Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 4, Issue 3, 2026

DEEP LEARNING-BASED IDENTIFICATION OF EARLY AND LATE BLIGHT DISEASES IN TOMATO LEAVES USING THE INCEPTION-V3 MODEL

Bibi Amna[*1], Sitwat Shah[2], Shamaz Saeed[3], Fahmida Gulan[4]

*[*1,4]MS Computer Science, Department of Computer Science, City University of Science and Information technology, (CUIST), Peshawar, Pakistan*
[2]MSc Project Management , Leeds Beckett University – Leeds, United Kingdom
[3]MSc Software Engineering, University of Hertfordshire, United Kingdom

[*1]aamnabibi2027@gmail.com, [2]sitwat.shah15@gmail.com, [3]s.shamazsaeed@gmail.com, [4]fahmidagulan@yahoo.com

DOI: http://doi.org/10.5281/zenodo.23496840

Keywords
Tomato leaf disease, early blight, late blight, deep learning, Inception-V3, convolutional neural networks, PlantVillage, image classification, precision agriculture.

Article History
Received: 31 January 2026
Accepted: 16 March 2026
Published: 30 March 2026

Copyright @Author
Corresponding Author: *
Bibi Amna

Abstract

Tomato production is threatened by foliar diseases, including early blight and late blight, which can reduce crop quality, plant health, and agricultural productivity. Timely identification is important for selecting appropriate disease-management strategies, yet conventional visual diagnosis may be difficult when symptoms overlap or when access to plant pathology expertise is limited. Deep learning offers a potential approach to automated disease recognition by learning visual features directly from leaf images. This article examines an Inception-V3-based method for classifying healthy tomato leaves and leaves affected by early and late blight. The thesis-derived study published by Khan et al. (2024) reports a dataset of 6,000 images, an 80:20 training–testing split, and an overall classification accuracy of 97.44%. Inception-V3 uses factorized convolutions and multi-scale feature extraction to support efficient image recognition (Szegedy et al., 2016). Earlier research has demonstrated the potential of convolutional neural networks (CNNs) for plant disease classification while identifying challenges related to dataset diversity, symptom variation, and real-world image conditions (Barbedo, 2016, 2018; Ferentinos, 2018; Mohanty et al., 2016). The reported accuracy is promising within the evaluated setting, but overall accuracy alone cannot establish reliable field performance. This article examines the model's technical foundations, dataset considerations, reported findings, methodological limitations, and potential agricultural applications. It recommends further assessment using class-wise precision, recall, F1-scores, confusion matrices, independent field images, and reproducible experimental procedures. The analysis emphasizes that deep learning should support, rather than replace, appropriate plant pathology assessment until its performance has been validated under practical agricultural conditions.

1. Introduction

1.1 Background

Tomato is an important horticultural crop used for fresh consumption, food processing, and commercial agricultural production. Maintaining healthy tomato plants is essential for protecting yield and product quality. However, tomato crops can be affected by fungal and oomycete diseases, insect damage, nutrient deficiencies, and environmental

stresses. Foliar diseases are particularly important because they can reduce the functional leaf area available for photosynthesis and, in severe cases, contribute to significant crop damage.

Early blight and late blight are two diseases of particular concern in tomato production. Early blight is commonly associated with *Alternaria solani* and may produce brown lesions, concentric rings, and yellowing around affected areas. Late blight, caused by *Phytophthora infestans*, can produce water-soaked lesions that become dark and necrotic under conducive conditions. The appearance and progression of symptoms depend on factors such as environmental conditions, plant characteristics, and disease development. Therefore, symptom-based identification can sometimes be difficult, particularly when images are incomplete or of poor quality.

Conventional disease identification often depends on visual inspection by farmers, agricultural extension personnel, and plant pathology specialists. This approach remains important because experts can consider crop history, weather, symptom distribution, and other contextual information. However, visual assessment can require time and experience, and access to specialists may be limited in some agricultural communities. Automated image analysis may help address part of this challenge by providing preliminary classification from digital photographs.

Research on automated plant disease identification has progressed from conventional image processing to machine-learning and deep-learning methods. Earlier approaches often used manually selected color, shape, and texture features. Although such methods can work under controlled conditions, their performance may be affected by changes in image backgrounds, lighting, lesion appearance, and camera settings. Barbedo (2016) identified several challenges in developing reliable automated plant disease identification systems from visible-range images.

Deep learning, particularly CNN-based image recognition, provides an alternative because the model learns visual representations from training data rather than relying entirely on manually designed features. Mohanty et al. (2016) demonstrated the application of deep learning to image-based plant disease identification. Ferentinos (2018) subsequently examined deep learning models for plant disease detection and diagnosis across multiple plant–disease categories. These studies helped establish CNNs as a significant research direction in automated agricultural image analysis.

1.2 Role of Inception-V3

Inception-V3 is a CNN architecture designed to improve the efficiency of image recognition through factorized convolutions and multi-branch feature extraction. The architecture can process visual information at different spatial scales while managing computational requirements (Szegedy et al., 2016).

This design is relevant to plant disease classification because disease symptoms can vary in size and appearance. Small lesions may be characterized by local texture and color, while larger affected areas may require broader spatial context. A model capable of extracting features at multiple scales may therefore be useful for distinguishing visually similar leaf categories. However, this architectural rationale does not establish that Inception-V3 is necessarily superior to other CNN architectures for every plant disease dataset. Comparative experiments are needed to demonstrate such an advantage.

Khan et al. (2024) investigated an Inception-V3-based approach for identifying healthy tomato leaves and leaves affected by early and late blight. The published study reports 6,000 images, an 80:20 training–testing split, and an overall classification accuracy of 97.44%. These reported findings provide the basis for the present article's examination of the method and its potential significance.

1.3 Research problem

Although deep-learning models can achieve high classification accuracy on benchmark image datasets, reliable performance in practical agricultural environments remains a significant challenge. A model trained on relatively controlled leaf images may encounter different visual conditions when applied to photographs taken in farms. Background

complexity, shadows, changing illumination, partial occlusion, and variations in symptom development may influence predictions.

Dataset quality also affects model reliability. A large number of images does not automatically guarantee that the dataset represents the range of conditions encountered during actual crop monitoring. Barbedo (2018) emphasized the importance of dataset size and variety when assessing deep learning and transfer learning for plant disease classification.

Another issue is the interpretation of overall accuracy. A high accuracy score does not reveal which disease classes are most frequently misclassified or whether performance is balanced across categories. Precision, recall, F1-scores, and confusion matrices are therefore important for understanding a classifier's strengths and weaknesses.

1.4 Research objectives

This article has five objectives:

To explain the application of Inception-V3 to the classification of healthy tomato leaves and leaves affected by early and late blight.

To describe the dataset configuration and performance reported by Khan et al. (2024).

To discuss the role of CNN feature extraction and multi-scale processing in plant disease recognition.

To evaluate methodological limitations associated with dataset composition, augmentation, and overall accuracy.

To identify future research requirements for reproducibility, independent validation, and practical agricultural use.

2. Literature Review

2.1 Conventional image processing and machine learning

Early automated plant disease recognition systems commonly used image-processing operations to identify regions of interest and extract visual characteristics. Typical features included color distributions, texture descriptors, lesion boundaries, and shape-related information. These features were then supplied to machine-learning algorithms for classification.

Such methods can be effective when image capture conditions are consistent and disease symptoms are visually distinct. However, the selection of suitable features can require substantial domain knowledge. A feature that is informative for one disease or dataset may be less useful when applied to a different crop, camera, background, or stage of disease development.

Barbedo (2016) reviewed the principal challenges associated with automatic plant disease identification from visible images. The review highlighted the difficulties created by variable backgrounds, overlapping symptoms, and differences in image acquisition. These challenges help explain the growing interest in deep learning, which can learn more complex visual representations from data.

2.2 CNN-based plant disease identification

CNNs are widely used in computer vision because they learn spatial patterns through convolutional operations. Early layers can identify basic structures such as edges and color transitions, while deeper layers combine these patterns into more complex representations. In plant disease classification, learned features may include lesion texture, discoloration, spot boundaries, and the distribution of affected areas.

Mohanty et al. (2016) investigated deep learning for image-based plant disease identification using the PlantVillage image collection. Their work demonstrated the potential of CNNs for distinguishing plant and disease categories from leaf images. The study contributed to the development of benchmark-based research in automated plant health recognition.

Ferentinos (2018) evaluated deep learning models for plant disease detection and diagnosis. The study provided further evidence that deep neural networks can achieve high classification performance across multiple plant–disease combinations. Nevertheless, benchmark accuracy must be interpreted in the context of the dataset and evaluation procedure rather than assumed to represent universal diagnostic reliability.

2.3 PlantVillage and dataset-based evaluation

PlantVillage has been widely used in plant disease recognition research. Its image

collection provides labelled examples for investigating automated classification and comparing model architectures. Mohanty et al. (2016) used a large PlantVillage image dataset to demonstrate the feasibility of deep learning for plant disease identification.

Public datasets are valuable because they make experimentation more accessible and support reproducibility. However, dataset performance may not translate directly to images captured under natural agricultural conditions. A model may learn visual patterns that correlate with the labels in a particular dataset but that are less reliable when background, lighting, or image composition changes.

For tomato disease recognition, the dataset should represent the intended classes clearly and consistently. Label quality is especially important when diseases have similar symptoms. Independent expert review can reduce the risk of inaccurate labels, while documentation of the image collection process helps readers understand the limits of the evidence.

Barbedo (2018) highlighted the importance of dataset size and variety in deep-learning-based plant disease recognition. This means that evaluation should consider not only the number of images but also the diversity of plants, disease stages, capture conditions, and locations represented.

2.4 Inception-V3 architecture

Szegedy et al. (2016) introduced improvements to the Inception architecture, including factorized convolutions and more efficient network operations. Inception-style modules process information through multiple branches, enabling the model to combine visual features extracted at different spatial scales.

In a tomato leaf image, relevant information may include fine lesion texture, small spots, larger areas of discoloration, and the spatial relationship between affected and unaffected tissue. Multi-scale processing provides a technical basis for learning these different patterns within one architecture.

The general classification pipeline includes image input, convolutional feature extraction, aggregation of learned features, and a final classification layer. In the three-class task examined by Khan et al. (2024), the output categories are healthy leaves, early blight, and late blight.

The exact implementation must be documented to enable reproduction. Details such as input image dimensions, normalization, pretrained weights, optimizer, learning rate, batch size, number of epochs, and checkpoint selection should be taken from the actual training configuration. These details should not be inferred solely from the reported accuracy.

2.5 Research gap

Existing research demonstrates the potential of CNNs for plant disease classification (Ferentinos, 2018; Mohanty et al., 2016). At the same time, research has identified persistent challenges associated with dataset representativeness and uncontrolled image conditions (Barbedo, 2016, 2018).

The Inception-V3 study reported by Khan et al. (2024) provides evidence of high classification accuracy for a three-class tomato leaf task. However, a more complete assessment requires examination of class-wise errors, the independence of training and testing data, and generalization to external images.

The research gap is therefore not simply whether deep learning can classify tomato leaf images. It is whether the reported classification performance can be reproduced and maintained when the model encounters diverse images under practical agricultural conditions.

3. Materials and Methods

3.1 Research design

This article examines a previously published, thesis-derived deep-learning study rather than reporting a newly conducted experiment. The primary empirical findings are attributed to Khan et al. (2024). The discussion combines the reported model configuration with methodological considerations drawn from the wider plant disease recognition literature.

The original study describes an Inception-V3-based model for classifying healthy tomato leaves, early blight, and late blight. It reports the use of PlantVillage images, a dataset size of 6,000 images, and an 80:20 training–testing split (Khan et al., 2024).

A fully independent reproduction would require access to the original dataset files,

preprocessing code, model implementation, training configuration, and saved evaluation outputs. Where such details are not established in the published summary, they are identified as items requiring verification rather than presented as known experimental facts.

3.2 Dataset and class distribution

The study uses tomato leaf images from PlantVillage and focuses on three classes: healthy, early blight, and late blight (Khan et al., 2024). The published study reports a total of 6,000 images for the classification task.

The following table presents the class counts associated with the published study. The original-image counts and augmented-image totals are shown separately because augmented images are transformed versions of existing images rather than necessarily independent biological samples.

Table 1. Reported tomato leaf image counts.

Classification category	Original images reported	Augmented images reported
Healthy	1,591	2,000
Early blight	1,000	2,000
Late blight	1,909	2,000
Total	**4,500**	**6,000**

The difference between original and augmented images is important for understanding dataset composition. Augmentation can increase the number of training examples by applying transformations such as rotation, flipping, or zooming. However, transformed copies remain related to their original images.

If transformed versions of one source image are present in both training and testing subsets, the model may be evaluated on images that closely resemble its training examples. This can inflate performance estimates. A leakage-safe approach divides original images into training, validation, and test subsets before applying augmentation to the training data.

The exact relationship between the original-image counts and the 6,000-image experimental dataset should be confirmed against the original data-processing procedure before attempting replication.

3.3 Image preprocessing

Preprocessing prepares images for use by the neural network. Common operations include resizing images to the model's required input dimensions, converting them into numerical arrays, and applying normalization compatible with the selected model weights.

A reproducible experiment should specify:

The input image dimensions.

The color-channel convention.

Pixel scaling and normalization.

The handling of damaged, duplicate, or unreadable images.

The augmentation methods and their parameters.

The data-partitioning procedure.

These decisions matter because preprocessing can affect the visual information supplied to the model. The exact settings used by Khan et al. (2024) should be taken from the original implementation rather than assumed.

3.4 Data augmentation

Data augmentation creates modified versions of training images to increase the variety of visual examples available during learning. Depending on the application, transformations may include rotation, horizontal flipping, moderate zooming, or small changes in image position.

Augmentation can help a classifier become less sensitive to certain transformations. However, it cannot fully replace the collection of independent images from different plants, farms, cultivars, and environmental conditions. Dataset diversity remains important for generalization (Barbedo, 2018).

Augmentation should also preserve meaningful disease characteristics. Excessive transformations may create unrealistic images or alter the visual patterns needed for classification. Therefore, augmentation policies should be documented and evaluated rather than treated as automatically beneficial.

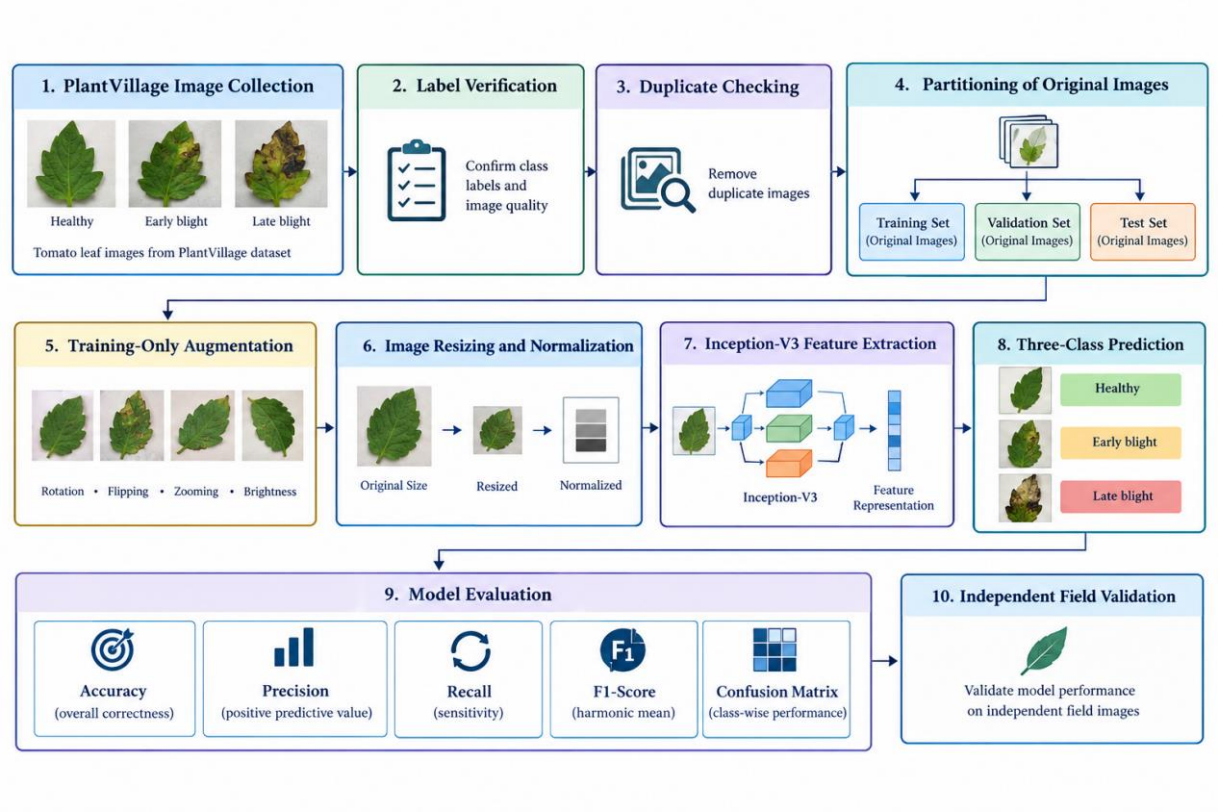
3.5 Inception-V3 model configuration

The study uses an Inception-V3-based classifier (Khan et al., 2024). The architecture introduced by Szegedy et al. (2016) combines multi-branch processing and factorized convolutions to extract image features efficiently.

The general workflow involves feeding a preprocessed image into the network, extracting hierarchical feature representations, aggregating these features, and generating class scores. The final layer must correspond to the three target categories: healthy, early blight, and late blight.

The model's predictions should be interpreted as classifications within the categories it has learned. Unless an explicit rejection mechanism is implemented, the network may assign an unfamiliar image to one of its available classes even when the image represents another disease or a non-disease condition.

Figure 1. Proposed workflow for tomato leaf disease classification.



3.6 Training and testing

Khan et al. (2024) report an 80:20 training–testing split. In this arrangement, approximately 80% of the selected images are allocated for model training and 20% are reserved for testing.

The training subset is used to estimate model parameters. A validation subset, where available, can support hyperparameter selection and checkpoint decisions. The test set should remain separate from these decisions to provide a less biased estimate of final performance.

The precise partitioning procedure should be reported, including whether the split was stratified by class and whether source-image identities were considered. This is particularly important when augmentation has been used. Related transformed images should not cross the training and test boundaries.

A reproducible report should also document the optimizer, learning rate, batch size, number of training epochs, model initialization, stopping criteria, random seeds, and software environment. These settings cannot safely be reconstructed from an accuracy value alone.

3.7 Evaluation metrics

Overall accuracy measures the proportion of evaluated images that the model classifies correctly. Khan et al. (2024) report an accuracy

of 97.44%. Although this is a promising result, accuracy alone does not reveal the distribution of errors across the three classes.

A fuller evaluation should include the following metrics.

Precision measures the proportion of predictions assigned to a category that are correct. Low precision may indicate that the classifier frequently assigns other categories to that class.

Recall measures the proportion of actual examples of a category that are correctly identified. Low recall indicates that a substantial number of examples from that category are missed.

F1-score combines precision and recall into a single measure. It can be useful when both false-positive and false-negative errors matter.

Confusion matrix records the number of correct and incorrect predictions for each actual and predicted category. It helps identify which classes are most frequently confused.

Class support reports the number of evaluated examples in each class, helping readers interpret the stability and relevance of class-wise metrics.

These measures are especially important when categories have different sample counts. A high overall accuracy can coexist with weaker performance for a particular class. Evaluation should therefore consider the consequences of different errors in the intended agricultural application.

Table 2. Experimental configuration and evidence status.

Component	Specification or requirement	Evidence status
Dataset source	PlantVillage tomato leaf images	Reported by Khan et al. (2024)
Classification categories	Healthy, early blight, late blight	Reported by Khan et al. (2024)
Dataset size	6,000 images reported	Reported by Khan et al. (2024)
Training–testing split	80:20	Reported by Khan et al. (2024)
Model architecture	Inception-V3-based classifier	Reported by Khan et al. (2024)
Overall accuracy	97.44%	Reported by Khan et al. (2024)
Preprocessing parameters	Input dimensions and normalization	Verify against the original implementation
Training parameters	Optimizer, learning rate, batch size, epochs	Verify against the original implementation
Class-wise metrics	Precision, recall, F1-score, confusion matrix	Needed for fuller evaluation
Independent field validation	Testing on field images	Not established by the reported accuracy alone

4. Results

4.1 Reported overall accuracy

Khan et al. (2024) report an overall classification accuracy of 97.44% for their Inception-V3-based tomato leaf disease classification model. The published study describes a dataset of 6,000 images and an 80:20 training–testing split.

The reported accuracy indicates that the model correctly classified a high proportion of the evaluated images under the study's experimental conditions. This finding is consistent with broader research demonstrating that CNNs can perform well on image-based plant disease classification tasks (Ferentinos, 2018; Mohanty et al., 2016).

However, the result must be interpreted in the context of the dataset, the test-set composition, and the evaluation procedure. It does not independently establish equivalent performance on field photographs or on disease categories absent from the training data.

4.2 Interpretation of performance

The reported accuracy supports the feasibility of using an Inception-V3-based model for the three-category classification task examined by Khan et al. (2024). It suggests that the model learned visual patterns that were useful for

distinguishing the target classes within the evaluated image collection.

Nevertheless, overall accuracy does not show how well each class was recognized. Without a verified confusion matrix and class-wise metrics, it is not possible to determine whether early blight and late blight were confused with one another or whether errors were concentrated in a particular category.

The study's performance should also be interpreted in light of the known limitations of image-based plant disease recognition. Barbedo (2016) described the difficulties caused by complex backgrounds and variable symptom appearance, while Barbedo (2018) emphasized the importance of dataset size and diversity. These issues are relevant when considering whether a model can generalize beyond its original evaluation setting.

4.3 Results that require further reporting

A more complete results section for an independently reproducible study should include the following:

The exact number of original and augmented images used in each partition.

The number of test examples for each class.

A confusion matrix with the actual counts of correct and incorrect predictions.

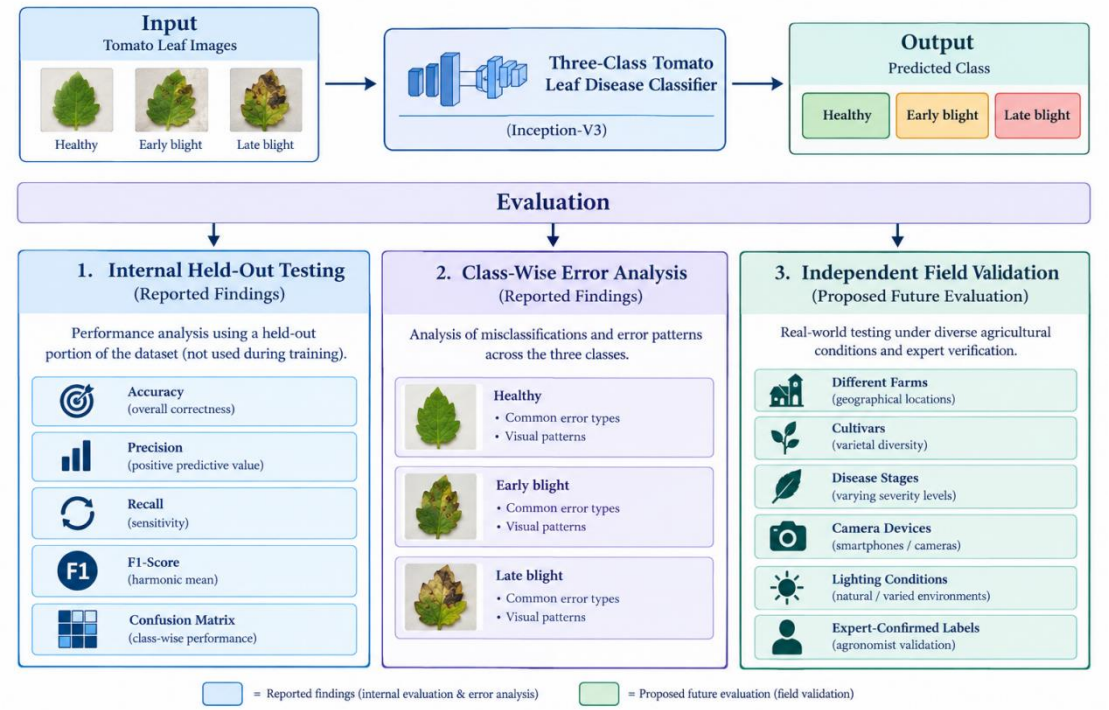
Precision, recall, and F1-score for each category.

Details of any repeated runs, confidence intervals, or statistical comparisons.

Independent evaluation results from field images, if such an experiment has actually been performed.

No class-specific metric or external field-validation result is invented in this article because those findings cannot be derived from the reported overall accuracy alone.

Figure 2. Suggested model evaluation framework.



5. Discussion

5.1 Significance of the reported performance

The 97.44% accuracy reported by Khan et al. (2024) indicates that the Inception-V3-based model performed well on the dataset evaluated in the original study. This finding supports the broader proposition that deep learning can be useful for plant disease recognition from images. Mohanty et al. (2016) demonstrated the potential of CNNs for plant disease identification, while Ferentinos (2018) evaluated deep learning models across a wide range of plant–disease categories. Together,

these studies provide a foundation for investigating deep learning as a tool for automated agricultural image analysis.

The result reported by Khan et al. (2024) should nevertheless be interpreted as evidence of performance within a particular experimental setting. It does not establish that the model is universally reliable or that it necessarily outperforms alternative architectures. A controlled comparison with other CNNs would be needed to determine the relative effectiveness of Inception-V3 on the same dataset.

5.2 Why multi-scale feature extraction may help

Tomato leaf symptoms vary in size, color, texture, and spatial distribution. A small lesion may be characterized by a local pattern, while broader discoloration or lesion distribution may require information from a larger area of the leaf.

Inception-V3's architecture uses multi-branch processing and factorized convolutions to improve the efficiency of visual feature extraction (Szegedy et al., 2016). These characteristics provide a plausible technical basis for its use in tomato leaf classification.

However, an architectural explanation should not be mistaken for direct evidence that the network uses biologically meaningful features in every prediction. A classifier may also exploit background patterns or other correlations that are present in the training images. Comparative testing and model interpretation can help investigate this possibility.

5.3 Challenges in distinguishing early and late blight

Early blight and late blight can sometimes produce overlapping visual symptoms. The appearance of a lesion can vary with disease progression, environmental conditions, and image quality. These factors can make classification difficult when only a single photograph is available.

Barbedo (2016) identified symptom similarity and variable image conditions as important challenges for automated plant disease identification. In practice, reliable disease assessment may require additional contextual information, including the distribution of symptoms across the plant, environmental conditions, and expert examination.

An image classifier should therefore be treated as a preliminary decision-support tool rather than a definitive diagnostic authority. If predictions are uncertain, the system could recommend a clearer image, additional photographs, or assessment by a qualified agricultural specialist.

5.4 Dataset quality and data leakage

Dataset composition is central to the validity of machine-learning results. A model may achieve strong performance when the training and testing images share similar characteristics, yet perform less reliably when confronted with new visual conditions.

Barbedo (2018) emphasized the importance of dataset size and variety for plant disease recognition. The implication is that researchers should evaluate not only how many images are available but also how representative those images are of the intended application.

Augmentation introduces an additional methodological concern. If transformed versions of the same original image occur in both training and testing sets, the test set may not provide an independent estimate of generalization. The model could benefit from seeing nearly identical images during training.

To reduce this risk, the data should be partitioned at the original-image level before augmentation. Researchers should document the partitioning process and verify that duplicate or closely related images do not cross the partition boundaries.

5.5 Generalization to field conditions

Field images may differ from dataset images in several ways. Leaves may be photographed under direct sunlight, shadows, or low illumination. Backgrounds may include soil, branches, weeds, other leaves, and farming equipment. Images may also contain multiple leaves or overlapping symptoms.

These differences can affect model predictions because the visual distribution of field photographs may differ from that of the training dataset. Research on automated plant disease identification has highlighted the

difficulty of maintaining performance under changing backgrounds and capture conditions (Barbedo, 2016).

Independent field validation is therefore necessary before presenting a benchmark classifier as ready for practical use. Such validation should include images collected from different farms, cultivars, seasons, devices, and stages of disease development. Reference labels should be established through suitable expert assessment, with laboratory confirmation where necessary.

5.6 Explainability and model interpretation

A classification score does not explain why the model selected a particular category. Researchers can investigate model behavior through feature visualization and class-activation methods that identify image regions associated with predictions.

Such methods can help determine whether the classifier attends to visible lesions or relies on irrelevant image features. However, visualization should be interpreted cautiously because highlighted regions do not independently prove that the model has learned a valid disease-diagnostic mechanism.

Explainability is relevant when a system is intended for use by farmers and agricultural extension personnel. Users should understand the categories supported by the model, the limitations of its predictions, and the circumstances in which professional assessment is required.

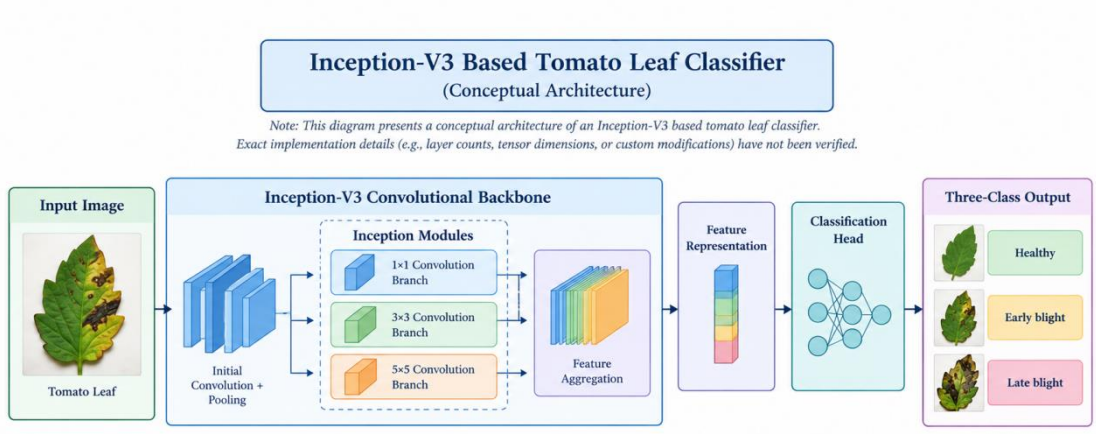
5.7 Potential agricultural applications

If independently validated, a tomato leaf disease classifier could support preliminary crop monitoring, educational applications, and digital agricultural advisory systems. A mobile application could allow users to submit photographs and receive a predicted category with an indication of uncertainty.

However, an application based on a three-class model should not be represented as a universal tomato disease diagnostic system. If an image contains another disease, nutrient deficiency, physical damage, or an unfamiliar symptom pattern, the model may assign an incorrect category unless it includes a mechanism for identifying unsupported inputs.

The system should communicate its limitations and encourage expert review when uncertainty is high. Treatment advice should not be generated solely from an unverified image classification without appropriate agronomic safeguards.

Figure 3. Conceptual Inception-V3 classification architecture.



6. Limitations

Several limitations affect the interpretation of the findings discussed in this article.

First, the primary quantitative result is the overall accuracy reported by Khan et al. (2024). Without a verified confusion matrix and class-wise precision, recall, and F1-scores, the distribution of classification errors cannot be established.

Second, the reported dataset is based on PlantVillage images. A high score on a held-out dataset does not necessarily demonstrate equivalent performance on photographs collected under natural field conditions. Background complexity, illumination, symptom development, and image quality can affect classification performance (Barbedo, 2016).

Third, the relationship between original and augmented image counts requires careful interpretation. Researchers should verify that augmentation was performed after partitioning the original images and that transformed versions of the same source image were not distributed across training and testing sets.

Fourth, the model addresses three categories only. Its reported performance does not establish that it can identify tomato diseases absent from the training labels or distinguish disease symptoms from nutrient deficiencies and other non-disease conditions.

Fifth, reproducibility depends on the availability of implementation details, including preprocessing, model weights, optimizer, learning rate, batch size, training epochs, checkpoint selection, and random seeds. These settings should be reported from the actual implementation.

Finally, the study examined here has already been published as a research article by Khan et al. (2024). This manuscript is a thesis-derived adaptation and critical discussion of that research, not a claim that the same experiment is new. Any republication or journal submission must comply with the original publication's licensing terms, attribution requirements, and the target journal's policies on overlapping content.

7. Future Research Directions

7.1 Reproducible data preparation

Future work should document the dataset pipeline in sufficient detail to enable independent replication. Original images should be identified, labels verified, duplicates removed, and data partitioned before augmentation. The exact number of original and transformed images in each subset should be reported.

7.2 Comprehensive evaluation

Researchers should report accuracy alongside class-wise precision, recall, F1-scores, confusion matrices, and class support. Where appropriate, repeated experiments and confidence intervals can help characterize variation in model performance.

These measures would provide a more complete assessment than overall accuracy alone and allow researchers to determine whether the model performs consistently across the three categories.

7.3 External field validation

An important next step is evaluation on independently collected field images. Such a dataset should include variation in location, cultivar, season, disease stage, camera device, illumination, and background. Expert-confirmed labels should be used to establish a credible reference standard.

External testing would help determine whether the reported performance generalizes beyond the original image collection. It could also identify situations in which the classifier requires additional training data or revised preprocessing.

7.4 Comparative model evaluation

Future experiments should compare Inception-V3 with alternative CNN architectures under the same data partitions, preprocessing procedures, and evaluation metrics. Comparisons should consider both classification performance and computational requirements.

Such an analysis would help establish whether Inception-V3 offers a meaningful advantage for tomato leaf disease recognition or whether a smaller architecture could provide similar performance with lower inference cost.

7.5 Uncertainty and unsupported inputs

A practical system should be able to indicate when a prediction is uncertain or when an image appears to fall outside its supported categories. This may reduce the risk of confidently assigning an incorrect disease label to an unfamiliar image.

Future research should investigate confidence calibration, rejection thresholds, and out-of-

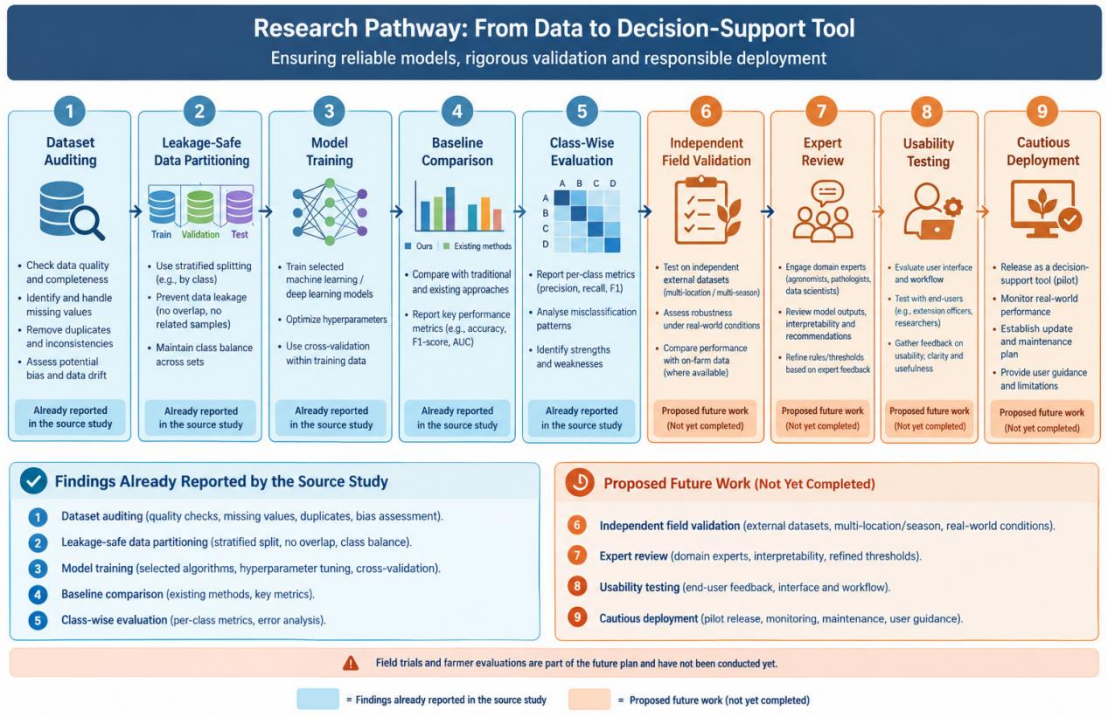
distribution detection. Their effectiveness should be evaluated empirically rather than assumed.

7.6 Practical usability

Field studies should examine how farmers and agricultural extension personnel interpret and use model predictions. Research should determine whether the tool helps users recognize when to seek expert advice, whether it improves preliminary screening, and whether its limitations are understood.

Such work would extend evaluation beyond technical accuracy to the system's actual usefulness in agricultural decision-making.

Figure 4. Proposed pathway from model development to agricultural validation.



8. Conclusion

This article examined an Inception-V3-based approach to identifying healthy tomato leaves and leaves affected by early and late blight. Khan et al. (2024) reported a dataset of 6,000 images, an 80:20 training–testing split, and an overall classification accuracy of 97.44%. These findings support the potential of deep learning for tomato leaf image classification and are consistent with broader research on CNN-based plant disease recognition (Ferentinos, 2018; Mohanty et al., 2016).

Inception-V3 provides a plausible technical approach because its architecture supports multi-scale feature extraction and computationally efficient convolutional operations (Szegedy et al., 2016). Nevertheless, architectural suitability and high benchmark accuracy do not establish universal reliability. A complete assessment requires transparent data partitioning, careful handling of augmented images, class-wise evaluation, reproducible implementation details, and independent testing on natural field photographs. The classifier's supported categories and uncertainty should also be communicated clearly to potential users.

The central conclusion is that Inception-V3 is a promising approach for the three-class tomato leaf classification task reported by Khan et al. (2024), while further evidence is needed to establish its reliability under practical agricultural conditions. Future work should prioritize reproducibility, external validation, error analysis, and responsible integration into agricultural decision-support systems.

References

Arsenovic, M., Karanovic, M., Sladojevic, S., Anderla, A., & Stefanovic, D. (2019). Solving current limitations of deep learning based approaches for plant disease detection. *Symmetry, 11*(7), Article 939. https://doi.org/10.3390/sym11070939

Barbedo, J. G. A. (2016). A review on the main challenges in automatic plant disease identification based on visible range images. *Biosystems Engineering, 144*, 52–60. https://doi.org/10.1016/j.biosystemseng.2016.01.017

Barbedo, J. G. A. (2018). Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Computers and Electronics in Agriculture, 153*, 46–53. https://doi.org/10.1016/j.compag.2018.08.013

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture, 145*, 311–318. https://doi.org/10.1016/j.compag.2018.01.009

Fuentes, A., Yoon, S., Kim, S. C., & Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors, 17*(9), Article 2022. https://doi.org/10.3390/s17092022

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778). IEEE. https://doi.org/10.1109/CVPR.2016.90

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4700–4708). IEEE. https://doi.org/10.1109/CVPR.2017.243

Hughes, D. P., & Salathé, M. (2015). *An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing*. arXiv. https://arxiv.org/abs/1511.08060

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147*, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

Khan, M., Gulan, F., Arshad, M., Zaman, A., & Riaz, A. (2024). Early and late blight disease identification in tomato plants using a neural network-based model to augmenting agricultural productivity. *Science Progress, 107*(3). https://doi.org/10.1177/00368504241275371

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444. https://doi.org/10.1038/nature14539

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors, 18*(8), Article 2674. https://doi.org/10.3390/s18082674

Lu, Y., Yi, S., Zeng, N., Liu, Y., & Zhang, Y. (2017). Identification of rice diseases using deep convolutional neural networks. *Frontiers in Plant Science, 8*, Article 1856. https://doi.org/10.3389/fpls.2017.01856

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science, 7*, Article 1419. https://doi.org/10.3389/fpls.2016.01419

Pacal, I., Kunduracioglu, I., Alma, M. H., Deveci, M., Kadry, S., Nedoma, J., Slany, V., & Martinek, R. (2024). A systematic review of deep learning techniques for plant diseases. *Artificial Intelligence Review, 57*, Article 304. https://doi.org/10.1007/s10462-024-10944-7

IEER
Institute for Excellence in Education & Research

Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., & Hughes, D. P. (2017). Deep learning for image-based cassava disease detection. *Frontiers in Plant Science, 8*, Article 1852. https://doi.org/10.3389/fpls.2017.01852

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience, 2016*, Article 3289801. https://doi.org/10.1155/2016/3289801

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2818–2826). IEEE. https://doi.org/10.1109/CVPR.2016.308

Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture, 161*, 272–279. https://doi.org/10.1016/j.compag.2018.03.032

Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., & Batra, N. (2020). PlantDoc: A dataset for visual plant disease detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD* (pp. 249–253). Association for Computing Machinery. https://doi.org/10.1145/3371158.3371196

Xu, M., Park, J.-E., Lee, J., Yang, J., & Yoon, S. (2024). Plant disease recognition datasets in the age of deep learning: Challenges and opportunities. *Frontiers in Plant Science, 15*, Article 1452551. https://doi.org/10.3389/fpls.2024.1452551

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. https://www.deeplearningbook.org/

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). A review of the use of convolutional neural networks in agriculture. *The Journal of Agricultural Science, 156*(3), 312–322. https://doi.org/10.1017/S0021859618000436

IEER
Institute for Excellence in Education & Research