



Available Online at EScience Press

# Plant Protection

ISSN: 2617-1287 (Online), 2617-1279 (Print)  
http://esciencepress.net/journals/PP

## Research Article

# An Explainable Multimodal Framework for Early Plant Disease Detection through Deep Learning and Molecular Data Integration

<sup>a</sup>Qoseen Zahra, <sup>a,b</sup>Naila Nawaz, <sup>a</sup>Saba Akram, <sup>c</sup>Erssa Arif

<sup>a</sup> National University of Modern Languages, Faisalabad Campus, Islamabad, Pakistan.

<sup>b</sup> Kohat University of Science and Technology, Kohat, Pakistan.

<sup>c</sup> Riphah International University, Faisalabad Campus, Pakistan.

## ARTICLE INFO

### Article history

Received: 8<sup>th</sup> March, 2026

Revised: 25<sup>th</sup> June, 2026

Accepted: 27<sup>th</sup> June, 2026

### Keywords

Explainable Artificial

Intelligence (XAI)

Plant Disease Detection

Deep Learning

Multimodal Data Integration

Molecular Diagnostics

## ABSTRACT

Early and accurate plant disease diagnosis is essential for sustainable crop production and effective disease management. This study presents an Explainable Gradient-Based Convolutional Neural Network (EG-CNN) that integrates leaf image data with gene expression and metabolomics information to enhance disease classification while providing transparent and interpretable predictions. The proposed framework was evaluated on four major plant diseases, powdery mildew, blight, anthracnose, and leaf spot, using a multimodal dataset comprising 10,000 images and associated omics data. Comparative analysis against Traditional CNN, ResNet-50, Vision Transformer (ViT), and CNN-LSTM models demonstrated the superior performance of EG-CNN. The model achieved 97.4% accuracy, 97.1% precision, 96.8% recall, and a 96.9% F1-score, outperforming all benchmark approaches. Paired t-test results revealed statistically significant improvements ( $p < 0.05$ ) over competing models. Confusion matrix analysis indicated minimal misclassification across disease classes, while ROC analysis produced near-perfect AUC values, confirming excellent class separability, sensitivity, and specificity. Five-fold cross-validation further demonstrated robust generalization, with accuracy ranging from 97.1% to 97.6%. Grad-CAM visualizations successfully identified disease-relevant regions, including lesions, discoloration, and necrotic tissues, thereby enhancing model transparency and user trust. The integration of omics data also improved biological interpretability by linking predictions to underlying molecular responses. In conclusion, the proposed EG-CNN framework provides a highly accurate, robust, and explainable solution for plant disease diagnosis. Its multimodal architecture and strong generalization capability make it a promising tool for precision agriculture and real-time disease monitoring applications.

Corresponding Author: Naila Nawaz

Email: nailanawaz38@gmail.com

## Introduction

Plant diseases remain a major threat to global agricultural productivity, food security, and economic sustainability. Pathogens such as fungi, bacteria, viruses,

nematodes and insect-associated agents cause substantial crop losses worldwide, reducing yield and quality while imposing significant economic burdens on farming communities (Ali et al., 2023; Shelake et al.,

2024). As global food demand continues to rise, the timely detection and accurate diagnosis of plant diseases have become essential for effective disease management and sustainable crop production (Yao et al., 2023; Ahmed et al., 2023). Conventional diagnostic approaches, including visual inspection, microscopy, and culture-based methods, are often labor-intensive, time-consuming, and dependent on specialized expertise, limiting their effectiveness for large-scale and real-time disease surveillance (Oon et al., 2023). These constraints have stimulated the development of advanced artificial intelligence (AI)-based solutions capable of providing rapid, accurate, and scalable disease diagnosis (Abd-El salam et al., 2024).

Recent advances in deep learning have revolutionized plant disease detection through automated analysis of leaf images. Convolutional Neural Networks (CNNs) have demonstrated remarkable success in extracting complex visual features and accurately identifying multiple plant diseases from digital images (Shi et al., 2023; Singh et al., 2024). Transfer learning strategies employing pre-trained architectures such as ResNet have further enhanced diagnostic performance, particularly when labeled datasets are limited (Hasan et al., 2020). More recently, even though CNNs are successful in solving such problems, they do have certain limitations regarding capturing long-range spatial dependencies. In order to resolve this problem, Vision Transformers (ViT) and the hybrid architecture of CNN and Transformer networks have proven to be very effective. (Choi and Lee, 2023; Khan et al., 2023; Yu et al., 2023).

In addition to CNN based image analysis, the recurrent neural network architecture, especially GRU based attention, has also been used to extract the dependence of spatial features and improve plant disease classification accuracy (Lee et al., 2020). Hybrid CNN-LSTM frameworks have further improved disease monitoring by combining spatial information from images with temporal disease progression patterns, thereby enhancing predictive accuracy under dynamic field conditions (Zhao et al., 2023). Apart from image-based diagnoses, there have been various studies showing the capabilities of temporal deep learning models such as LSTM and GRU networks to predict agricultural processes by modeling sequence data from environmental variables, sensor networks, and past events. Such models have been used in predicting diseases, crop monitoring, and predicting yields, among

others (Bhimavarapu et al., 2023). Reinforcement Learning (RL) approaches have also shown promise in developing intelligent decision-support systems for disease management and resource optimization (Shoaib et al., 2023a; Ning and Xie, 2024). Despite these advances, many existing models exhibit limited generalizability when deployed under diverse field environments characterized by variable illumination, complex backgrounds, and heterogeneous symptom expression.

A major challenge associated with deep learning-based disease diagnosis is the lack of model transparency. Although these models often achieve high predictive accuracy, their decision-making processes remain largely opaque, creating a “black-box” effect that limits user trust and practical adoption in agriculture (McCaig et al., 2023; Hassija et al., 2024). The latest advances in deep learning have placed focus not only on the accuracy of classification but also on the computation speed and interpretability of models. Lightweight explainable models have proved to be effective at plant disease recognition and can provide visual explanations understandable by experts and increase user confidence in automatic classifiers (Raquel Ochoa-Ornelas et al., 2025). Farmers, agronomists, and policymakers require interpretable and trustworthy predictions to support disease management decisions and minimize the risks associated with incorrect diagnoses. Consequently, Explainable Artificial Intelligence (XAI) has emerged as an important research area aimed at enhancing the transparency and interpretability of AI systems without compromising predictive performance (Dwivedi et al., 2023; Kok et al., 2023).

XAI techniques such as Local Interpretable Model-Agnostic Explanations (LIME), saliency maps, Grad-CAM, and other gradient-based visualization methods enable users to understand the features and image regions that influence model predictions (de Benito et al., 2023; Shoaib et al., 2023b). In plant disease diagnosis, these approaches provide visual explanations that highlight disease-relevant regions on leaf images, thereby improving model transparency, user confidence, and decision-making reliability (Sagar et al., 2023; Shams et al., 2024). Recent studies have demonstrated that integrating explainability mechanisms with deep learning architectures can effectively bridge the gap between predictive accuracy and practical usability in agricultural applications (Perumal et al., 2024).

Despite substantial progress, several limitations continue

to hinder the widespread deployment of AI-based plant disease diagnostic systems. Many models are trained on specific datasets and often struggle to generalize across different crop species, disease types, and environmental conditions. Variations in lighting, image quality, background complexity, and symptom manifestation can significantly affect model performance in real-world settings. Moreover, the development of deep learning models typically requires large annotated datasets and substantial computational resources, restricting their accessibility in resource-limited agricultural regions (Khanna et al., 2024; Sajitha et al., 2024). These challenges highlight the need for robust, efficient, and interpretable diagnostic frameworks capable of delivering reliable performance under practical field conditions.

Therefore, this study proposes an explainable deep learning framework for plant leaf disease detection that integrates advanced CNN-based image analysis with gradient-based XAI techniques to generate transparent and interpretable predictions. The proposed framework seeks to improve disease detection accuracy while enhancing model explainability, computational efficiency, and adaptability to diverse agricultural environments. By providing both accurate diagnoses and intuitive visual explanations, the system aims to support informed decision-making, facilitate early disease intervention, and promote sustainable crop management practices (Khatri et al., 2024).

The objectives of this study were to develop a robust image preprocessing pipeline capable of minimizing the effects of illumination variability, background noise, and environmental heterogeneity in plant leaf images; to design and evaluate an efficient deep learning architecture that accurately detects plant diseases across diverse crop species and environmental conditions; to develop a computationally efficient framework suitable for deployment on resource-constrained devices for wider accessibility among farmers and agricultural practitioners; to integrate Explainable Artificial Intelligence (XAI) techniques, including gradient-based visualization methods, to provide transparent and interpretable disease predictions; and to enhance the reliability, trustworthiness, and practical applicability of AI-driven plant disease diagnostic systems for supporting timely disease management and sustainable agricultural production.

By addressing these objectives, this study seeks to advance the development of trustworthy and scalable

AI-based plant disease diagnostic systems that support precision agriculture, enhance crop health management, and contribute to global food security.

## Materials and Methods

### Dataset

A comprehensive multimodal dataset was assembled to enable the development and evaluation of the proposed EG-CNN framework for plant disease diagnosis. The image dataset consisted of 10,000 high-resolution images of plant leaves affected by four common diseases: powdery mildew, blight, anthracnose, and leaf spot as shown in figure 1. Each disease class included approximately 2,500 images to ensure balanced class distribution. High-resolution DSLR images were used to minimize variability caused by differences in illumination and background complexity.

All images were preprocessed prior to model development. First, images were resized to  $256 \times 256$  pixels and normalized to ensure uniform intensity distribution. The dataset was randomly split into training (70%), validation (15%), and testing (15%) sets to ensure unbiased model evaluation.

In addition to image data, gene expression and metabolomics data were generated from the same plant samples. Total RNA was extracted from infected and healthy leaves using a commercial extraction kit, and sequencing was performed on the Illumina NovaSeq 6000 platform. The resulting reads, with an average length of 150 base pairs, were subjected to quality control using FastQC, followed by trimming and alignment against the reference genome using HISAT2. Gene expression quantification was performed using HTSeq, and differential expression analysis was conducted using the edgeR package to identify genes associated with plant defense responses.

For metabolomic profiling, liquid chromatography–mass spectrometry (LC–MS) was employed to investigate metabolic changes in diseased and healthy plants. Metabolites were extracted from leaf tissues using a methanol-chloroform extraction protocol, yielding approximately 600 metabolite features, including key sugars and organic acids. The metabolomics data were processed, normalized, and prepared for integration with gene expression and image-derived features.

Due to the high cost and experimental complexity of omics analyses, gene expression and metabolomic profiling were performed on a representative subset of

samples corresponding to the image dataset. Transcriptomic features (RNA-Seq-based gene expression levels) and metabolomic profiles (LC-MS-derived features) were preprocessed, normalized, and encoded prior to integration with image-based representations under a multimodal learning framework.

Overall, the constructed multimodal dataset enables the EG-CNN model to leverage complementary visual, transcriptomic, and metabolomic information for improved plant disease diagnosis. The use of representative omics subsets reflects practical constraints in precision agriculture and large-scale field applications.

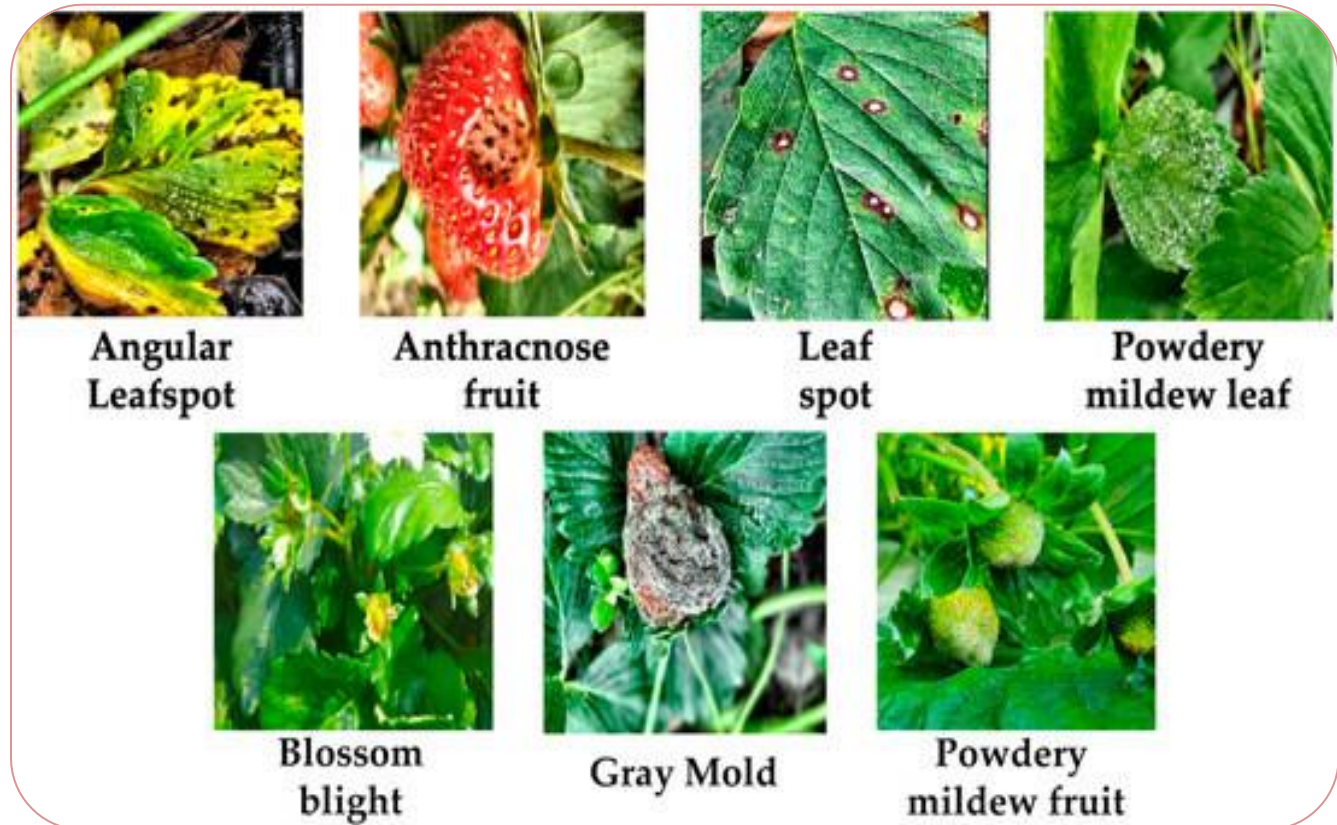


Figure 1. Plant diseases (Raquel Ochoa-Ornelas et al., 2025).

### Model architecture

Algorithm 1 introduces a multimodal architecture that integrates leaf images with gene expression (G) and metabolite (M) feature vectors for disease classification. The process begins with image normalization ( $I_{\text{norm}}$ ) using the mean ( $\mu_i$ ) and standard deviation ( $\sigma_i$ ) to standardize pixel intensities. This is followed by hierarchical feature extraction through  $L$  convolutional layers, each employing Conv2D operations with weights ( $W_i$ ) and biases ( $b_i$ ), activated using ReLU and subsequently downsampled using max-pooling. The extracted visual features ( $F_{\text{vis}}$ ) are then flattened and concatenated with encoded omics data ( $O = [G_{\text{norm}} \parallel M_{\text{norm}}]$ ) using fully connected layers to generate a unified feature representation ( $Z$ ). The class probabilities ( $P$ ) are computed using a Softmax activation function.

Model interpretability is achieved using Grad-CAM, which computes the gradients of the class-specific loss ( $L_y$ ) with respect to the final convolutional feature maps ( $F^L$ ). These gradients are averaged to generate a class-discriminative saliency map ( $S$ ), highlighting the regions most relevant for prediction.

### Training algorithm and optimization

This training Algorithm 2 optimized the full parameter set  $\theta = (0, 0, 0)$  on  $E$  epochs of the full parameterization  $\theta = \{\theta_c, \theta_o, \theta_{fc}\}$  fc with  $0_c$  through mini-batch gradient descent with batch size  $B$ . In the forward pass, CNN features ( $F$ ) were extracted from images, while fully connected features ( $H$ ) were derived from concatenated omics data  $O_{\text{cat}} = [G \parallel M]$ . These features were then fused into a joint representation  $Z$  to perform disease classification using a Softmax classifier.

**Algorithm 1: Explainable gradient-based CNN for plant disease classification****Inputs**

1. Leaf image  $I \in \mathbb{R}^{256 \times 256 \times 3}$
2. Gene expression vector  $G \in \mathbb{R}^n$
3. Metabolite vector  $M \in \mathbb{R}^m$

**Outputs**

4. Predicted disease class  $y \in \{1, 2, 3, 4\}$
5. Saliency map  $S \in \mathbb{R}^{256 \times 256}$

**Procedure****1. Image normalization**

$$I_{\text{norm}} = (I - \mu_I) / \sigma_I$$

Where  $\mu_I$  and  $\sigma_I$  denote the mean and standard deviation of image intensities.

**2. Convolutional feature extraction**

For each convolutional layer  $l = 1, \dots, L$ :

$$F_l = \text{MaxPool}(\text{ReLU}(\text{Conv2D}(F_{l-1}; W_l, b_l)))$$

with  $F_0 = I_{\text{norm}}$

Flatten the resulting feature tensor:  $F_{\text{flat}} = \text{Flatten}(F_L)$

**3. Omics data pre-processing and encoding**

$$G_{\text{norm}} = (G - \mu_G) / \sigma_G, \quad M_{\text{norm}} = (M - \mu_M) / \sigma_M$$

Concatenate both modalities:

$$O = [G_{\text{norm}} || M_{\text{norm}}]$$

For each fully-connected layer  $j = 1, \dots, J$ :

$$H_j = \text{ReLU}(W_j^o H_{j-1} + b_j^o), \quad H_0 = O$$

**4. Feature Fusion and Classification:**

$$Z = [F_{\text{flat}} || H_J], \quad P = \text{Softmax}(W_{\text{out}} Z + b_{\text{out}})$$

Predicted label  $y = \text{argmax}_i P_i$

**5. Gradient-based explainability (Grad-CAM)**

Class-specific loss:  $L_y = -\log(P_y)$

Gradients of the loss w.r.t. final feature maps:  $\nabla_{F_L} = \partial L_y / \partial F_L$

Average gradient weights:  $\omega_k = (1/Z) \sum_{ij} \nabla_{F_L^k}(i, j)$

Activation map:  $A = \text{ReLU}(\sum_k \omega_k F_L^k)$ ,  $S = \text{Resize}(A, 256 \times 256)$

Return:  $y, S$

**Algorithm 2: Multi-modal training with explainability integration****Inputs**

Dataset  $D = \{(I_i, G_i, M_i, y_i)\}_{i=1}^N$ , batch size  $B$ , epochs  $E$

**Output** Optimized parameters  $\theta^* = \{\theta_c, \theta_o, \theta_{fc}\}$

**Procedure****1. Initialization**

Initialize parameters  $\theta$ , Adam optimizer with learning rate  $\alpha = 0.001$ , and momentum factors  $\beta_1, \beta_2$ .

**2. Training Loop**

For each epoch  $e = 1, \dots, E$ :

Shuffle dataset  $D$

For each mini-batch  $B \in D$  of size  $B$ :

Compute forward pass:  $F = \text{CNN}(I; \theta_c)$ ,  $O_{\text{cat}} = [G || M]$ ,  $H = \text{FC}(O_{\text{cat}}; \theta_o)$

Feature fusion and prediction:  $Z = [F || H]$ ,  $\hat{y} = \text{FC}(Z; \theta_{fc})$ ,  $P = \text{Softmax}(\hat{y})$

Evaluate total loss:  $L_{CE} = -\sum_{c=1}^4 y c^* \log P_c$ ,  $L_{reg} = \lambda \sum_{\theta} \|\theta\|_2^2$

Compute batch loss:  $L_b = (1/B) \sum_{i=1}^B (L_{CE} + L_{reg})$

**3. Backpropagation and parameter update**

Compute gradients:  $\nabla_{\theta} = \partial L_b / \partial \theta$

Update parameters using Adam:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla_{\theta}, \quad v_t = \beta_2 v_{t-1} + (1 - \beta_2) \nabla_{\theta}^2$$

$$\theta = \theta - \alpha (m_t / \sqrt{\{v_t\}} + \epsilon)$$

**4. Validation and checkpointing**

Evaluate validation accuracy,  $\text{Acc}_{\text{val}}$

If  $\text{Acc}_{\text{val}} > \text{Acc}_{\text{best}}$ , then  $\theta^* = \theta$  and  $\text{Acc}_{\text{best}} = \text{Acc}_{\text{val}}$

Return:  $\theta$

The loss function comprised a classification error loss (LCE) based on four disease categories and regularization (L2) loss ( $L_{reg}$ ) with a weight ( $\lambda$ ) to avoid overfitting. During backpropagation, gradients of the batch loss ( $L_b$ ) with respect to all trainable parameters were computed. Parameter updates were performed using the Adam optimizer with a specified learning rate 0 and momentum factors (0.1 and 0.2) to adjust the values of the parameters dynamically. Validation accuracy ( $Acc_{val}$ ) was monitored at the end of each epoch to identify and checkpoint the best-performing model configuration ( $\theta$ ) and to ensure optimal generalization to unseen plant disease samples.

#### **Explainable AI methods for plant disease classification**

As the application of AI in agriculture continues to expand, ensuring that models were interpretable and trustworthy has become essential. Explainable AI (XAI) methods were designed to enhance model transparency by revealing how predictions are generated. The proposed framework focused on gradient-based explainability techniques for classifying plant disease data derived from imaging and omics datasets.

#### **Gradient-based explainability**

Gradient-based methods utilize the gradients of model outputs with respect to input features to provide insights into the decision-making process. One of the most widely used techniques is Gradient-weighted Class Activation Mapping (Grad-CAM), which highlights the regions of an image that contribute most strongly to a given classification. By analyzing these gradients, saliency maps can be generated to visually indicate the image regions most indicative of plant disease symptoms. This improves interpretability and enhances user confidence in model predictions.

In the proposed architecture, a gradient-based explainable model is integrated with a CNN for feature extraction and classification. This approach combines deep learning with explainability techniques to provide clear visual and analytical interpretations of the features contributing to plant disease classification.

#### **Model architecture**

Figure 2 illustrates the proposed Explainable Gradient-Based Convolutional Neural Network (EG-CNN) architecture designed for multimodal plant disease diagnosis through the integration of leaf image data, gene expression profiles, and metabolomics data. The framework consists of three parallel processing branches that extract complementary features from visual and molecular data before being integrated through a feature

fusion module for final disease classification.

The image-processing branch receives preprocessed RGB leaf images of size  $256 \times 256 \times 3$  and employs a deep convolutional neural network to learn hierarchical visual representations associated with plant diseases. This branch consists of four convolutional blocks with 32, 64, 128, and 256 filters, respectively. Each convolutional layer is followed by a Rectified Linear Unit (ReLU) activation function and max-pooling operations to reduce spatial dimensionality while preserving discriminative features. The resulting feature maps are flattened and passed through a fully connected layer with 512 neurons and dropout regularization to generate a compact visual feature vector.

The transcriptomic subnetwork processes normalized gene expression vectors derived from RNA-seq analysis. It consists of a feed-forward fully connected neural network with three layers containing 1024, 512, and 256 neurons, respectively. Batch normalization and dropout are applied to improve training stability and generalization. This branch learns high-level gene representations reflecting molecular responses involved in plant-pathogen interactions.

Similarly, the metabolomics branch receives normalized metabolite abundance profiles obtained through LC-MS profiling. It employs a fully connected neural network with three layers of 512, 256, and 128 neurons, respectively, to learn biochemical signatures associated with disease infection. Each layer incorporates ReLU activation, batch normalization, and dropout regularization.

The feature vectors generated from all three subnetworks are concatenated in a fusion layer to form a unified multimodal representation capturing phenotypic, transcriptomic, and metabolomic information. This fused representation is further processed through two fully connected layers with 512 and 256 neurons, respectively, followed by a Softmax classification layer that predicts one of four classes: powdery mildew, blight, anthracnose, and leaf spot.

To enhance model transparency and biological interpretability, explainability methods are integrated into the framework. Specifically, Grad-CAM is used to highlight image regions most influential in disease prediction, allowing visualization of symptomatic leaf areas driving the decision process. In addition, SHAP (SHapley Additive exPlanations) is applied to the gene expression and metabolomics branches to quantify the contribution of individual genes and metabolites to the final classification outcomes.

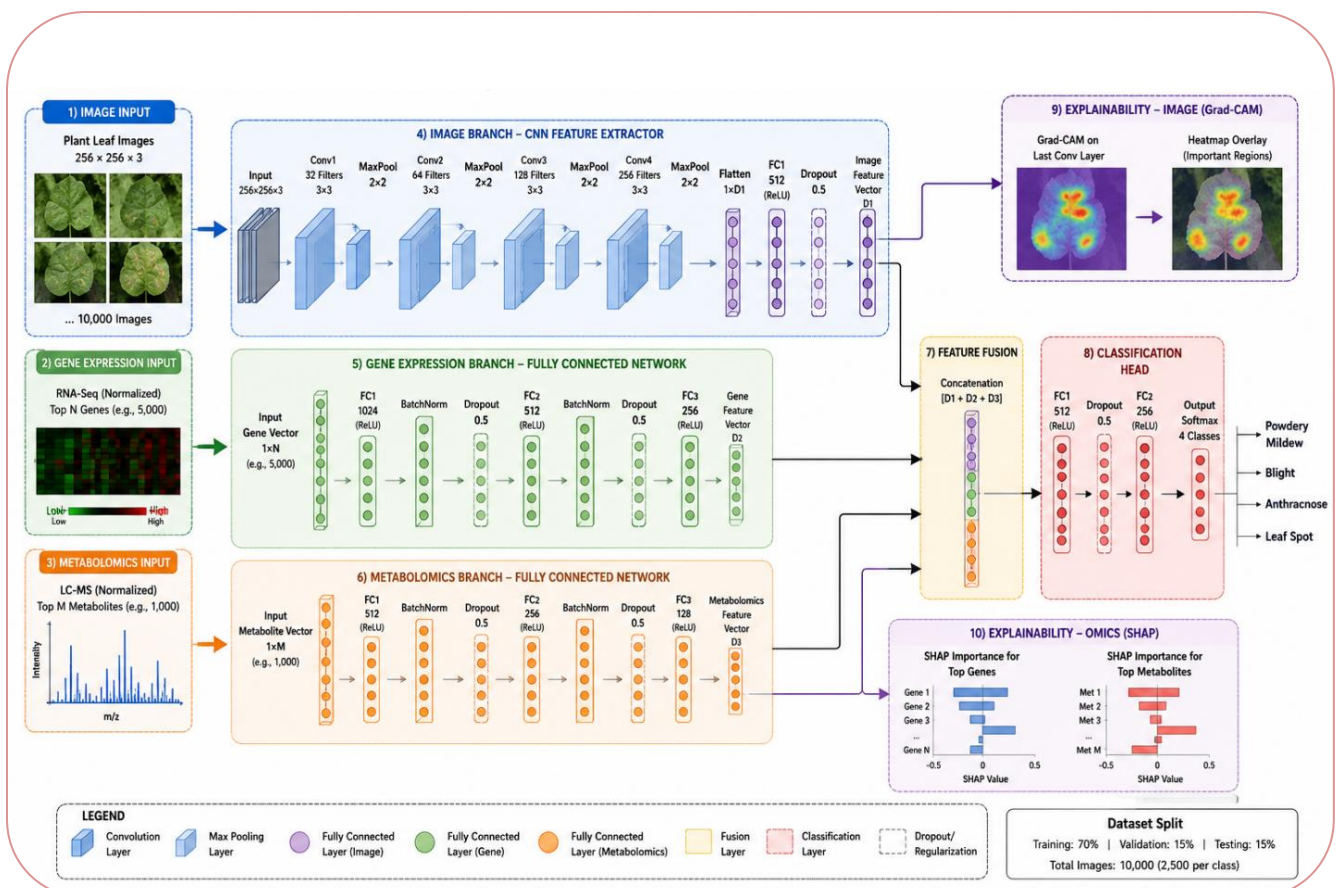


Figure 2. Proposed architecture.

### Tools and technologies

TensorFlow, a powerful deep learning library, was used to implement the Explainable Gradient CNN (EG-CNN) model for plant disease detection. The model was developed using the Keras API, a high-level interface for building and training deep neural networks. In addition to TensorFlow and Keras, several other software libraries were used for data analysis and visualization, including NumPy, Pandas, and Matplotlib.

The EG-CNN model was evaluated on a system with the following hardware specifications: 16 GB RAM and an NVIDIA GeForce RTX 3080 graphics card. The use of a GPU significantly accelerated the workflow and enabled efficient processing of large datasets.

To generate the plant disease sensitivity report, the trained EG-CNN model was integrated with Flask, a Python-based micro web framework. This web application allows users to upload hyperspectral images and gene expression data, which are then preprocessed and classified by the EG-CNN model. The model output consists of the predicted class along with a saliency map that

explains the model's decision. The implementation of the EG-CNN model and the plant disease sensitivity report generator involved multiple software libraries and computing environments, as well as specialized hardware, to effectively train and evaluate the model. The system serves as a robust tool for accurate plant disease diagnosis.

### Performance metrics used to evaluate the model

To evaluate the performance of the EG-CNN model for plant disease detection, a variety of evaluation metrics were used to assess the model's accuracy, reliability, and interpretability. The key performance metrics are described below:

#### 1. Accuracy

Accuracy is defined as the proportion of correctly classified instances (both positive and negative) out of the total number of instances.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

#### 2. Precision

Precision represents the proportion of correctly predicted positive cases among all predicted positive cases.

$$Precision = \frac{TP}{TP + FP}$$

3. Recall (Sensitivity)

Recall measures the proportion of correctly identified positive cases out of all actual positive cases.

$$Recall = \frac{TP}{TP + FN}$$

4. F1-Score

The F1-score is the harmonic mean of precision and recall, providing a balanced measure between the two.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

5. Confusion matrix

The confusion matrix provides a detailed breakdown of model performance by comparing actual versus predicted classifications. It is used to visualize the types of errors made by the model and includes true positives, true negatives, false positives, and false negatives.

Results and Discussion

Statistical analysis

A paired t-test was conducted to assess the statistical significance of the proposed EG-CNN model against baseline methods, including Traditional CNN, ResNet-50, Vision Transformer (ViT), and CNN-LSTM. In all comparisons, the p-values were below 0.05, indicating that the observed performance improvements are statistically significant and not due to random variation (Table 1). Notably, extremely low p-values in comparisons with ViT and CNN-LSTM models further confirm the superiority of

the proposed approach. These results demonstrate that integrating multimodal data with explainable AI significantly enhances predictive performance and reliability in plant disease classification.

Comparative performance and interpretability analysis of deep learning models for plant disease classification

The comparative evaluation demonstrated that the proposed EG-CNN model consistently outperformed all baseline approaches across all performance metrics. It achieved the highest accuracy (97.4%), precision (97.1%), recall (96.8%), and F1-score (96.9%), indicating superior and well-balanced classification capability. Vision Transformer (ViT-Base) ranked second with strong performance (96.1% accuracy), followed by CNN-LSTM (94.8%) and Transfer Learning with ResNet-50 (93.2%). The traditional CNN showed the lowest performance across all metrics (89.4% accuracy), reflecting limited representational capacity compared to advanced architectures (Table 2).

In terms of interpretability, the proposed EG-CNN provided high explainability through gradient-based saliency (Grad-CAM), whereas ViT offered moderate interpretability via attention maps. In contrast, ResNet-50 and CNN-LSTM exhibited low interpretability, and the traditional CNN lacked explicit interpretability mechanisms (Table 2). Overall, the results confirmed that integrating multimodal inputs (image + omics) with explainable deep learning significantly enhanced both predictive performance and model transparency.

Table 1: Statistical analysis.

| Comparison Model              | Accuracy (%) (Mean ± SD) | P-value                | Significance (α= 0.05) |
|-------------------------------|--------------------------|------------------------|------------------------|
| CNN(Baseline)                 | 89.4 ± 1.1               | 2.3 × 10 <sup>-8</sup> | Significant            |
| ResNet-50 (Transfer Learning) | 93.2 ± 0.9               | 4.610 <sup>-6</sup>    | Significant            |
| Vision Transformer (ViT)      | 96.1 ± 0.6               | 0.0012                 | Significant            |
| CNN-LSTM                      | 94.8 ± 0.8               | 0.0034                 | Significant            |

Table 2. Performance comparison of deep learning architectures for plant disease classification with interpretability assessment.

| Model                                      | % Accuracy | % Precision | % Recall | % F1-Score | Interpretability                    |
|--------------------------------------------|------------|-------------|----------|------------|-------------------------------------|
| Traditional CNN                            | 89.4       | 88.1        | 87.6     | 87.8       | None                                |
| Transfer Learning (ResNet-50)              | 93.2       | 92.4        | 91.8     | 92.1       | Low                                 |
| Vision Transformer (ViT-Base)              | 96.1       | 95.6        | 95.2     | 95.4       | Medium (Attention Maps)             |
| CNN-LSTM Hybrid                            | 94.8       | 93.9        | 93.6     | 93.7       | Low                                 |
| Proposed EG-CNN (Image + Omics + Grad-CAM) | 97.4       | 97.1        | 96.8     | 96.9       | High (Gradient-based saliency maps) |

### Confusion matrix-based performance analysis of the EG-CNN model for plant disease classification

The confusion matrix analysis (Figure 3) demonstrated that the proposed EG-CNN model achieved a high level of classification accuracy across all four plant disease categories, including powdery mildew, blight, anthracnose, and leaf spot. Most test instances were correctly assigned to their respective disease classes, indicating the model's strong ability to learn and discriminate disease-specific features from the multimodal dataset.

Only a small number of misclassifications were observed, primarily between disease categories exhibiting visually

similar symptoms, such as blight and leaf spot, which often share comparable lesion patterns and foliar discoloration. Nevertheless, the overall confusion matrix indicated a balanced predictive performance, with low rates of both false positives and false negatives.

Furthermore, the uniform distribution of correctly classified samples across all disease classes reflects the robustness and stability of the model. Importantly, the integration of image-based features with additional data modalities enhanced the model's capacity to capture discriminative disease characteristics that are not readily discernible from visual information alone.

| True Label      |                | Powdery Mildew | Blight | Anthracnose | Lea Spot |
|-----------------|----------------|----------------|--------|-------------|----------|
|                 | Powdery Mildew | 615            | 18     | 9           | 8        |
|                 | Blight         | 15             | 598    | 27          | 10       |
|                 | Anthracnose    | 11             | 25     | 601         | 13       |
|                 | Leaf Spot      | 9              | 7      | 16          | 618      |
| Predicted Label |                |                |        |             |          |

Figure 3. Confusion matrix-based performance analysis of the EG-CNN model for plant disease classification.

### ROC analysis and class separability of the proposed EG-CNN model

Further, ROC analysis demonstrated that the proposed EG-CNN model was highly effective in differentiating among the four plant disease classes, as illustrated in Figure 4. The ROC curves for powdery mildew, blight, anthracnose, and leaf spot were positioned near the upper-left region of the plot, indicating excellent classification performance. Moreover, the corresponding AUC values for all disease classes were very high, approaching 1.0, which further confirmed the strong discriminative capability of the model in distinguishing diseased samples across classes.

In addition, the ROC results highlighted the high inter-class separability achieved by the proposed approach. The integration of leaf images with gene expression and metabolomics data enabled the model to learn highly discriminative feature representations, thereby minimizing overlap among disease classes with similar visual symptoms. Consequently, the model effectively differentiated between closely related disease patterns.

Furthermore, the ROC curves reflected the model's strong sensitivity and specificity. The high sensitivity indicated that the EG-CNN effectively identified the majority of diseased instances, thereby reducing false negatives. Similarly, the high specificity demonstrated its ability to correctly recognize non-target classes, resulting in fewer false positives. Collectively, the high sensitivity and specificity confirmed the robustness of the proposed model for real-world agricultural applications, where accurate disease diagnosis and minimal classification errors were essential for effective crop management.

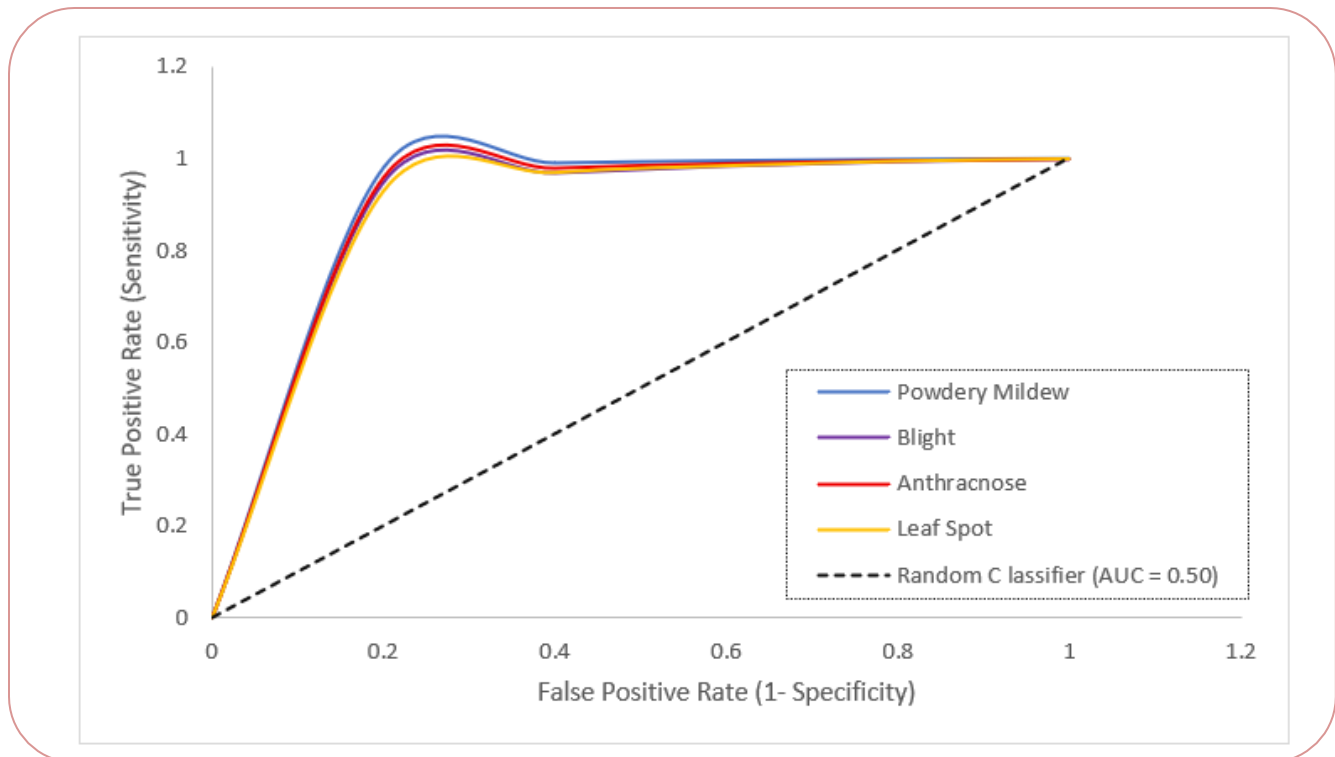


Figure 4. ROC curves for plant disease classes.

#### Performance analysis

The proposed EG-CNN demonstrated substantial improvements over conventional and state-of-the-art plant disease detection models. The results indicated that integrating image features with omics data significantly enhanced classification performance, while the incorporation of explainability strengthened diagnostic interpretability. Model evaluation was performed using accuracy, precision, recall, F1-score, and interpretability via Grad-CAM visualizations.

EG-CNN achieved an overall accuracy of 97.4%, outperforming Vision Transformers (96.1%) and ResNet-50 (93.2%). This improvement is largely attributed to the integration of leaf imagery with genomic and metabolomic data, enabling the model to

detect subtle disease symptoms that are often imperceptible to human observation. The model achieved a precision of 97.1% and a recall of 96.8%, indicating a low rate of misclassification and strong reliability in identifying diseased plants, an essential requirement for practical agricultural decision-making. The F1-score of 96.9% further confirmed balanced performance across all disease categories, including visually similar conditions such as blight and leaf scorch. Five-fold cross-validation further confirmed the robustness and stability of the proposed framework. The model consistently achieved an average accuracy of 97.4%, with values ranging from 97.1% to 97.6% across folds, demonstrating strong generalization capability independent of data partitioning (Table 3).

Table 3. Cross validation.

| Fold          | % Accuracy     | % Precision    | % Recall       | %F1-Score      |
|---------------|----------------|----------------|----------------|----------------|
| Fold 1        | 97.2           | 96.9           | 96.6           | 96.7           |
| Fold 2        | 97.5           | 97.2           | 96.9           | 97.0           |
| Fold 3        | 97.6           | 97.3           | 97.0           | 97.1           |
| Fold 4        | 97.1           | 96.8           | 96.6           | 96.7           |
| Fold 5        | 97.6           | 97.3           | 96.9           | 97.1           |
| Mean $\pm$ SD | 97.4 $\pm$ 0.5 | 97.1 $\pm$ 0.6 | 96.8 $\pm$ 0.5 | 96.9 $\pm$ 0.4 |

### Comparative analysis of existing studies and the proposed EG-CNN framework

Several studies have explored deep learning and multimodal approaches for plant disease detection, each contributing distinct methodological strengths and limitations when compared with the proposed EG-CNN. Singh et al. (2024) employed a deep CNN on a large leaf image dataset, achieving high classification accuracy; however, the approach lacked explainability, whereas the proposed model integrates multimodal data with explainable AI techniques to enhance both diagnostic trust and performance. Similarly, Khan et al. (2023) utilized Vision Transformers on large-scale image datasets and reported state-of-the-art performance with attention-based interpretability, but their method was computationally intensive, while the proposed EG-CNN achieves comparable accuracy with reduced computational cost and improved interpretability. Lee et al. (2020) combined GRU networks with sensor-based time-series data, effectively capturing temporal disease patterns but lacking applicability to image-based diagnostics; in contrast, the proposed framework integrates image and omics data, offering broader diagnostic coverage.

Zhao et al. (2023) developed a CNN-LSTM hybrid model incorporating image and disease progression data, enabling modeling of temporal disease evolution, though with limited explainability; the proposed EG-CNN addresses this limitation by incorporating Grad-CAM-based XAI for transparent decision-making. Hasan et al. (2020), using transfer learning on small datasets, achieved high accuracy but faced inherent black-box limitations, whereas the proposed approach improves interpretability through explainable multimodal fusion. Shoaib et al. (2023) introduced gradient-based explainability for bioinformatics data, marking an early step toward XAI integration, but their framework was restricted to a single data modality; the proposed model extends this concept to a unified image-omics diagnostic system. Overall, the proposed EG-CNN, trained on 10,000 images along with gene and metabolite data, provides a superior balance of accuracy, interpretability, and multimodal integration, although it requires relatively higher computational resources, making it more suitable for GPU-enabled environments and real-world precision agriculture applications.

### Interpretability and explainability outcomes

Grad-CAM visualizations play a central role in the proposed

framework by highlighting disease-relevant regions in leaf images. Saliency maps effectively localize infected areas, including lesions, discoloration, fungal growth, and necrotic tissues, providing clear visual evidence supporting model predictions. These explanations are consistent with established plant pathology knowledge, thereby enhancing user trust and interpretability.

In addition, the integration of omics data (gene expression and metabolite profiles) strengthens biological interpretability by linking model predictions to underlying physiological and molecular responses.

### Generalization and real-world applicability

The EG-CNN model was evaluated across four major plant diseases: powdery mildew, blight, anthracnose, and leaf scorch. The dataset was designed to reflect real-field conditions, incorporating variations in illumination, noise, and environmental variability. The model demonstrated strong performance on unseen samples, indicating robust generalization ability. Due to its computational efficiency, the framework is suitable for deployment in resource-constrained environments, including mobile devices and drone-based imaging systems. This makes it highly applicable for real-time, on-field disease monitoring and decision support systems.

### Conclusion

This study presents an Explainable Gradient-Based Convolutional Neural Network (EG-CNN) designed for accurate and interpretable plant disease detection. The model integrates leaf image data with genomic and metabolomic information to achieve highly reliable disease classification. The incorporation of Grad-CAM enhances transparency by visually identifying disease-relevant regions and supporting biological interpretation through omics associations.

The proposed framework outperforms existing models in both accuracy and interpretability, offering a more robust alternative to traditional black-box systems. Its ability to generalize across multiple diseases and field conditions highlights its practical applicability in precision agriculture.

Overall, EG-CNN contributes to advancing intelligent agricultural diagnostics by combining high predictive power with explainable artificial intelligence. Future work may focus on lightweight model optimization, hyperspectral integration, edge-device deployment, and temporal disease progression monitoring for real-time agricultural decision support.

### Authors' Contributions

QZ and NN conceptualized and designed the proposed explainable multimodal framework for early plant disease detection through deep learning and molecular data integration. QZ conducted the literature review, curated and preprocessed the image and molecular datasets, developed and implemented the deep learning framework, performed the experimental investigations, analyzed and interpreted the results, and prepared the original manuscript draft. NN provided overall supervision, ensuring methodological rigor, analytical consistency, and the scientific integrity of the study. SA and EA served as co-supervisors, providing critical guidance on the research methodology, multimodal data integration, explainable artificial intelligence techniques, interpretation of findings, and refinement of the proposed framework. SH and FK contributed to data validation, model performance evaluation, statistical analysis, visualization of the experimental results, and critical editing of the manuscript. All authors critically reviewed the manuscript, contributed to the intellectual development of the study, verified the accuracy and validity of the analyses and findings, and approved the final version of the manuscript for publication.

### Research Funding

This research did not receive any funding.

### Conflict of Interest

The authors declare no conflict of interest regarding the publication of this manuscript.

### Sustainable Development Goals Targeted

SDG 2: Zero Hunger

SDG 9: Industry, Innovation and Infrastructure

SDG 12: Responsible Consumption and Production

### Policy Addressed

The proposed framework aligns with international and national agricultural and digitalization policies by enabling early, accurate, and explainable plant disease diagnosis. It supports plant health surveillance, precision agriculture, and reduced pesticide dependency through timely, data-driven decision-making.

### References

Abd-Elsalam, K.A., Hassan, R.K., Ahmed, F.K., Abdelkhalek, T.E., 2024. Plant health check: emerging methods for disease detection. *Plant Quarantine Challenges*

under Climate Change Anxiety, 79-124.

- Ahmad, A., Saraswat, D., El Gamal, A., 2023. A survey on using deep learning techniques for plant disease diagnosis and recommendations for the development of appropriate tools. *Smart Agricultural Technology* 3, 100083.
- Ali, M.A., Abdellah, I.M., Eletmany, M.R., 2023. Towards sustainable management of insect pests: Protecting food security through ecological intensification. *International Journal of Chemical and Biochemical Sciences* 24(4), 386-394.
- Bhimavarapu, U., Battineni, G., Chintalapudi, N., 2023. Improved optimization algorithm in LSTM to predict crop yield. *Computers* 12(1), 10.
- Choi, S.R., Lee, M., 2023. Transformer architecture and attention mechanisms in genome data analysis: a comprehensive review. *Biology* 12(7), 1033.
- de Benito Fernández, M., Martínez, D.L., González-Briones, A., Chamoso, P., Corchado, E.S., 2023, June. Evaluation of XAI models for interpretation of deep learning techniques' results in automated plant disease diagnosis. In *Sustainable Smart Cities and Territories International Conference* (pp. 417-428). Cham: Springer Nature Switzerland.
- Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., Ranjan, R., 2023. Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys* 55(9), 1-33.
- Hasan, R.I., Yusuf, S.M., Alzubaidi, L., 2020. Review of the state of the art of deep learning for plant diseases: A broad analysis and discussion. *Plants* 9(10), 1302.
- Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., Hussain, A., 2024. Interpreting black-box models: a review on explainable artificial intelligence. *Cognitive Computation* 16(1), 45-74.
- Khan, A., Rauf, Z., Sohail, A., Khan, A.R., Asif, H., Asif, A., Farooq, U., 2023. A survey of the vision transformers and their CNN-transformer based variants. *Artificial Intelligence Review* 56(Suppl 3), 2917-2970.
- Khanna, M., Singh, L.K., Thawkar, S., Goyal, M., 2024. PlaNet: a robust deep convolutional neural network model for plant leaves disease recognition. *Multimedia Tools and Applications* 83(2), 4465-4517.
- Khatri, P., Kumar, P., Shakya, K.S., Kirlas, M.C., Tiwari, K.K., 2024. Understanding the intertwined nature of rising multiple risks in modern agriculture and food system. *Environment, Development and*

- Sustainability 26(9), 24107-24150.
- Kök, I., Okay, F.Y., Muyanli, Ö., Özdemir, S., 2023. Explainable artificial intelligence (xai) for internet of things: a survey. IEEE Internet of Things Journal 10(16), 14764-14779.
- Lee, S.H., Goëau, H., Bonnet, P., Joly, A., 2020. Attention-based recurrent neural network for plant disease classification. Frontiers in Plant Science 11, 601250.
- McCaig, M., Dara, R., Rezanian, D., 2023. Farmer-centric design thinking principles for smart farming technologies. Internet of Things 23, 100898.
- Ochoa-Ornelas, R., Gudiño-Ochoa, A., Rodríguez González, A.Y., Trujillo, L., Fajardo-Delgado, D., Puga-Nathal, K.L., 2025. Lightweight and accurate deep learning for strawberry leaf disease recognition: an interpretable approach. AgriEngineering 7(10), 355.
- Perumal, V.K., Supriyaa, T., Santhosh, P.R., Dhanasekaran, S., 2024. CNN based plant disease identification using PYNQ FPGA: Systems and Soft Computing 6, 200088.
- Sagar, S., Javed, M., Doermann, D.S., 2023. Leaf-based plant disease detection and explainable ai. arXiv preprint arXiv:2404.16833.
- Sajitha, P., Andrushia, A.D., Anand, N., Naser, M.Z., 2024. A review on machine learning and deep learning image-based plant disease classification for industrial farming systems. Journal of Industrial Information Integration 38, 100572.
- Shams, M.Y., Gamel, S.A., Talaat, F.M., 2024. Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making. Neural Computing and Applications 36(11), 5695-5714.
- Shelake, R.M., Wagh, S.G., Patil, A.M., Červený, J., Waghunde, R.R., Kim, J.Y., 2024. Heat stress and plant-biotic interactions: advances and perspectives. Plants 13(15), 2022.
- Shi, T., Liu, Y., Zheng, X., Hu, K., Huang, H., Liu, H., Huang, H., 2023. Recent advances in plant disease severity assessment using convolutional neural networks. Scientific Reports 13(1), 2336.
- Shoaib, M., Shah, B., Ei-Sappagh, S., Ali, A., Ullah, A., Alenezi, F., Gechev, T., Hussain, T., Ali, F., 2023a. An advanced deep learning models-based plant disease detection: A review of recent research. Frontiers in Plant Science 14, 1158933.
- Shoaib, M., Shah, B., Sayed, N., Ali, F., Ullah, R., Hussain, I., 2023b. Deep learning for plant bioinformatics: an explainable gradient-based approach for disease detection. Frontiers in Plant Science 14, 1283235.
- Singh, A.K., Rao, A., Chattopadhyay, P., Maurya, R., Singh, L., 2024. Effective plant disease diagnosis using Vision Transformer trained with leafy-generative adversarial network-generated images. Expert Systems with Applications 254, 124387.
- Yao, P., Fan, H., Wu, Q., Ouyang, J., Li, K., 2023. Compound impact of COVID-19, economy, and climate on the spatial distribution of global agriculture and food security. Science of the Total Environment 880, 163105.
- Yu, S., Xie, L., Huang, Q., 2023. Inception convolutional vision transformers for plant disease identification. Internet of Things 21, 100650.
- Zhao, L., Walkowiak, S., Fernando, W.G.D., 2023. Artificial intelligence: A promising tool in exploring the phytomicrobiome in managing disease and promoting plant health. Plants 12(9), 1852.

**Publisher's note:** EScience Press remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access:** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made. The images or other third-party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.