenced the study design, data analysis, interpretation of results, or preparation of the manuscript.

DATA AVAILABILITY

The data associated with this study are publicly available online in the replication package. [<https://github.com/spMohanty/PlantVillage-Dataset/tree/master/raw/color>]

AUTHOR CONTRIBUTIONS

Frenky Riski Gilang Pratama: Conceptualization; Programming and coding implementation; Methodology; Writing-Original Draft. **Sugiarto Surono:** Conceptualization; Methodology; Supervision; Writing-Review & Editing. **Aris Thobirin:** Proofreading Paper; Writing-Review & Editing; Funding Acquisition.

REFERENCES

- [1] C. Pal, S. Karmakar, I. Mukherjee, and P. P. Chakrabarti, "A lightweight and explainable CNN model for empowering plant disease diagnosis," *Sci. Rep.*, vol. 15, no. 1, pp. 1–14, 2025, doi: 10.1038/s41598-025-94083-1.
- [2] D. Shah, V. Trivedi, V. Sheth, A. Shah, and U. Chauhan, "ResTS: Residual Deep interpretable architecture for plant disease detection," *Inf. Process. Agric.*, vol. 9, no. 2, pp. 212–223, 2022, doi: 10.1016/j.inpa.2021.06.001.
- [3] J. Lu, L. Tan, and H. Jiang, "Review on convolutional neural network (CNN) applied to plant leaf disease classification," *Agric.*, vol. 11, no. 8, pp. 1–18, 2021, doi: 10.3390/agriculture11080707.
- [4] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanekaran, "Using deep transfer learning for image-based plant disease identification," *Comput. Electron. Agric.*, vol. 173, no. November 2019, p. 105393, 2020, doi: 10.1016/j.compag.2020.105393.
- [5] I. Pacal *et al.*, "A systematic review of deep learning techniques for plant diseases," *Artif. Intell. Rev.*, vol. 57, no. 11, p. 304, 2024, doi: 10.1007/s10462-024-10944-7.
- [6] K. Hong, Y. Zhou, and H. Han, "The pipelines of deep learning-based plant image processing," *Quant. Plant Biol.*, vol. 6, p. e23, 2025, doi: 10.1017/qpb.2025.10018.
- [7] T. D. Salka, M. B. Hanafi, S. M. S. A. A. Rahman, D. B. M. Zulperi, and Z. Omar, "Plant leaf disease detection and classification using convolution neural networks model: a review," *Artif. Intell. Rev.*, vol. 58, no. 10, 2025, doi: 10.1007/s10462-025-11234-6.
- [8] J. Xiong, D. Yu, S. Liu, L. Shu, X. Wang, and Z. Liu, "A Review of Plant Phenotypic Image Recognition Technology Based on Deep Learning," pp. 1–19, 2021, doi: 10.3390/electronics10010081.
- [9] W. Shafik, A. Tufail, C. De Silva Liyanage, and R. A. A. H. M. Apong, "Using transfer learning-based plant disease classification and detection for sustainable agriculture," *BMC Plant Biol.*, vol. 24, no. 1, pp. 1–19, 2024, doi: 10.1186/s12870-024-04825-y.
- [10] J. Pan *et al.*, "ChatLeafDisease: a chain-of-thought prompting approach for crop disease classification using large language models," *Plant Phenomics*, vol. 7, no. 3, p. 100094, 2025, doi: <https://doi.org/10.1016/j.plaphe.2025.100094>.

- [11] E. Amri, Y. Gulzar, A. Yeafi, S. Jendoubi, F. Dhawi, and M. S. Mir, "Advancing automatic plant classification system in Saudi Arabia: introducing a novel dataset and ensemble deep learning approach," *Model. Earth Syst. Environ.*, vol. 10, no. 2, pp. 2693–2709, 2024, doi: 10.1007/s40808-023-01918-9.
- [12] H. K. Kondaveeti and C. G. Simhadri, "Evaluation of deep learning models using explainable AI with qualitative and quantitative analysis for rice leaf disease detection," *Sci. Rep.*, vol. 15, no. 1, p. 31850, 2025, doi: 10.1038/s41598-025-14306-3.
- [13] V. Hassija *et al.*, "Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence," *Cognit. Comput.*, vol. 16, no. 1, pp. 45–74, 2024, doi: 10.1007/s12559-023-10179-8.
- [14] Y. Wang *et al.*, "A large language model for multimodal identification of crop diseases and pests," *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-01908-0.
- [15] H. Li, H. Wu, Q. Li, and C. Zhao, "A review on enhancing agricultural intelligence with large language models," *Artif. Intell. Agric.*, vol. 15, no. 4, pp. 671–685, 2025, doi: <https://doi.org/10.1016/j.aiia.2025.05.006>.
- [16] Y. Arun and G. S. Viknesh, "Leaf Classification for Plant Recognition Using EfficientNet Architecture," *Proc. IEEE 2022 4th Int. Conf. Adv. Electron. Comput. Commun. ICAECC 2022*, vol. 10, no. 11, pp. 68–72, 2022, doi: 10.1109/ICAEC54045.2022.9716637.
- [17] J. Li *et al.*, "Foundation models in smart agriculture: Basics, opportunities, and challenges," *Comput. Electron. Agric.*, vol. 222, 2024, doi: 10.1016/j.compag.2024.109032.
- [18] M. Y. Shams, S. A. Gamel, and F. M. Talaat, "Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making," *Neural Comput. Appl.*, vol. 36, no. 11, pp. 5695–5714, 2024, doi: 10.1007/s00521-023-09391-2.
- [19] D. K. Saha, M. R. Ahmed, T. D. Nath, R. I. Boby, M. J. Hossen, and M. F. Mridha, "Fusing explainable deep learning ensembles and LLM recommendations for real-time plant leaf disease diagnosis," *Intell. Syst. with Appl.*, vol. 28, p. 200596, 2025, doi: <https://doi.org/10.1016/j.iswa.2025.200596>.
- [20] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Front. Plant Sci.*, vol. 7, no. September, pp. 1–10, 2016, doi: 10.3389/fpls.2016.01419.
- [21] S. Murugesan, J. Chinnadurai, S. Srinivasan, S. K. Mathivanan, R. R. Chandan, and U. Moorthy, "Robust multiclass classification of crop leaf diseases using hybrid deep learning and Grad-CAM interpretability," *Sci. Rep.*, vol. 15, no. 1, p. 29955, 2025, doi: 10.1038/s41598-025-14847-7.
- [22] J. H. Sinamenye, A. Chatterjee, and R. Shrestha, "Potato plant disease detection: leveraging hybrid deep learning models," *BMC Plant Biol.*, vol. 25, no. 1, p. 647, 2025, doi: 10.1186/s12870-025-06679-4.
- [23] S. Lee and S. Lee, "Efficient Data Augmentation Methods for Crop Disease Recognition in Sustainable Environmental Systems," *Big Data Cogn. Comput.*, vol. 9, no. 1, 2025, doi: 10.3390/bdcc9010008.
- [24] S. Timilsina, S. Sharma, and S. Kondo, "Advancements in maize leaf disease detection, segmentation and classification: A review," *Biosyst. Eng.*, vol. 255, p. 104162, 2025, doi: <https://doi.org/10.1016/j.biosystemseng.2025.104162>.
- [25] M. Tan and Q. Le, "EfficientNetV2: Smaller Models and Faster Training," in *Proceedings of the 38th International Conference on Machine Learning*, M. Meila and T. Zhang, Eds., in *Proceedings of Machine Learning Research*, vol. 139. PMLR, 2021, pp. 10096–10106. [Online]. Available: <https://proceedings.mlr.press/v139/tan21a.html>
- [26] S. Kumar *et al.*, "Spatial attention-guided pre-trained networks for accurate identification of crop diseases," *Sci. Rep.*, vol. 15, no. 1, pp. 1–29, 2025, doi: 10.1038/s41598-025-08004-3.
- [27] S. Fuji, B. Nugroho, and I. G. Susrama, "Implementation Of Hybrid EfficientNet V2 And Vision Transformer for Apple Leaf Diseases Classification," vol. 3, no. 1, pp. 1–11, 2025, doi: [doi: 10.32985/ijeces.16.8.5](https://doi.org/10.32985/ijeces.16.8.5).
- [28] R. S. S. Devi, V. R. V. Kumar, and P. Sivakumar, "EfficientNetV2 Model for Plant Disease Classification and Pest Recognition," 2023. doi: 10.32604/csse.2023.032231.

- [29] P. Ramachandran, B. Zoph, and Q. V Le, "Searching for Activation Functions," *ArXiv*, vol. abs/1710.0, 2018, [Online]. Available: <https://api.semanticscholar.org/CorpusID:10919244>
- [30] C.-C. Lee, H.-H. (Hank) Chau, H.-L. Wang, Y.-F. Chuang, and Y. Chau, "Mild cognitive impairment prediction based on multi-stream convolutional neural networks," *BMC Bioinformatics*, vol. 22, no. 5, p. 638, 2024, doi: 10.1186/s12859-024-05911-6.
- [31] G. Tariku, I. Ghiglieno, G. Gilioli, F. Gentilin, S. Armiraglio, and I. Serina, "Automated Identification and Classification of Plant Species in Heterogeneous Plant Areas Using Unmanned Aerial Vehicle-Collected RGB Images and Transfer Learning," *Drones*, vol. 7, no. 10, 2023, doi: 10.3390/drones7100599.
- [32] J. F. Hernández Santa Cruz, "An ensemble approach for multi-stage transfer learning models for COVID-19 detection from chest CT scans," *Intell. Med.*, vol. 5, p. 100027, 2021, doi: <https://doi.org/10.1016/j.ibmed.2021.100027>.
- [33] L. Yang, "Theoretical Analysis of Adam Optimizer in the Presence of Gradient Skewness," *Int. J. Appl. Sci.*, vol. 7, no. 2 SE-Articles, p. p27, 2024, doi: 10.30560/ijas.v7n2p27.
- [34] M. N. Razali, N. Arbaiy, P.-C. Lin, and S. Ismail, "Optimizing Multiclass Classification Using Convolutional Neural Networks with Class Weights and Early Stopping for Imbalanced Datasets," *Electronics*, vol. 14, no. 4, 2025, doi: 10.3390/electronics14040705.
- [35] W. Liu, G. Wu, H. Wang, and F. Ren, "Cross-Modal Data Fusion via Vision-Language Model for Crop Disease Recognition," *Sensors*, vol. 25, no. 13, 2025, doi: 10.3390/s25134096.
- [36] T. Savage, A. Nayak, R. Gallo, E. Rangan, and J. H. Chen, "Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine," *npj Digit. Med.*, vol. 7, no. 1, p. 20, 2024, doi: 10.1038/s41746-024-01010-1.