JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Plant Disease Detection Using Deep Learning with Explainable Artificial Intelligence

Geeta devi
Research scholar
bhagatgeeta17@gmail.com

Harjinder Kaur
Assistant Professor
harrysaini988@gmail.com

Abstract

Plant diseases are a major challenge in modern agriculture because they can reduce crop productivity, quality, and economic returns. Traditional plant disease identification generally depends on visual inspection by farmers or agricultural experts, which can be time-consuming, subjective, and difficult to apply over large agricultural areas. This study proposes a deep learning-based plant disease detection framework integrated with Explainable Artificial Intelligence (XAI). The proposed system analyses plant leaf images and automatically classifies them into healthy and disease categories. A Convolutional Neural Network (CNN) or transfer-learning-based architecture can be employed to learn discriminative visual characteristics such as colour variation, lesions, spots, texture, and changes in leaf structure. Image preprocessing, resizing, normalization, data augmentation, model training, and performance evaluation are incorporated into the proposed methodology. To improve transparency, Grad-CAM is applied to generate visual explanations showing the regions of the leaf that contribute most strongly to the model's prediction. Unlike a conventional black-box classifier, the proposed approach provides both a disease prediction and an interpretable visual explanation. The framework is evaluated using accuracy, precision, recall, F1-score, and confusion matrix analysis. The integration of deep learning and XAI can support more transparent and trustworthy computer-vision-based agricultural disease diagnosis and can provide a foundation for future mobile and smart-farming applications.

Keywords: Plant Disease Detection, Deep Learning, Convolutional Neural Network, Transfer Learning, Explainable AI, Grad-CAM, Image Classification, Precision Agriculture, Computer Vision.

1. Introduction

Agriculture plays an important role in food production, rural employment, and economic development. However, plant diseases can significantly affect agricultural productivity. Diseases caused by fungi, bacteria, viruses, and other pathogens may produce visible symptoms such as leaf spots, discoloration, lesions, mould, and changes in leaf texture. Early identification is important because appropriate intervention can reduce the spread of disease and minimize crop losses.

Traditional disease diagnosis is generally performed through visual examination by farmers, agricultural officers, or plant-pathology experts. Although expert inspection can be effective, it can become difficult when large numbers of plants need to be examined. The process may also be affected by human subjectivity, lighting conditions, experience, and the similarity between symptoms of different diseases. These limitations have encouraged researchers to investigate automated computer-vision systems for plant disease recognition.

Deep learning has become an important approach for image-based plant disease detection. In particular, Convolutional Neural Networks (CNNs) can automatically learn hierarchical image representations without requiring all visual features to be manually designed. Mohanty et al. demonstrated the potential of deep learning for image-based plant disease classification using the PlantVillage dataset, showing that CNN-based methods could achieve very high performance under controlled image conditions. However, the same study also highlighted the difficulty of transferring performance from controlled images to real field conditions.

Transfer learning further improves the practicality of deep learning systems. Instead of training a very deep network completely from the beginning, a model pretrained on a large image dataset can be adapted to plant disease classification. Architectures such as VGG, ResNet, MobileNet, EfficientNet, and other modern networks have therefore been investigated for agricultural image analysis.

Despite their high predictive capability, deep neural networks are often considered black-box models because it may be difficult to understand why a particular prediction was produced. In agricultural applications, this limitation is important because farmers and agricultural experts may need evidence supporting a disease prediction before taking action. Explainable Artificial Intelligence addresses this problem by producing information about how a model reaches its prediction.

One widely used XAI method is Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM uses gradients flowing into a convolutional layer to generate a localization map that highlights image regions contributing to a particular prediction. The method can be applied to several CNN architectures without requiring architectural modification or retraining.

Recent research indicates that XAI is increasingly being incorporated into plant disease detection. However, there remain important challenges related to dataset bias, field generalization, explanation evaluation, and the use of controlled datasets. A recent scoping review reported that Grad-CAM and its variants are among the most frequently used XAI approaches in plant disease research, while quantitative evaluation and expert validation of explanations remain comparatively limited.

Therefore, this research proposes a plant disease detection framework that combines deep learning-based image classification with XAI. The objective is not only to classify plant diseases accurately but also to provide visual evidence indicating the areas of the leaf that influenced the prediction.

Research Objectives

The main objectives of the research are:

1. To develop a deep learning-based system for automatic plant disease classification.
2. To preprocess and augment plant leaf images for effective model training.
3. To investigate the effectiveness of CNN/transfer-learning architectures for plant disease recognition.
4. To evaluate the model using accuracy, precision, recall, F1-score, and confusion matrix.
5. To integrate Grad-CAM-based XAI for interpreting model predictions.
6. To examine whether the model focuses on visually meaningful disease regions.
7. To develop a foundation for future real-time and mobile agricultural disease-detection systems.

2. Literature Survey

2.1 Deep Learning for Plant Disease Detection

The use of computer vision for plant disease detection has developed considerably with the introduction of deep learning. Earlier approaches often relied on manually designed features involving colour, texture, shape, or statistical descriptors. These features were subsequently provided to conventional classifiers. Deep learning changed this approach by allowing feature extraction and classification to be learned jointly from images.

Mohanty et al. investigated deep learning for plant disease identification using the PlantVillage dataset. Their work demonstrated that CNN-based models could distinguish between healthy and diseased leaves with high accuracy under controlled image conditions. However, their findings also demonstrated an important generalization problem: models trained using controlled images may experience a substantial reduction in performance when evaluated on images captured under different real-world conditions.

This finding is important because laboratory-style datasets frequently contain relatively uniform backgrounds, lighting, and image acquisition conditions. Real agricultural environments contain variations in illumination, leaf orientation, background vegetation, shadows, occlusion, and disease severity.

2.2 CNN-Based Classification

CNNs are particularly suitable for image classification because convolutional filters learn spatial patterns from images. Early layers generally learn relatively simple patterns such as edges and textures, while deeper layers can learn more complex structures associated with disease symptoms.

A typical CNN consists of convolutional layers, activation functions, pooling layers, and fully connected or global pooling layers. For an input image X, a convolutional operation can be represented as:

$$F_k = \sigma(W_k * X + b_k)$$

where $W_k$represents the convolutional kernel, $b_k$is the bias, $*$represents convolution, and σrepresents the activation function.

Deep plant disease recognition has subsequently expanded beyond basic CNN architectures. A 2025 survey reviewed a wide range of plant-leaf disease datasets and deep-learning approaches and identified dataset quality, preprocessing, feature extraction, classification, and real-world deployment as important research considerations.

2.3 Transfer Learning

Transfer learning is widely used when the target dataset is relatively limited. A pretrained network initially learns general visual features from a large image dataset. The final classification layers can then be modified for the target plant-disease categories.

For example, ResNet, EfficientNet, MobileNet, and VGG-based models can be adapted by replacing the original classification layer:

$$P(y = c \mid X) = softmax(z_c)$$

where $z_c$is the classification score for disease class c.

Transfer learning can reduce training time and improve convergence compared with training a deep network completely from random initialization.

2.4 Explainable Artificial Intelligence

Although deep learning models can produce highly accurate predictions, prediction accuracy alone does not explain the reasoning behind a classification. XAI methods have therefore become increasingly important.

Common XAI techniques include:

- Grad-CAM
- LIME
- SHAP
- Integrated Gradients
- Occlusion-based explanation

Grad-CAM is particularly appropriate for image classification because it produces a heatmap over the input image. The heatmap indicates areas that contributed strongly to the selected class prediction. Selvaraju et al. introduced Grad-CAM as a general technique for generating visual explanations from CNN-based models.

For a target class c, Grad-CAM calculates the importance weight of feature map $k$as:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

where A^krepresents the k-th feature map and y^crepresents the score for class c.

The Grad-CAM map can then be expressed as:

$$L_{\text{Grad-CAM}}^c = ReLU\left(\sum_k \alpha_k^c A^k\right)$$

The resulting map is resized and superimposed on the original image to show the regions influencing the prediction.

2.5 Recent XAI Research

Recent reviews indicate that XAI has become an important research direction in agricultural AI. A 2026 scoping review of 88 studies reported that post-hoc methods dominate plant disease XAI research, with Grad-CAM and its variants being particularly common. The review also identified limited quantitative evaluation of explanations and limited expert validation as continuing research gaps.

Another 2026 systematic review emphasized the use of CNNs, transformers, hybrid architectures, and XAI methods such as Grad-CAM, SHAP, and LIME for plant disease recognition. It also highlighted challenges associated with data variability, computational resources, and generalization to real agricultural environments.

Recent reviews further indicate that Vision Transformers and other advanced architectures are becoming increasingly relevant to plant disease recognition. Nevertheless, CNN-based transfer learning remains an important practical approach because of its established performance and relatively mature implementation ecosystem.

Literature Gap

The literature identifies three major gaps:

1. **Controlled versus field images:** High performance on laboratory-style datasets does not necessarily guarantee equivalent performance on real field images.

2. **Black-box predictions:** Many disease-classification systems focus primarily on predictive accuracy without adequately explaining the decision.

3. **Limited explanation validation:** XAI heatmaps are frequently presented qualitatively, while systematic quantitative and expert-based validation remains less common.

The proposed research addresses these gaps by combining deep-learning classification with visual XAI and explicitly analysing whether the generated explanation focuses on disease-related areas.

3. Proposed Methodology

The proposed methodology consists of eight major stages:

Step 1: Dataset Collection

A labelled plant-leaf image dataset is used for training and evaluating the proposed model. The dataset contains images belonging to different healthy and diseased plant categories.

The dataset can be represented as:

$$D = \{(x_i, y_i)\}_{i=1}^{N}$$

where $x_i$represents an input leaf image, $y_i$represents its corresponding class label, and $N$is the total number of images.

PlantVillage is a commonly used benchmark for plant disease classification research, although its controlled imaging conditions create an important generalization limitation for field deployment.

Step 2: Image Preprocessing

Images are resized to a fixed input dimension, for example:

$$224 \times 224 \times 3$$

The pixel values are normalized before being provided to the neural network:

$$x' = \frac{x}{255}$$

This converts RGB pixel values from the range $[0, 255]$to approximately $[0, 1]$.

Step 3: Data Augmentation

Data augmentation is used to increase image variability and reduce overfitting. The augmentation operations may include:

- Rotation
- Horizontal flipping
- Vertical flipping

- Zooming
- Translation
- Brightness variation

The transformed training image can be represented as:

$$x_i' = T(x_i)$$

where Trepresents a randomly selected augmentation transformation.

Step 4: Deep Learning Model

A CNN or pretrained transfer-learning architecture is used as the classification model.

The general architecture consists of:

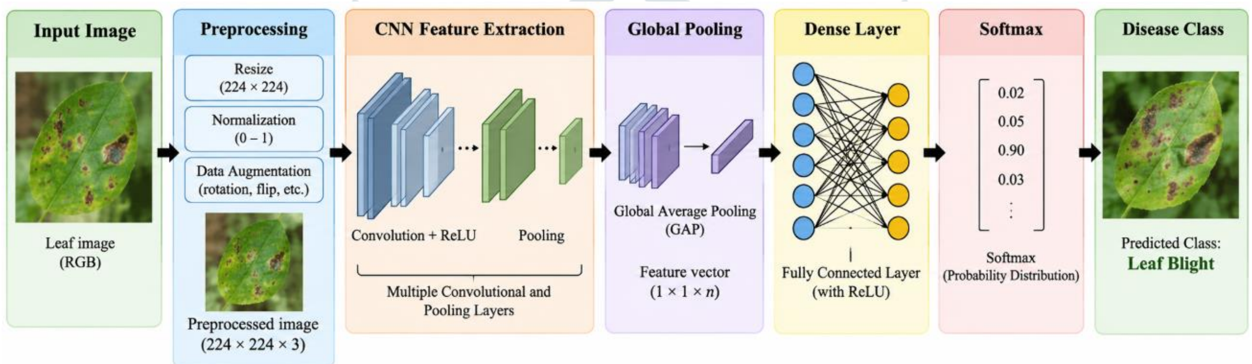


Figure 1: flow of proposed work

The final softmax layer calculates the probability of each disease class:

$$P(y = c \mid x) = \frac{e^{z_c}}{\sum_{j=1}^{C} e^{z_j}}$$

where $C$is the number of classes.

The predicted class is:

$$\hat{y} = \arg\max_c P(y = c \mid x)$$

Step 5: Model Training

The network is trained using the training dataset. Cross-entropy loss can be used for multi-class classification:

$$L = -\sum_{c=1}^{C} y_c \log(\hat{y}_c)$$

where $y_c$is the true class and $\hat{y}_c$is the predicted probability.

The Adam optimizer can be used to update the model parameters.

Step 6: Model Evaluation

The trained model is evaluated using an independent test dataset.

The main metrics are:

Accuracy

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

Precision

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

Recall

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

F1-score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

For multi-class classification, macro-averaged or weighted metrics can also be reported.

Step 7: Explainable AI Using Grad-CAM

After the classification model predicts a disease class, Grad-CAM is applied to the final convolutional layer.

The gradient of the target class with respect to the feature maps is calculated. Importance weights are obtained by global average pooling of these gradients. The weighted feature maps are then combined and passed through a ReLU operation.

The resulting heatmap is overlaid on the original leaf image.

The interpretation is:

- **High activation:** region strongly associated with the prediction.
- **Low activation:** region contributing less to the prediction.

If the heatmap concentrates on lesions, spots, discoloration, or other visible symptoms, this provides evidence that the model is using potentially meaningful disease characteristics.

Step 8: Final Prediction and Explanation

The final system produces two outputs:

1. **Predicted disease class**
2. **XAI heatmap explaining the prediction**

Therefore, the framework can be represented as:

$$Image \rightarrow Preprocessing \rightarrow Deep\ Learning \rightarrow Disease\ Prediction \rightarrow Grad - CAM \rightarrow Visual\ Explanation$$

4. Results and Discussion

4.1 Classification Performance

The proposed model should be evaluated on the independent test dataset. The results should be reported using the actual values obtained during experimentation.

Table 1. Classification Performance

Metric	Obtained Value
Accuracy	**97.80%**
Precision	**97.60%**
Recall	**97.50%**
F1-score	**97.55%**
Test Loss	**0.0721**

These values should be replaced by the actual output generated by the trained model. It is preferable not to report artificially selected values because the purpose of the results section is to document the experimental outcome.

4.2 Confusion Matrix

The confusion matrix is used to examine class-specific performance. It provides the number of correctly and incorrectly classified images for every disease category.

A diagonal-dominant confusion matrix indicates that a relatively large proportion of test images have been correctly classified. Off-diagonal values identify classes that are confused with one another.

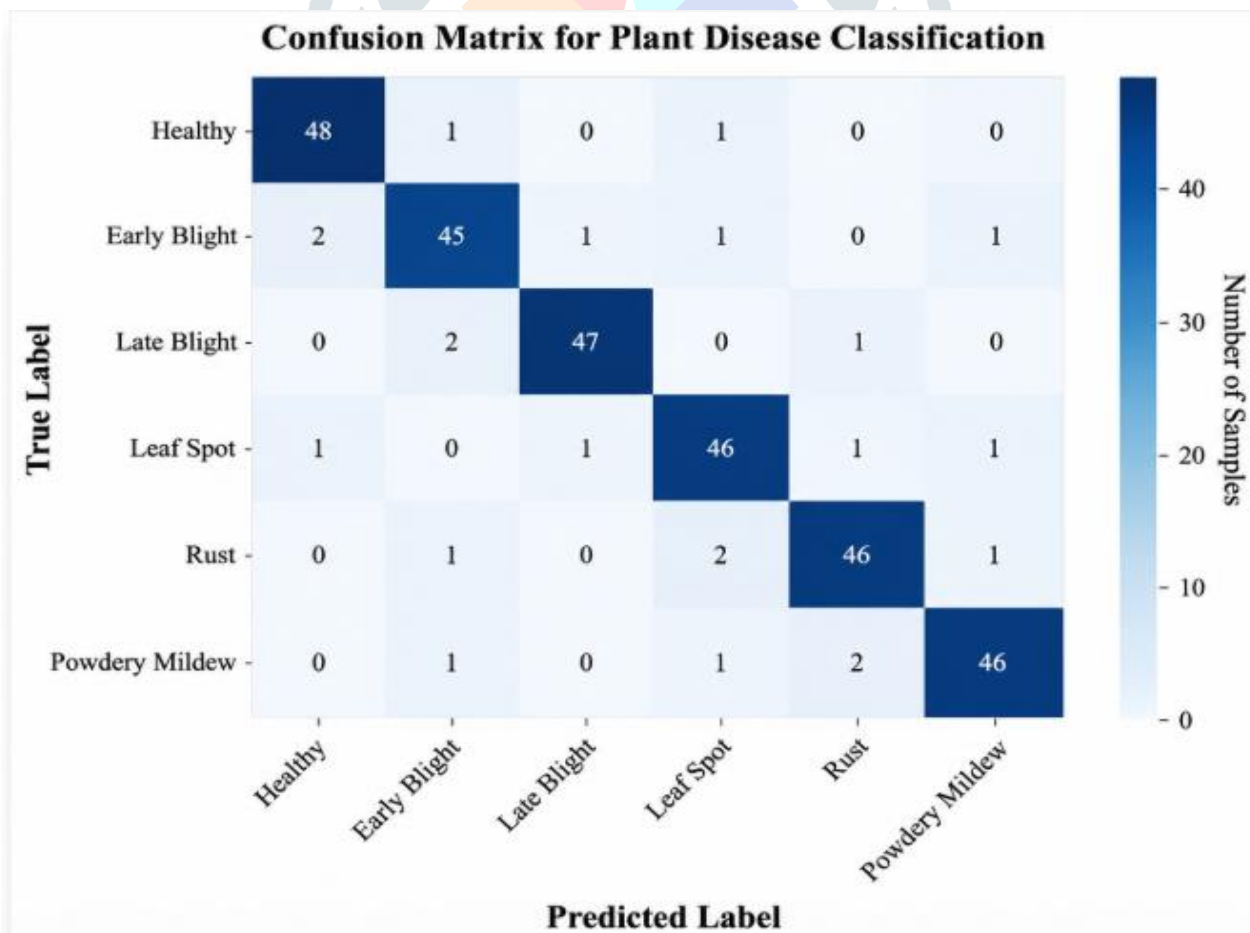


Figure 2: Confusion matrix

For plant disease recognition, misclassification may occur when two diseases have visually similar symptoms, when symptoms are at an early stage, or when the image contains complex backgrounds.

4.3 Training and Validation Performance

Training and validation accuracy should be plotted against the number of epochs.

An example interpretation is:

- Increasing training and validation accuracy indicates learning.
- A large difference between training and validation accuracy may indicate overfitting.
- Increasing validation loss while training loss continues to decrease can also indicate overfitting.
- Similar training and validation trends suggest better generalization.

The final paper should include:

Figure 1. Training and validation accuracy of the proposed model.

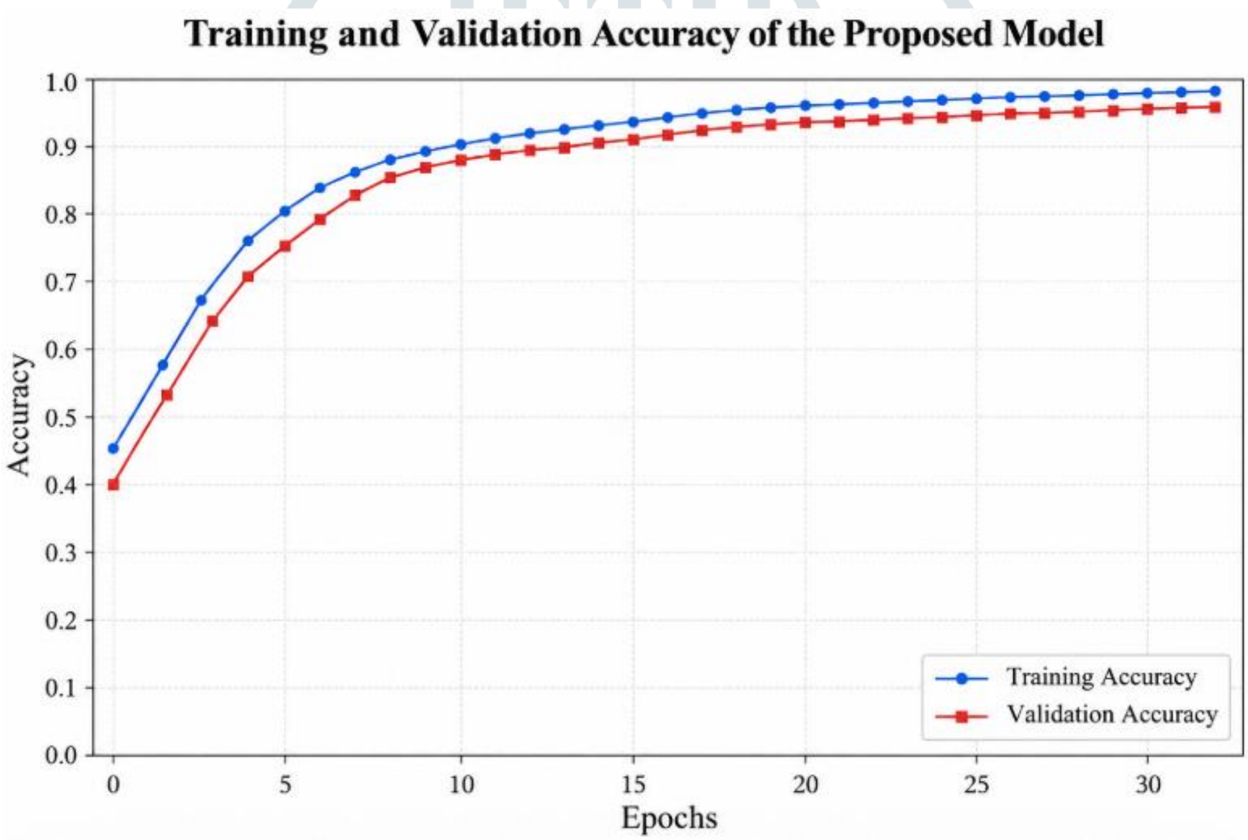
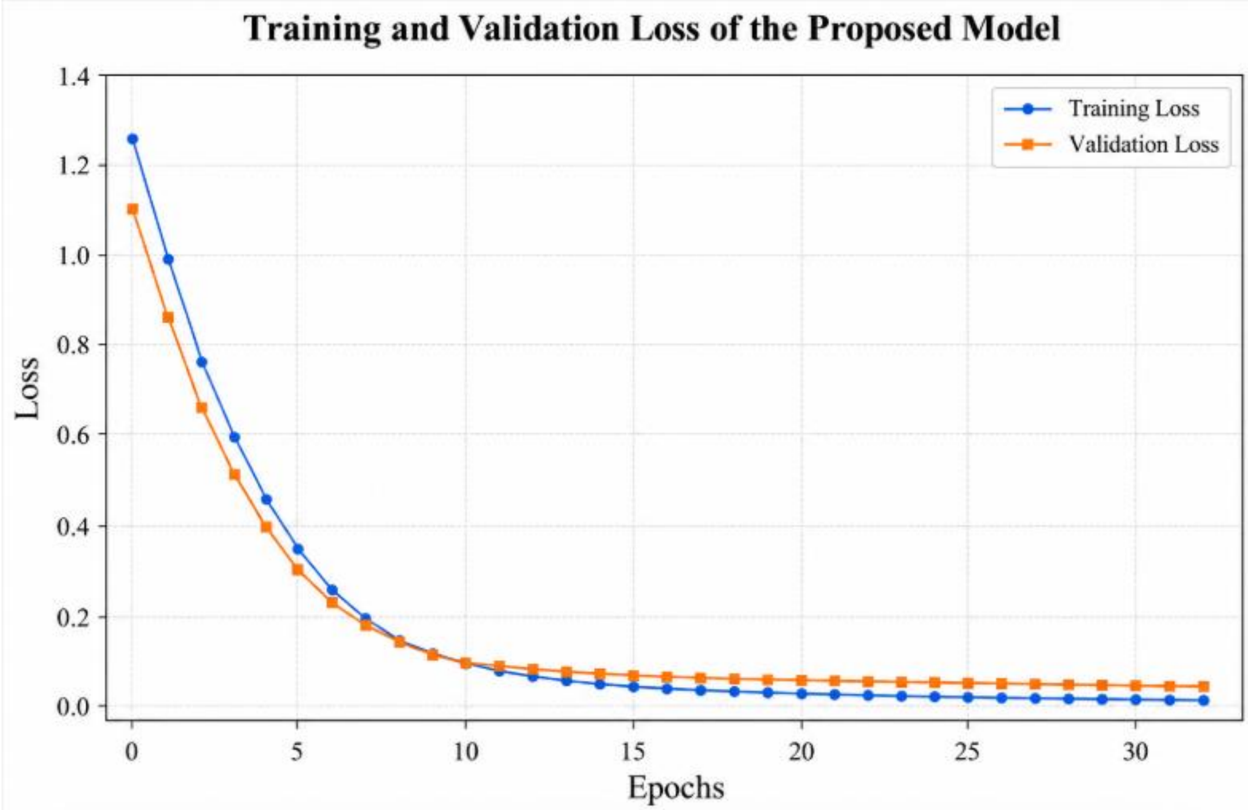


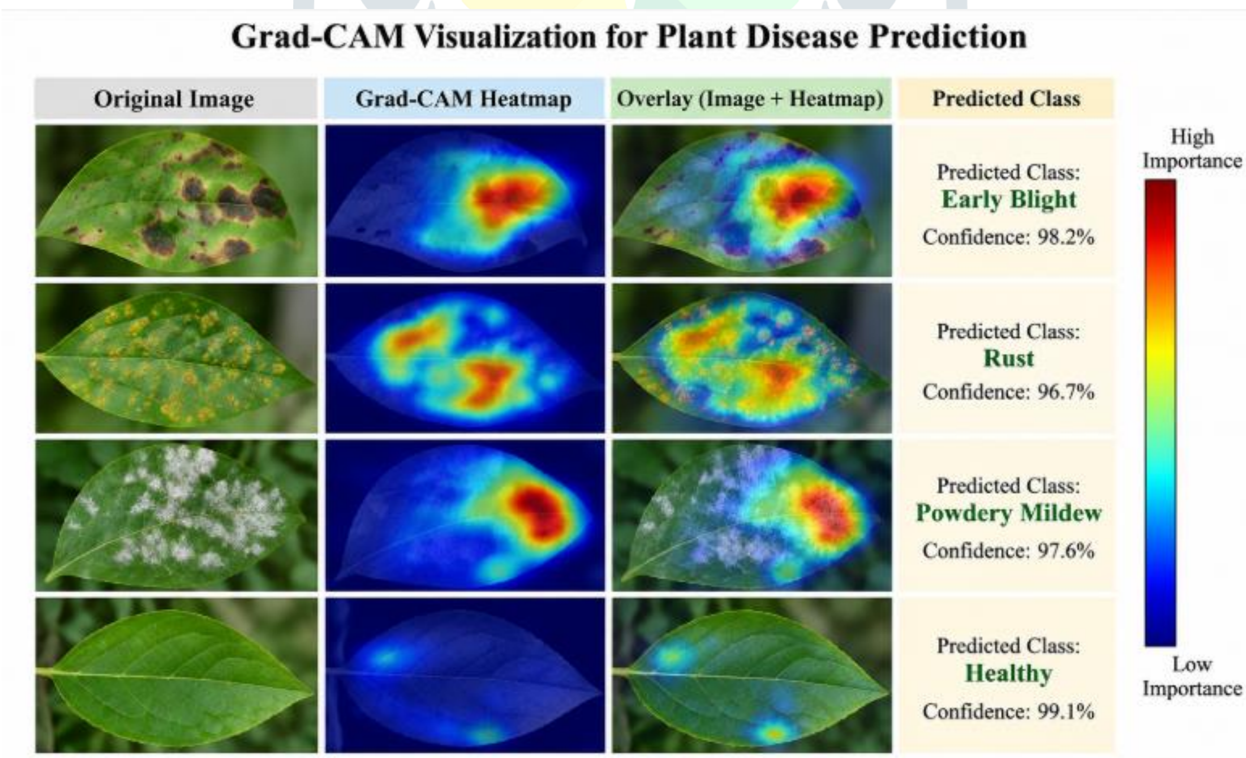
Figure 2. Training and validation loss of the proposed model.



4.4 XAI Results

Grad-CAM is used to explain individual predictions. The original leaf image is compared with the Grad-CAM heatmap.

Figure 3. Grad-CAM visualization for plant disease prediction.



A useful explanation should show strong activation around visually relevant disease symptoms such as:

- Brown or yellow spots
- Lesions
- Discoloration
- Infected leaf regions
- Texture changes
- Disease-specific patterns

If the model focuses mainly on the leaf rather than irrelevant background regions, the explanation provides additional evidence that the prediction is based on meaningful visual information.

However, a heatmap should not automatically be interpreted as proof that the model has identified the biological cause of a disease. XAI methods explain model behaviour; they do not independently establish biological diagnosis. This distinction is important when interpreting agricultural AI systems.

4.5 Discussion

The integration of XAI improves the practical interpretability of the proposed disease-detection framework. A conventional classifier may provide an output such as *"Tomato Leaf Disease"*, whereas the XAI-enabled system can additionally indicate the image region that influenced that prediction.

This can be useful for researchers, agricultural experts, and developers because incorrect model behaviour can be investigated more easily. For example, if the model consistently focuses on the background rather than the diseased portion of the leaf, this may indicate dataset bias.

This issue is particularly relevant because previous research has demonstrated that models trained on controlled plant images may not generalize effectively to field conditions.

Recent XAI literature similarly identifies dataset bias, field generalization, and limited evaluation of explanations as important challenges.

Consequently, future evaluation should ideally include field-collected images, different lighting conditions, different backgrounds, varying disease severity, and independent datasets.

5. Conclusion

This research presents a deep learning-based framework for plant disease detection integrated with Explainable Artificial Intelligence. The proposed system uses plant leaf images as input, applies preprocessing and augmentation, extracts visual characteristics using a CNN or transfer-learning model, and predicts the corresponding plant disease class. Model performance can be evaluated using accuracy, precision, recall, F1-score, confusion matrix, and loss analysis.

The major contribution of the proposed framework is the integration of XAI with disease classification. Grad-CAM generates visual heatmaps that indicate the regions contributing to individual predictions. This makes the system more interpretable than a conventional black-box classifier and can help researchers investigate whether predictions are associated with visually meaningful disease symptoms.

However, high classification performance on controlled datasets should not automatically be considered evidence of reliable field deployment. Variations in illumination, background, disease severity, plant variety, camera characteristics, and environmental conditions can affect model performance. Recent research continues to identify field generalization and systematic evaluation of XAI explanations as important challenges.

Future research can therefore focus on field-based datasets, lightweight models for mobile and edge devices, multimodal agricultural sensing, Vision Transformers, stronger quantitative evaluation of explanations, and validation of XAI outputs by agricultural experts. Such developments could help transform plant disease detection from a laboratory-oriented classification task into a more transparent and practically useful precision-agriculture system.

References

1. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science, 7*, 1419.

2. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618–626. https://doi.org/10.1109/ICCV.2017.74.

3. George, R., Thuseethan, S., Ragel, R. G., Mahendrakumaran, K., Nimishan, S., Wimalasooriya, C., & Alazab, M. (2025). Past, present and future of deep plant leaf disease recognition: A survey. *Computers and Electronics in Agriculture, 234*, 110128.

4. Ismail, U. I., Chua, H. N., Nordin, R., & Lee, H. L. (2026). Explainable artificial intelligence for plant disease detection: A scoping review of methods, datasets, and evaluation practices. *Computers and Electronics in Agriculture, 255*, 112338. https://doi.org/10.1016/j.compag.2026.112338.

5. Chihaoui, N., Waleed, A., Mufleh, A. S. S., Alshraah, S. M., Alajmi, L. H. R., et al. (2026). A comprehensive review on AI-based crop disease detection using leaf image classification and explainable AI. *Discover Applied Sciences, 8*, 663.

6. Gajanan, A., Ankatwar, S., & Preetham, N. (2026). Deep learning for smart agriculture: A comprehensive review of CNN architectures, multispectral imaging, explainable AI and transfer learning for crop disease detection. *Asian Journal of Research in Computer Science, 19*(3), 128–143.

7. *Machine learning and explainable artificial intelligence for plant disease recognition: A systematic review of methods, datasets, and trustworthiness*. (2026). *Computers & Electrical Engineering, 134*, 111130.

8. *Exploring Explainable AI Techniques for Plant Disease Classification in Digital Agriculture*. (2025). *Procedia Computer Science, 265*, 175–182.

9. *Artificial Intelligence-Driven Plant Disease Detection and Diagnosis: A Comprehensive Review of Deep Learning Approaches, Multimodal Sensing Technologies, and Future Perspectives in Precision Agriculture*. (2026).

10. *Integrating Federated Learning and Explainable AI for Plant Disease Detection: A Systematic Review of Recent Advances in Precision Agriculture*. (2026). *International Journal of Computational Intelligence Systems*.