INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

Deep Learning-Based Potato Leaf Disease Detection Using Transfer Learning and Explainable AI Techniques

Aradhy Tiwari[1], **Amit Kumar Saxena**[1*], **Chandrashekhar**[1]

1 Department of CSIT, Guru Ghasidas Vishwavidyalaya, Bilaspur, India

OPEN ACCESS

Received: 01/05/2026

Accepted: 14/07/2026

Published: 05/08/2026

Citation: Tiwari A, Saxena AK, Chandrashekhar (2026) Deep Learning-Based Potato Leaf Disease Detection Using Transfer Learning and Explainable AI Techniques. Indian Journal of Science and Technology 19(29): 2085-2097. https://doi.org/10.17485/IJST/v19i29.716

* **Corresponding author**.

amitsaxena65@rediffmail.com

Funding: None

Competing Interests: None

Copyright: © 2026 Tiwari et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: To create a multi-class image classification system to automate the detection of potato crop diseases using deep learning algorithms to classify images of potato leaves. **Method**: This study involves an implementing and comparing of six deep learning models to classify potato leaves as diseased or infected with pests. The models included a custom CNN as the baseline and five transfer-learning models: VGG16, DenseNet121, MobileNetV2, Xception, and InceptionV3. The final selected model was InceptionV3 due to its ability to extract strong features and achieve superior overall classification performance among all evaluated models. To enhance model's performance and improve generalization to unseen data, several techniques were implemented, including data augmentation, Batch Normalization, Dropout regularization, and selective fine-tuning of deeper layers. **Findings**: The proposed model achieved the highest test accuracy (94%) and macro-average F1-score (0.94) compared to other baseline models. The importance of fine-tuning is reflected in the high accuracy of the proposed model. An ablation study found that accuracy dropped to 84.67% without fine-tuning, which demonstrates how critical it is for this model's domain adaptation. The Grad-CAM analysis showed that the model focuses on biologically relevant areas of the leaves with infection and does not concentrate on backgrounds; therefore, the results indicate the model's potential for interpretability and deployment in real-world settings. **Novelty**: This study improves potato leaf disease detection using a fine-tuned InceptionV3 with data augmentation and dropout, while Grad-CAM visualizations enhance model interpretability, reliability, and practical utility for accurate agricultural disease diagnosis.

Keywords: PotatoLeaf Disease Detection; Deep Learning; Transfer Learning; InceptionV3; Image Classification; Grad-CAM; Sustainable Agriculture

1 Introduction

The potato (*Solanum tuberosum*) is one of the most important food crops in the world and contributes significantly to global food security. With annual production exceeding 370 million metric tonnes, it serves as a major source of carbohydrates, vitamins, minerals, and dietary fiber for millions of people, especially in developing regions [1], [2]. However, potato cultivation is highly vulnerable to several diseases and pest-related stresses, including Early Blight, Late Blight, fungal infections, nematode attacks, and other leaf disorders [3]. These problems can cause serious yield and quality losses if not detected early. Historically, Late Blight was responsible for catastrophic crop failures such as the Irish Potato Famine, showing the severe impact plant diseases can have on society [4].

In conventional agricultural practice, disease diagnosis mainly depends on visual inspection by trained experts. Although this method is widely used, it is often slow, expensive, subjective, and impractical for large farming areas. Accurate diagnosis also becomes more difficult because many diseases and pest damage show similar visual symptoms on leaves. As a result, farmers may apply incorrect treatments or fail to respond in time, leading to reduced productivity and unnecessary economic losses [5]. Therefore, there is a strong need for automated, fast, and reliable systems for potato disease identification.

Recent advancements in artificial intelligence [6], [7] and computer vision have created new opportunities for plant disease detection. Deep learning, particularly Convolutional Neural Networks (CNNs), has shown remarkable success in image classification tasks because these models can automatically learn complex visual patterns such as lesion texture, color variation, and shape directly from raw images [8], [9]. Unlike traditional machine learning approaches [10], [11], deep learning approaches reduce the need for manual feature engineering and provides more robust performance under varying conditions [12]. Several studies have already demonstrated the usefulness of architectures such as VGG16, DenseNet, MobileNetV2, Xception, and InceptionV3 in agricultural applications [9], [13].

Although previous studies reported promising results, important limitations still remain. Sholihati et al. [14] used VGG16 and VGG19 for potato disease classification, achieving 91% accuracy without fine-tuning or data augmentation. Erlin et al. [15] compared five CNN models and observed repeated misclassification of the Healthy class, with VGG19 performing the worst at 90%. Mishra et al. [16] evaluated CNN, AlexNet, ResNet, and EfficientNet, where AlexNet achieved only 90.2%, while explainability methods were not considered. Charisma and Adhinata [17] applied DenseNet201 but did not include fine-tuning or Grad-CAM analysis. Bhavani and Chalapathi [18] compared MobileNetV2 and InceptionV3, but did not explore selective layer fine-tuning. Similarly, Alhammad et al. [19] combined transfer learning with explainable AI, but their VGG-based models achieved only 91% accuracy and lacked stronger regularisation techniques. These findings indicate clear research gaps in optimization, generalization, and interpretability.

To overcome these limitations, the present study develops a comprehensive deep learning framework for six-class classification of potato leaf diseases. Multiple architectures are systematically compared, and the proposed fine-tuned InceptionV3 model achieves the highest test accuracy. In addition, an ablation study is conducted to analyze the effects of augmentation, dropout, and fine-tuning on performance. Grad-CAM-based visual explanations are also incorporated to show that the model focuses on biologically meaningful regions of the leaf. By combining high accuracy with transparency and scalability, this research provides a practical and trustworthy solution for intelligent disease management in modern precision agriculture.

1.1 Research Gap

There are still significant gaps in the new literature on detecting plant diseases using deep learning techniques, even though great strides have been made in this area. The majority of previous investigations on this topic [14], [15] did not use selective fine-tuning; therefore, they did not enable effective domain adaptation. There have been few studies that have compared results across multiple networks in terms of detection accuracy, and explainability techniques like Grad-CAM have been almost non-existent in previous studies on potato leaf disease [16], [17]. Additionally, no structured ablation studies have been conducted to determine the relative contributions of augmentation, dropout, and fine-tuning to the overall performance of these networks. This paper aims to fill those gaps by providing a thorough comparison analysis of multiple architectures using both explainability methods and ablation studies to quantify how much each aspect of network training affects overall performance. The Section 2 of the paper presents the Methodology adopted in this paper with experiment. Section 3 presents Results and Discussions whereas the Conclusions are presented in Section 4.

2 Methodology

To provide an objective and reproducible evaluation of six deep learning architectures, we implemented all models using a standardized experimental protocol that included a baseline CNN, MobileNetV2, VGG16, Xception, DenseNet121, and the proposed InceptionV3. The models were trained on the same processed dataset while ensuring consistency between augmentation methods, but not on case-by-case input dimensionality due to architectural differences. We used a systematic approach to create reliable data preparation processes, build robust models, and provide fair model evaluations. The strategy for optimization, verification, and comparison was used to improve generalization and identify the best architecture for image classification. All components of the research process are shown in Figure 1 to illustrate the complete workflow, from online dataset collection and preprocessing to model training, evaluation, and comparative performance analysis of six models.

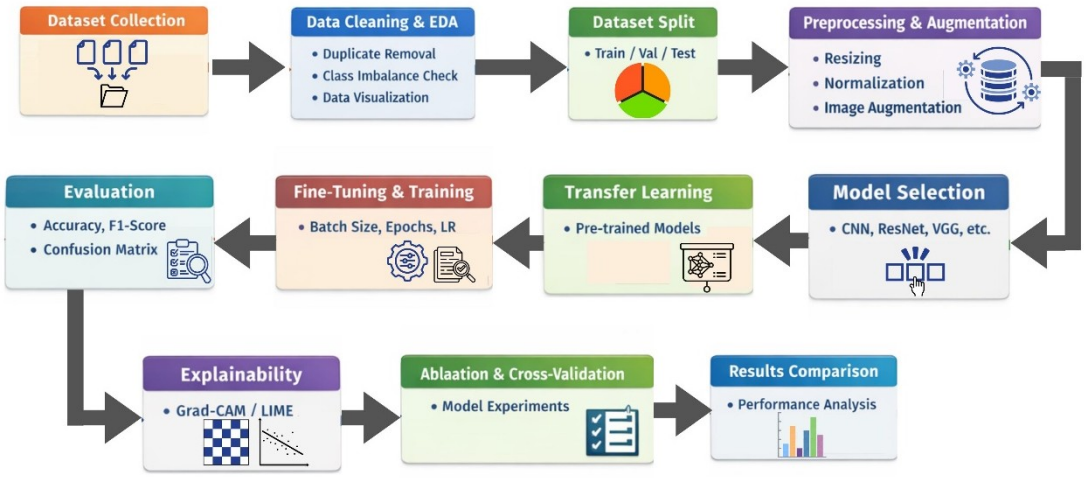


Fig 1. Overall research workflow.

2.1 Dataset Description

The potato leaf image dataset was used in the experiments and is publicly available through Kaggle (https://www.kaggle.com/). The dataset contains 3,000 RGB images across 6 classes, with 500 images per class. The six categories include Early Blight, Fungal Diseases, Late Blight, Plant Pests, Potato Cyst Nematode, and Healthy Leaves.

The images can be represented in .jpg or .jpeg format and are stored in folders corresponding to each class, allowing simple extraction of image labels by loading from directories.The dataset was divided into training (80%), validation (10%), and testing (10%) sets using a stratified split to preserve the class distribution across all subsets. The dataset contains images of leaves taken under a variety of conditions, providing diversity in symptoms of disease, background differences, and appearance characteristics. The combination of balanced and diverse images makes this dataset suitable for training and evaluating multi-class potato leaf disease classification using deep learning methods.

2.2 Image Preprocessing

The potato leaf dataset was preprocessed to improve image consistency and quality prior to training multiple deep learning models. Using OpenCV, images from all six classes were read from their respective folders and resized according to model input requirements. Images were resized to $224 \times 224 \times 3$ for models such as MobileNetV2, VGG16, and DenseNet121, while the proposed InceptionV3 model used an input size of $299 \times 299 \times 3$.

To reduce noise and improve visual consistency, a Gaussian blur with a 5×5 kernel was applied. Further enhancement was performed in the LAB color space using CLAHE (clip limit = 2.0, tile grid size = 8×8) on the L-channel to improve contrast and highlight disease-affected regions. The enhanced channels were then recombined and converted back to a standard color image. All preprocessing operations and real-time data augmentation techniques were applied exclusively to the training dataset, while the validation and test datasets were only resized and preprocessed using the corresponding model-specific preprocessing functions for evaluation. This approach prevented data leakage and ensured an unbiased assessment of the model's performance.

During training, model-specific preprocessing functions and real-time data augmentation techniques were applied to improve generalization and robustness.

2.3 Deep Learning Models and Comparative Experimental Setup

To assess the capability of different deep learning architectures for potato leaf disease classification, six models were developed and evaluated under a consistent experimental framework. These included one baseline CNN model and five transfer learning models: MobileNetV2, DenseNet121, VGG16, Xception, and the proposed InceptionV3 model.All models were trained using the same dataset split, preprocessing pipeline, and comparable training settings to ensure a fair comparison. Performance was assessed using classification accuracy, class-wise precision, recall, F1-score, and confusion matrix analysis, with particular attention to the separation of visually similar disease categories. The comparative results highlighted the strengths and limitations of each architecture, while the proposed InceptionV3 model achieved the best overall performance among the evaluated models.

2.3.1. Baseline CNN Model

A custom Convolutional Neural Network (CNN) was developed to establish a baseline for evaluating the effectiveness of deep learning in classifying potato leaf diseases. The model accepts input images of size $224 \times 224 \times 3$ and consists of three convolutional blocks with increasing numbers of filters (32, 64, and 128), each followed by a max-pooling layer for spatial down-sampling. The extracted feature maps are flattened and passed through a dense layer with 128 neurons, followed by a dropout layer (0.5) to reduce overfitting. The final layer uses the softmax activation function for six-class classification.

The model was trained using the Adam optimizer with categorical cross-entropy loss and minimal preprocessing through pixel rescaling. The CNN achieved an overall test accuracy of 76%, indicating moderate classification performance. It performed well on visually distinctive classes such as Early Blight and Potato Cyst Nematode, but showed weaker performance on Fungal Diseases (F1-score: 0.55), suggesting difficulty in distinguishing subtle visual differences among similar disease symptoms. These findings indicate that while the baseline CNN can learn useful patterns, more advanced architectures are required to achieve greater robustness and accuracy.

2.3.2. MobileNetV2 Model

MobileNetV2 is a lightweight convolutional neural network designed for efficient image classification with low computational cost. In this study, MobileNetV2 was employed as a transfer learning model for potato leaf disease classification. A pretrained MobileNetV2 model with ImageNet weights was used as the feature extractor with an input size of $224 \times 224 \times 3$. The original top layers were removed and replaced with a custom classification head consisting of Global Average Pooling, a dense layer with 256 neurons using ReLU activation, a Dropout layer of 0.4, and a softmax output layer for six-class classification.

The model was trained using augmented image data and optimized with the Adam optimizer and categorical cross-entropy loss. In the fine-tuning stage, the last 20 layers of the base network were unfrozen and retrained using a low learning rate (5×10^{-6}) with label smoothing to adapt pretrained features to potato leaf disease patterns. The MobileNetV2 model achieved a test accuracy of 91%, demonstrating strong performance across most classes. It showed excellent results for visually distinct categories such as Early Blight and Potato Cyst Nematode, while comparatively lower performance was observed for more challenging classes such as Fungal Diseases and Plant Pests, where visual symptoms are more similar.

2.3.3. DenseNet121 Model

DenseNet121 is a deep convolutional neural network that uses a dense connectivity mechanism, where each layer receives feature maps from all preceding layers and passes its own outputs to subsequent layers. This architecture promotes feature reuse, improves gradient flow, and helps reduce the vanishing gradient problem during training. In this study, a DenseNet121 model pre-trained on ImageNet was employed as a transfer-learning model for potato leaf disease classification, using input images of size $224 \times 224 \times 3$.

The original top layers of the pretrained network were replaced with a custom classification head consisting of a Global Average Pooling layer, a dense layer with 128 neurons using ReLU activation, and a softmax output layer for six-class prediction. All layers of the base DenseNet121 model were initially frozen to retain and transfer the learned ImageNet features to the target task. The model was trained using the Adam optimizer with categorical cross-entropy loss.

The DenseNet121 model achieved an overall test accuracy of 90%. It showed excellent performance for Early Blight (F1-score: 0.99) and Potato Cyst Nematode (F1-score: 0.97), while comparatively lower performance was observed for Plant Pests

(F1-score: 0.80). These results indicate that DenseNet121 can learn strong and discriminative feature representations, although certain visually similar classes remain more challenging to classify.

2.3.4. VGG16 Model

VGG16 is a popular deep convolutional neural network (CNN) known for its straightforward and uniform architecture, consisting of multiple stacked convolutional layers with 3×3 filters. In this study, transfer learning was applied using a VGG16 model pre-trained on ImageNet for potato leaf disease classification with an input size of 224 × 224 × 3.

The original fully connected layers were removed and replaced with a custom classification head comprising a Global Average Pooling layer, a Dense layer with 128 neurons and ReLU activation, and a softmax output layer for six-class prediction. The convolutional base was kept frozen to preserve the rich visual features learned during ImageNet pre-training. The model was trained using the Adam optimizer with categorical cross-entropy loss.

The final test results showed an overall accuracy of 86%. The model achieved strong performance for Early Blight (F1-score: 0.99) and Potato Cyst Nematode (F1-score: 0.94), while showing lower performance for Fungal Diseases (F1-score: 0.72), suggesting that some disease classes share similar visual characteristics. Overall, VGG16 provided stable feature extraction and moderate classification performance.

2.3.5. Xception Model

Xception is a deep convolutional neural network architecture that uses depth-wise separable convolutions to improve computational efficiency while maintaining strong feature-extraction capabilities. In this study, a pre-trained Xception model (trained on ImageNet) was applied via transfer learning to potato leaf disease classification, using input images of size 224 × 224 × 3.

The original top layers were removed and replaced with a custom classification head consisting of a Global Average Pooling layer, a Dense layer with 256 neurons using ReLU activation, a Dropout layer of 0.5, and a softmax output layer for six-class prediction. Initially, all layers of the pretrained backbone were frozen, and only the custom head was trained, achieving a test accuracy of 85%.

To further improve performance, fine-tuning was performed by unfreezing the last 20 layers of the Xception backbone and retraining the model using a lower learning rate. After fine-tuning, the model achieved an overall test accuracy of 93%. Notable improvements were observed for Fungal Diseases (F1-score: 0.84) and Plant Pests (F1-score: 0.89). These findings demonstrate that fine-tuning helps the model better adapt to domain-specific visual patterns and significantly enhances classification performance.

2.3.6. Proposed Method: InceptionV3

The image classification algorithm is designed to classify potato leaf diseases using an adapted version of the InceptionV3 transfer learning neural network, which was previously trained on the large-scale ImageNet dataset. As illustrated in Figure 2, the model accepts an input potato leaf image of size 299 × 299 × 3 and passes it through the InceptionV3 deep convolutional feature extractor, where multiple convolutional and inception modules generate discriminative feature maps for classification of the six potato leaf disease classes.

The original classification layers of the InceptionV3 network were removed and replaced with a customized classifier head consisting of a Global Average Pooling layer, Batch Normalization layer, Dense layer with 256 neurons using the ReLU activation function, Dropout layer (0.5), and a Softmax output layer corresponding to the six disease categories, as shown in Figure 2. To improve generalization capability, data augmentation was applied to the training images, including rotation (±25°), zooming, width/height shifting, and horizontal flipping.

The model was trained using the Adam optimization algorithm with a learning rate of 0.0001 and categorical cross-entropy as the loss function. During the fine-tuning phase, the last 30 layers of the pre-trained InceptionV3 backbone were unfrozen and retrained so that the high-level learned features could be better adapted to potato leaf disease characteristics.

The proposed model achieved an overall test accuracy of 94%. Its performance was comprehensively evaluated using confusion matrix analysis, precision, recall, F1-score, ROC curves, misclassification analysis, and Grad-CAM visual explanations. Furthermore, the complete model attained 98.00% validation accuracy in the ablation study, while removing the fine-tuning stage reduced performance to 84.67%, confirming the critical importance of transfer learning adaptation.

The complete network contains approximately 22.3 million parameters and required around 775 seconds for training. When compared with other evaluated architectures such as VGG16, DenseNet121, MobileNetV2, and Xception, the proposed approach demonstrated strong and robust performance, making it suitable for practical smart agriculture applications.

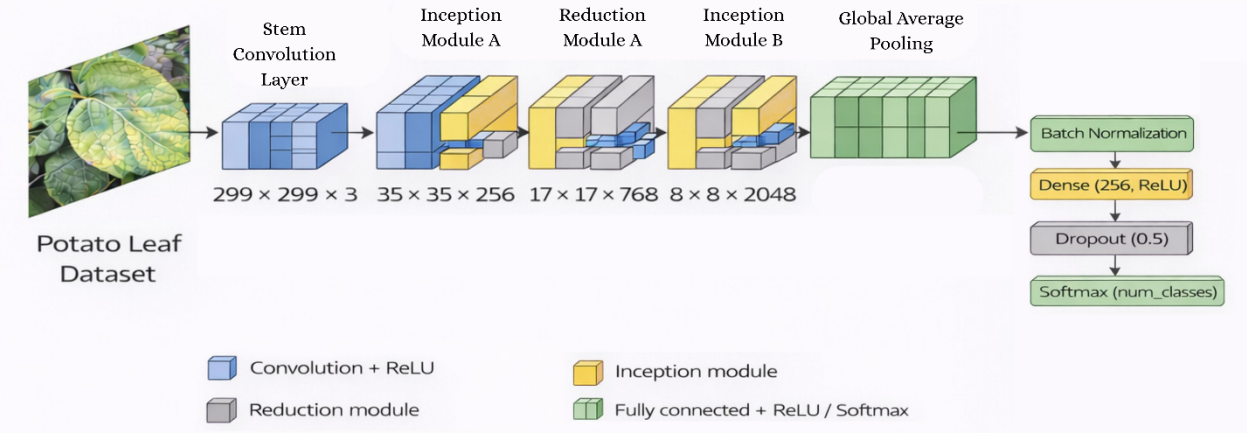


Fig 2. Proposed InceptionV3-based transfer learning architecture for six-class potato leaf disease classification

3 Results and Discussion

3.1 Comparative Performance Analysis

Each model in the comparative evaluation was assessed using four performance metrics: accuracy, precision, recall, and F1-score. The results are presented in Table 1, while the comparison of overall accuracy is illustrated in Figure 3.

Table 1. Performance comparison of evaluated models using accuracy, precision, recall, and F1-score.

Model	Accuracy (%)	Precision	Recall	F1-Score
CNN	76	0.77	0.76	0.76
VGG16	86	0.86	0.86	0.86
DenseNet121	90	0.90	0.90	0.90
MobileNetV2	91	0.91	0.91	0.91
Xception	93	0.94	0.93	0.93
Proposed	94	0.94	0.94	0.94

The accuracy for the baseline CNN Model was 76%, suggesting that the baseline model lacks sufficient capability to extract complex discriminative features from potato leaf images. Both VGG16 and DenseNet121 achieved accuracies of 86% and 90%, respectively, using transfer learning, significantly improving upon the baseline model. The MobileNetV2 model achieved a classification accuracy of 91%, indicating that lightweight architectures can deliver very good classification performance while increasing computational efficiency.

The fine-tuned Xception model's accuracy (93%) made it one of the highest-performing benchmark models. However, the highest test accuracy (94%) was achieved by the proposed InceptionV3-based model, outperforming all other compared architectures. Additionally, under optimized validation conditions, the proposed model achieved an accuracy of 98%.

In addition to accuracy, the proposed model demonstrated other characteristics of effectiveness, with a precision of 0.94, a recall of 0.94, and an F1 score of 0.94, indicating strong robustness and relatively balanced performance and generalization across all disease classes.

The proposed InceptionV3 model's test accuracy of 94% surpasses the 91% accuracy reported by Sholihati et al.[14] using VGG16 and VGG19, which were trained without fine-tuning or augmentation, thereby confirming that selective layer fine-tuning is a critical factor in domain adaptation for agricultural image classification. This finding is further supported by Erlin et al.[15], who observed repeated misclassification of the Healthy class across five CNN models, with the best model achieving only 90% accuracy, a limitation that the present study addresses through structured fine-tuning and data augmentation. Similarly, Mishra et al.[16] evaluated CNN, AlexNet, ResNet, and EfficientNet, achieving a maximum of 90.2% accuracy, but did not incorporate any explainability method; in contrast, the present study integrates Grad-CAM to validate that predictions are grounded in biologically meaningful leaf regions rather than background noise. Charisma and Adhinata[17] applied DenseNet201 without fine-tuning or Grad-CAM analysis, which limits both performance and interpretability — gaps

that are directly addressed in the current framework. Although Bhavani and Chalapathi [18] compared MobileNetV2 and InceptionV3, they did not explore selective layer fine-tuning, which the ablation study in the present work confirms is the single most impactful component, reducing accuracy from 98% to 84.67% upon its removal. Furthermore, while Alhammad et al. [19] combined transfer learning with explainable AI, their VGG-based models achieved only 91% accuracy and lacked stronger regularization techniques such as Batch Normalization and Dropout — both of which are incorporated in the proposed InceptionV3 framework to enhance generalization and robustness.

The CNN model performed poorly with a value of 76%, which showed that little discrimination could be made when looking at the images of the potato leaves and finding complex features. Both VGG16 and DenseNet121 performed well with scores above 86% and 90%, respectively; this indicates that transfer learning works very well with the images in agriculture. The MobileNetV2 showed great improvements from the previous models to have a performance of 91%, showing that it not only can classify well, but does so while being lightweight with low computational overhead.

The fine-tuned Xception model achieved a score of 93%, which makes it one of the strongest benchmarks when looking at other models. The proposed model based on InceptionV3 outperformed all other models with the highest value of 94%. Furthermore, when using optimal validation conditions, the InceptionV3 model had an accuracy of 98%, showing exceptional learning capabilities and generalization. From all aspects related to the performance of the proposed model, the scores of precision, recall, and F1 all equal 0.94, which indicates that there is strong reliability when identifying all the types of diseases present in the dataset.

The InceptionV3 model's test accuracy of 94% is greater than the 91% accuracy of VGG16 without fine-tuning, as demonstrated by Sholihati et al. [14]. This suggests that selective fine-tuning of layers is important for effective domain adaptation for agricultural image classification. Alhammad et al. [19] reported similar findings and also pointed out VGG architecture's limitations in performance. The current research uses Grad-CAM visualization to provide biological significance and interpretability, whereas Mishra et al. [16] provided no means of explaining their model predictions.

3.2 Class-wise Performance Analysis

The proposed InceptionV3 model was evaluated using precision, recall, and F1-score metrics to analyze class-wise performance across all potato leaf categories. The results demonstrate consistently strong classification performance across most classes, as shown in Fig. 4.

Among all categories, Early Blight and Potato Cyst Nematode achieved the highest performance, with F1-scores of 0.98, indicating that these classes contain more distinctive visual characteristics and are easier for the model to identify correctly. Late Blight also performed strongly, with an F1-score of 0.94.

Moderately lower results were observed for Healthy leaves, which obtained an F1-score of 0.90, and Plant Pests, which achieved 0.93. These slightly reduced scores may be caused by background variations, symptom overlap, or similarities in visual patterns. Fungal Diseases also achieved an F1-score of 0.93, suggesting minor confusion with visually similar disease symptoms.

Overall, the proposed InceptionV3 model demonstrated balanced and reliable performance across all classes, achieving a test accuracy of 94% and a macro-averaged F1-score of approximately 0.94. These results confirm the model's strong generalization capability for multiclass classification of potato leaf diseases.

3.3 Confusion Matrix Analysis

The proposed InceptionV3 model was further evaluated using both the confusion matrix and the normalized confusion matrix to examine its class-wise classification performance across all potato leaf categories. The confusion matrix shows the number of correctly and incorrectly classified samples, while the normalized confusion matrix presents the same results as percentages for clearer interpretation of class-level prediction accuracy.

As shown in Figure 4, strong diagonal dominance is observed, indicating that most test samples were correctly classified into their respective categories. The highest performance was achieved for Potato Cyst Nematode, where all samples were correctly classified (50/50). Similarly, Early Blight performed excellently with 49 correct predictions, while Late Blight achieved 47 correct predictions. The Fungal Diseases and Healthy classes each obtained 46 correct predictions, whereas Plant Pests achieved 45 correct classifications. Minor confusion was mainly observed among visually similar classes, including fungal diseases, healthy leaves, and plant pests.

The normalized confusion matrix further confirms these findings by presenting class-wise prediction accuracy as percentages, as shown in Fig. 6. Potato Cyst Nematode obtained a perfect normalized score of 1.00, while Early Blight reached 0.98, Late Blight achieved 0.94, and both Fungal Diseases and Healthy recorded 0.92. Plant Pests obtained a normalized score

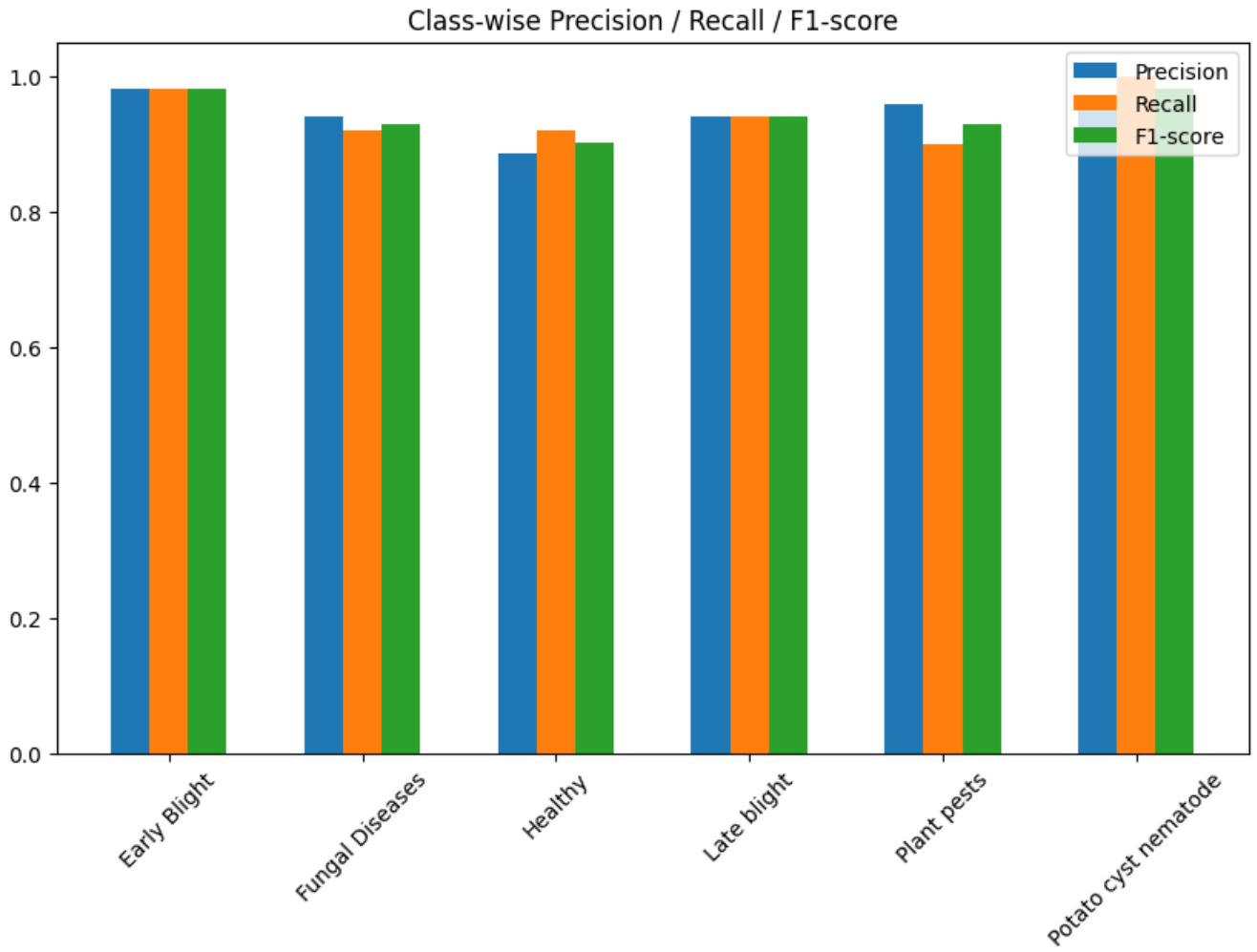


Fig 3. Class-wise precision, recall, and F1-score of the proposed InceptionV3 model.

of 0.90. Most classification errors were small in proportion (0.02–0.06), indicating balanced predictive behavior across all categories.

Overall, both matrices confirm that the proposed InceptionV3 model provides accurate, reliable, and well-balanced multiclass classification performance, supporting its suitability for practical potato leaf disease detection applications.

3.4 Ablation Study

An ablation study was conducted to evaluate the individual contributions of key components of the proposed InceptionV3 model, including data augmentation, dropout regularization, and fine-tuning. The results showed that the full model, with all components enabled, achieved the highest validation accuracy of 98%, confirming the effectiveness of the complete architecture.

When fine-tuning was removed, performance dropped significantly to 84.67%, demonstrating that adapting the pre-trained high-level features to the potato leaf disease dataset is the most critical factor for achieving optimal classification performance.

In comparison, removing dropout or augmentation produced only minor performance variations under the current experimental setting, suggesting that their direct effect on peak accuracy was smaller than that of fine-tuning. However, these components may still contribute to improved robustness, stability, and generalization during training.

Overall, the ablation results indicate that fine-tuning is the most influential component, while augmentation and regularization provide supportive benefits for building a more reliable model.

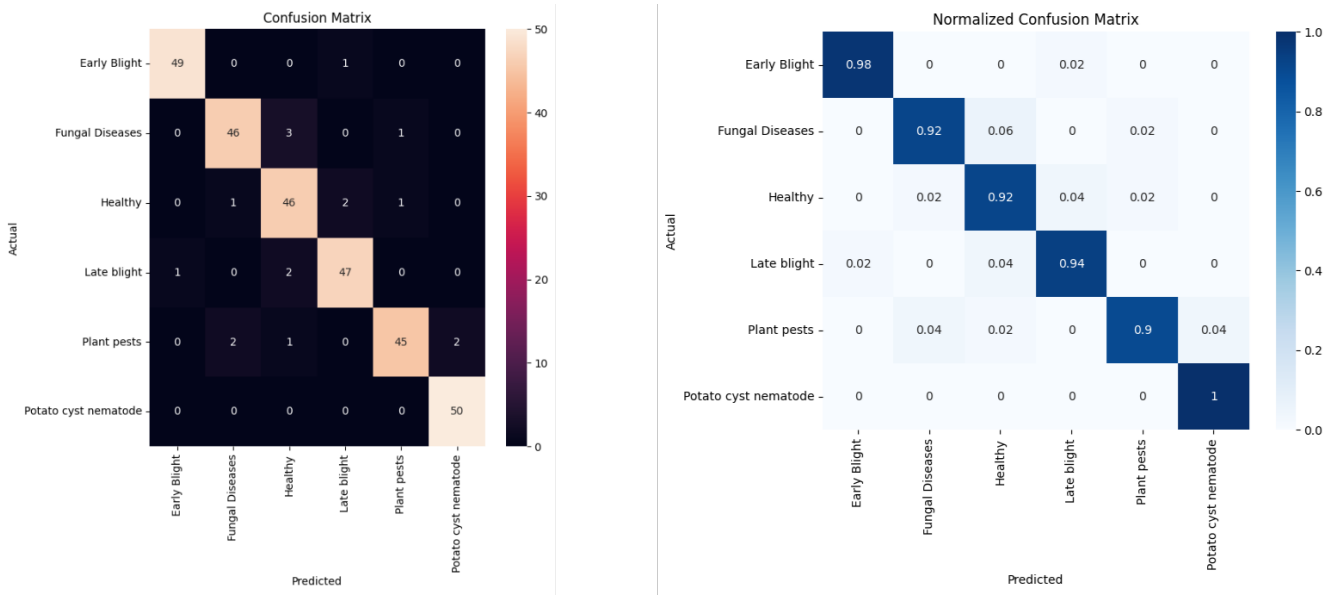


Fig 4. Confusion matrix and normalized confusion matrix of the proposed InceptionV3 model for potato leaf disease classification.

3.5 Explainability Analysis (Grad-CAM)

To enhance the interpretability of the proposed model, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to visualize the image regions that contributed most strongly to the classification decision. As illustrated in Figure 5, the generated heatmap highlights the symptomatic portion of the potato leaf, particularly the discolored and infected lesion area, while paying minimal attention to the surrounding background.

This observation indicates that the InceptionV3 model learned to focus on biologically relevant disease features rather than irrelevant patterns or image noise. The concentration of activation over the affected tissue suggests that the model successfully captured discriminative visual cues associated with the target disease class.

Furthermore, the Grad-CAM overlay provides visual evidence that the model's predictions are based on meaningful pathological characteristics present in the leaf image. Overall, this explainability analysis confirms that the proposed InceptionV3 framework learned robust and relevant representations for accurate classification of potato leaf diseases.

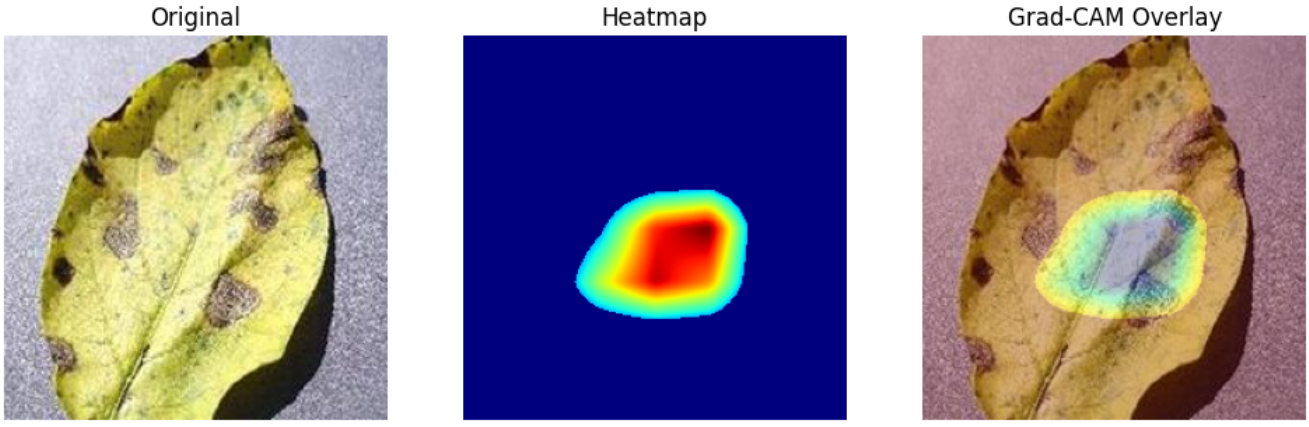


Fig 5. Grad-CAM visualization showing the model's attention map and the highlighted diseased region on a potato leaf image.

3.6 Qualitative Prediction Analysis

To further evaluate the classification behavior of the proposed InceptionV3 model, qualitative prediction examples from the test dataset were examined. As shown in Figure 6, representative samples from all six classes were correctly identified, where the predicted labels exactly matched the ground-truth categories. These results demonstrate that the model successfully learned discriminative visual features for recognizing both healthy and diseased potato leaves across multiple classes.

In contrast, Figure 7 presents representative misclassified samples in which the predicted labels differ from the actual classes. Most of these errors occurred between visually similar categories, such as Early Blight and Late Blight, or between mildly infected fungal leaves and healthy leaves. Such confusion is expected when symptoms are subtle, overlapping, or appear early in the course of infection. Variations in lighting conditions, leaf orientation, background texture, and image quality may also contribute to incorrect predictions.

Overall, the qualitative analysis confirms that the proposed model performs reliably on most test images while highlighting a small number of challenging cases that can guide future dataset enhancement and model improvement.

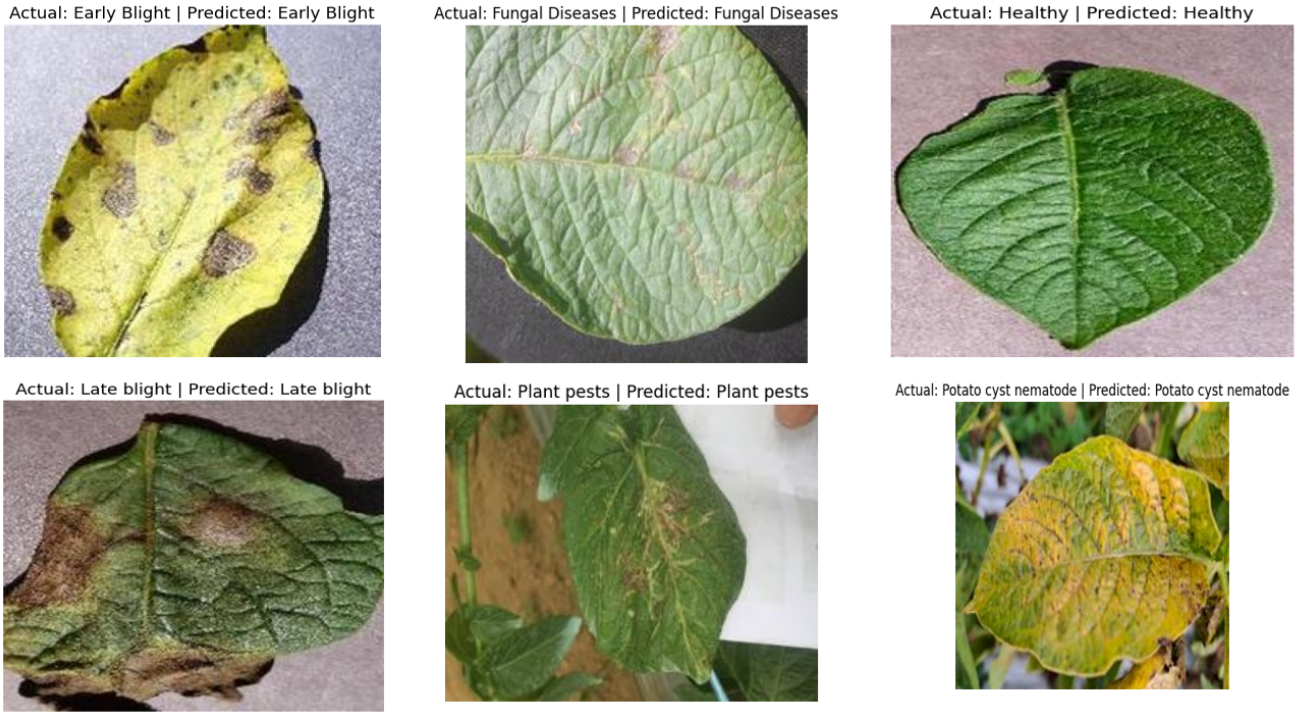


Fig 6. Representative correctly classified potato leaf images showing actual and predicted labels for all six classes

3.7 Computational Cost Analysis

The proposed InceptionV3 model contains approximately 22.3 million parameters, indicating a relatively high-capacity deep learning architecture. Based on the experimental setup, the total training time required was approximately 775 seconds on a CPU-based environment.

Although the model is computationally heavier than lightweight architectures, it achieved strong classification performance with high predictive accuracy and reliable generalization. This demonstrates a practical trade-off between computational cost and predictive performance. Therefore, the model is suitable for real-world deployment in applications where classification accuracy is prioritized over minimal resource usage.

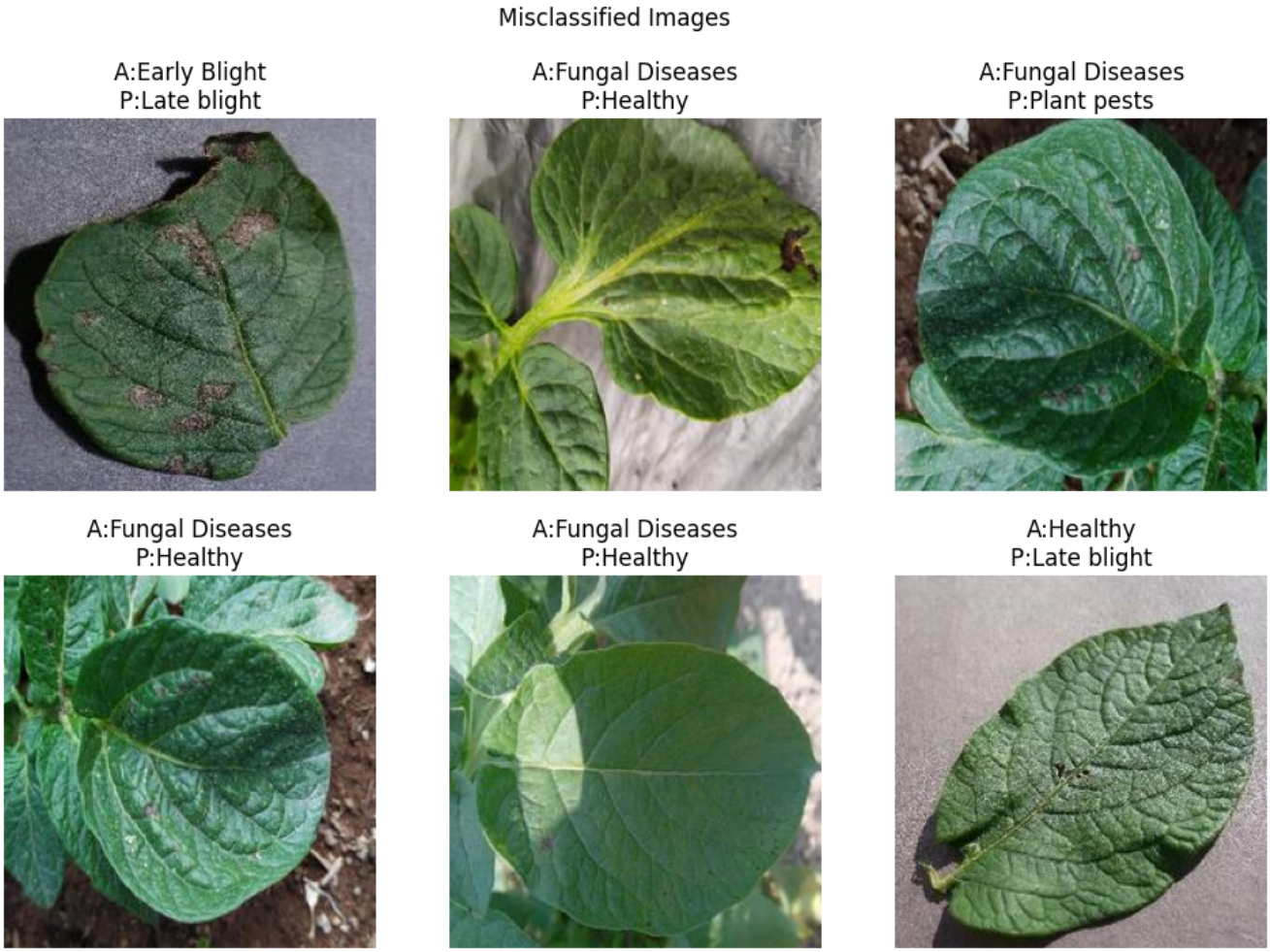


Fig 7. Representative misclassified potato leaf images showing actual (A) and predicted (P) labels

3.8 Training and ROC Analysis

The training and validation accuracy curves shown in Fig. 8 exhibit a steady upward trend throughout training, reaching approximately 98–99% training accuracy and 96–97% validation accuracy. The close alignment between the two curves indicates effective learning with only minor signs of overfitting, demonstrating strong generalization performance.

Furthermore, the Receiver Operating Characteristic (ROC) curves presented in Figure 8 demonstrate excellent discriminative capability across all six classes. The class-wise Area Under the Curve (AUC) values range from 0.99 to 1.00, indicating that the proposed model can distinguish between healthy and diseased potato leaf categories with very high confidence and accuracy.

4 Conclusion and Future Work

This research demonstrates how algorithms based on deep learning can distinguish diseased potato leaves simply by looking at a picture of the leaf. When the classifications were compared across all the various models tested, the best-performing model was based on transfer learning using a pre-trained model (InceptionV3) to extract relevant visual characteristics from the image of the leaf, and then finished tuning that model to perform specifically on potato leaves.

The accuracy achieved by the proposed framework during testing was 94%, with a corresponding macro-average F1-score of 0.94, meaning it outperformed all five of the alternative architectures being tested: CNN (76%), VGG16 (86%), DenseNet121 (90%), MobileNetV2 (91%), and Xception (93%).

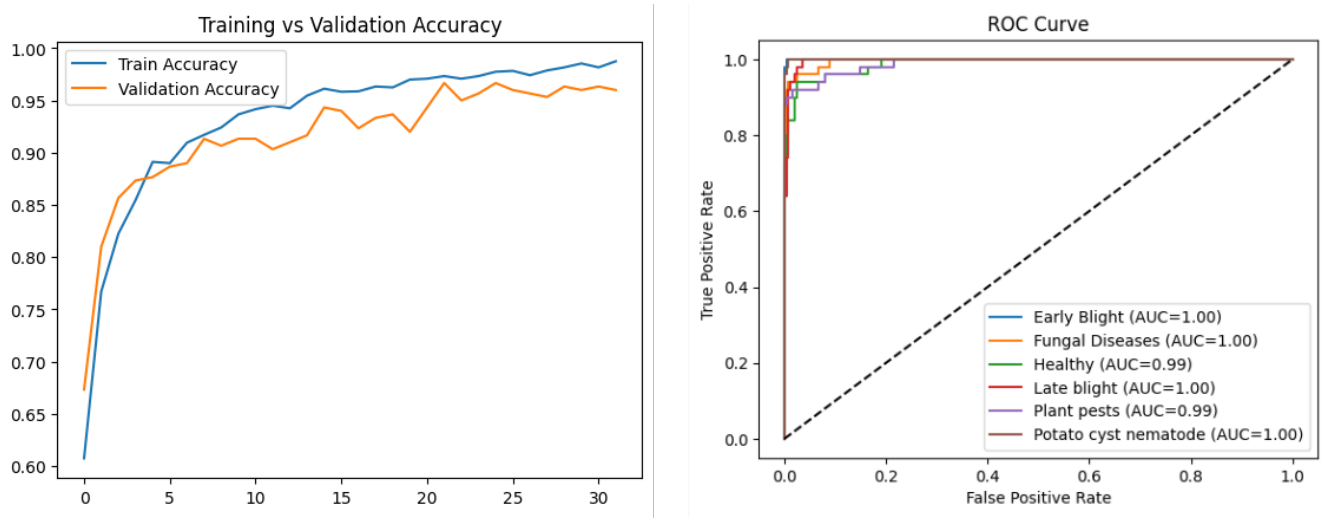


Fig 8. Training and validation accuracy curves and multi-class ROC curves with class-wise AUC scores for potato leaf disease classification using the proposed InceptionV3 model.

In addition to classification accuracy, other methods of evaluation were utilized to confirm the effectiveness of the model through multiple metrics (confusion matrix analysis, ROC curves, and Grad-CAM visual explanations). The resulting predictive performance of the model was consistent between all six classes, including the fact that the actual classification was based on biologically significant disease areas rather than background noise.

In contrast to previous studies[14],[19] that have reported maximum accuracies of only 91% without interpretability tools, this study incorporates an explanation of classification using Grad-CAM, resulting in improved clarity and reliability of the classification process.

An additional step taken in an ablation study was to assess the impact of removing the proposed framework's fine-tuned layers. Removal of these adapted Fine-Tuning layers resulted in a large decline in accuracy, from 98% to 84%, for certain agricultural image classifications across the three study farms. As a result, it further verified that fine-tuning is vital for providing contribution and added value to selective matching of layer adaptation to the specific domain(s) being used to improve the model's learning capabilities for that agricultural image classification.

From the error analysis performed on the misclassified images, it was determined that the majority of misclassifications for all three farms were due to visually similar classes (i.e., images of very similar diseases) or images of plants that exhibit only subtle/very minor symptoms. This indicates that, while there may still be errors in agricultural disease recognition in real-world settings, fine-tuning will be critical to improving model performance.

Overall, the results support the index as a viable, valid, and reliable framework, offering appropriate accuracy and interpretability for future automated diagnosis of potato leaf diseases.

Results from the developed InceptionV3 model showed that it performed well across all agricultural classification tasks but required significantly more computational resources than other lightweight architectures. Thus, additional work will be needed to modify the architecture of InceptionV3 to make it more computationally efficient for deployment in lower-resource environments, mobile devices, and/or edge devices in Precision Agriculture Applications.

Adding more diverse images of fieldwork, which are taken in the real world, will help us expand the database of images and increase their accuracy and generalizability. Real-time applications for agricultural monitoring can be achieved by developing mobile and cloud-based systems and using edge computing for real-time disease detection. Alternative techniques, such as attention mechanisms, ensemble learning, and multimodal data integration, will also improve the likelihood that the system will accurately, reliably, and effectively classify and deploy classifications from the dataset.

5 Acknowledgements

The research was not funded by any external funding agency.

Author details: Aradhy Tiwari & Chandrashekhar are Research Scholars; Amit Kumar Saxena serves as Professor.

References

1) FAOSTAT. FAOSTAT Database. 2026. Available from: https://www.fao.org/faostat/en/#data/QCL.
2) Rauf S, Khalil F, Ortiz R. Future Food: The Possible Impact of Potato Biofortification on Climate Resilience and Space Food. *Agriculture*. 2026;16(4):461. https://doi.org/10.3390/agriculture16040461.
3) Nawaz M, Javed A, Saudagar AKJ. PotatoGuardNet: a refined deep learning framework for potato leaf disease detection. *Frontiers in Plant Science*. 2026;17:1720276. https://doi.org/10.3389/fpls.2026.1720276.
4) Lamichhane S, Neupane S, Timsina S, Chapagain B, Paudel P, Rimal A. Potato Late Blight Caused by Phytophthora infestans; An Overview on Pathology, Integrated Disease Management Approaches, and Forecasting Models. *Plant Physiology and Soil Chemistry*. 2024;4:105–118. https://doi.org/10.26480/ppsc.02.2024.105.118.
5) Isinkaye FO, Olusanya MO, Singh PK. Deep learning and content-based filtering techniques for improving plant disease identification and treatment recommendations: A comprehensive review. *Heliyon*. 2024;10(9):e29583. https://doi.org/10.1016/j.heliyon.2024.e29583.
6) Grover P, Chaturvedi K, Zi X, Saxena A, Prakash S, Jan T, et al. Ensemble transfer learning for distinguishing cognitively normal and mild cognitive impairment patients using MRI. *Algorithms*. 2023;16(8):377. https://doi.org/10.3390/a16080377.
7) Patel D, Saxena AK, Laha S, Ansari GM. A novel scheme for feature selection using filter approach. *Proceedings of the 7th International Conference on Computing, Communication and Security (ICCCS)*. 2022;p. 1–4. https://doi.org/10.1109/ICCCS55188.2022.10079604.
8) Ghosh H, Rahat IS, Shaik K, Khasim S, Yesubabu M. Potato Leaf Disease Recognition and Prediction using Convolutional Neural Networks. *EAI Endorsed Transactions on Scalable Information Systems*. 2023;10(6). https://doi.org/10.4108/eetsis.3937.
9) Asif M, Abbas T, Iqbal S, et al. MapReduce-based deep learning framework for potato leaf disease detection in sustainable precision agriculture. *Scientific Reports*. 2026;16:1607. https://doi.org/10.1038/s41598-025-30940-3.
10) Paul H, Ghatak S, Chakraborty S, et al. A study and comparison of deep learning based potato leaf disease detection and classification techniques using explainable AI. *Multimedia Tools and Applications*. 2024;83:42485–42518. https://doi.org/10.1007/s11042-023-17235-3.
11) Pare S, Mittal H, Sajid M, Bansal JC, Saxena A, Jan T, et al. Remote sensing imagery segmentation: A hybrid approach. *Remote Sensing*. 2021;13(22):4604. https://doi.org/10.3390/rs13224604.
12) Chandrashekhar, Chandanan AK. Hybrid Quantum–Classical Machine Learning Framework for Village-Level Soil Environmental Sustainability and Land Degradation Risk Assessment in Bilaspur District, India. *International Journal for Multidisciplinary Research (IJFMR)*. 2026;8(2):1–21.
13) Tiwari A, Saxena AK, Patel D, Chandrashekhar. Evaluating Custom and Pre-trained CNN Architectures for Scalable Tomato Leaf Disease Classification. *International Journal of Computer Applications*. 2025;187(49):20–27.
14) Sholihati RA, Sulistijono IA, Risnumawan A, Sulistyo E. Potato leaf disease classification using deep learning approach. *Proceedings of the 2020 International Electronics Symposium (IES)*. 2020;p. 392–397. https://doi.org/10.1109/IES50839.2020.9231784.
15) Erlin E, Fuadi I, Putri RN, Nasien D, Gusrianty, Oktarina D. Deep learning approaches for potato leaf disease detection: Evaluating the efficacy of convolutional neural network architectures. *Revue d'Intelligence Artificielle*. 2024;38(2):236–245. https://doi.org/10.18280/ria.380236.
16) Mishra U, Pandey A, Logeswari G, Tamilarasi K. Deep learning-based disease detection in potato and mango leaves: A comparative study of CNN, AlexNet, ResNet, and EfficientNet. *Scientific Reports*. 2025;16:2788. https://doi.org/10.1038/s41598-025-32607-5.
17) Charisma RA, Adhinata FD. Transfer learning with DenseNet201 architecture model for potato leaf disease classification. *Proceedings of the 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*. 2023;p. 738–743. https://doi.org/10.1109/ICCoSITE57641.2023.10127772.
18) Bhavani GD, Chalapathi MMV. PotatoLeafNet: Two-stage convolutional neural networks for effective potato leaf disease identification and classification. *Frontiers in Artificial Intelligence*. 2026;8:1668839. https://doi.org/10.3389/frai.2025.1668839.
19) Alhammad SM, Khafaga DS, El-Hady WM, Samy FM, Hosny KM. Deep learning and explainable AI for classification of potato leaf diseases. *Frontiers in Artificial Intelligence*. 2025;7:1449329. https://doi.org/10.3389/frai.2024.1449329.