Enhancing Tomato Leaf Disease Classification Integrating DenseNet201 and InceptionV3 Models with Fine Tuning and Feature Fusion

Abhiram Sharma[1[0009-0004-0822-0603]] , R Srivats[1[0009-0001-5791-4255]],

Dr. S Abirami and Dr. T Nathezhtha

[1] School of Computer Science and Engineering, Vellore Institute of Technology, Chennai Campus, Tamilnadu, India
abirami.s@vit.ac.in

Abstract. Accurate and timely classification of tomato leaf diseases are crucial for effective disease management and food security. Traditional methods relying on expert knowledge and manual inspection are time-consuming, subjective, and often inaccurate. Advanced models achieve high accuracy but are limited by high computational demands and poor generalizability across crops and diseases. Furthermore, dependence on sophisticated hardware can limit practical application in less accessible regions. This paper bridges these gaps by developing a highly efficient and precise hybrid model for classification of several diseases in tomato plants. The proposed model integrates DenseNet201 and InceptionV3, utilizing feature fusion and fine tuning to enhance performance. The experimental setup involves a comprehensive dataset of over 32,000 classified images of tomato leaves, augmented to increase size and variability. The hybrid model excelled through feature fusion and fine-tuning, achieving a high accuracy of 98.5%, precision of 97.3%, recall of 97.8%, and an F1 score of 97.6%. The study highlights the hybrid model's potential for practical agricultural use, offering an adaptable solution for farmers and researchers.

Keywords: Disease classification, InceptionV3, Feature Fusion, DenseNet201, Inception, Model comparison, Transfer learning, Fine tuning

1 Introduction

Tomato plants hold a prominent place in global agriculture, being a vital source of nutrients and economic value. Healthy tomato plants have vibrant green foliage, robust stems, and abundant flowering and fruit set. Leaves are spotless and fruits are blemish free. However, they are highly susceptible to a plethora of diseases that can significantly reduce yield and quality. The accurate and timely classification of these diseases is essential for effective management and ensuring food security. Traditionally, plant disease identification has relied on expert knowledge and manual inspection, which, although valuable, is often time consuming, subjective, and highly prone to inaccuracies.

Machine learning and deep learning have revolutionized disease identification, providing rapid, accurate, and consistent solutions for early detection and intervention, crucial for protecting crop yield. In particular, convolutional neural networks - CNNs like DenseNet and Inception have shown exceptional performance in image classification tasks, making them suitable for diagnosing plant diseases from leaf images. This section details ten major tomato leaf diseases, including their symptoms, causes, impact with an image of all the diseases provided above.

Late blight, caused by *Phytophthora infestans*, manifests as dark lesions on leaves, stems, and fruits, spreading rapidly in cool, wet conditions. It propagates through wind, rain, and contaminated equipment, with white mold on leaf undersides in humidity. Plants infected with **Tomato Yellow Leaf Curl Virus** exhibit yellowing and curling leaves, stunted growth, and decreased fruit production, with flowers often falling before fruit develops. Spread by the whitefly (*Bemisia tabaci)*, the virus remains in the vector for life and rapidly transfers between plants. These two highly destructive diseases lead to significant economic losses due to reduced yield and fruit quality.

Septoria leaf spot, caused by *Septoria lycopersici*, starts with small, water-soaked spots on lower leaves that merge, causing yellowing and defoliation. It thrives in warm, moist conditions and spreads via water splashes, contaminated tools, and plant debris, reducing yield and exposing fruits to sunscald and secondary infections. **Early blight**, caused by the fungus Alternaria solani, produces brown to black "target spots" on leaves, stems, and fruits, with older leaves affected first, showing V-shaped yellowing. It survives in soil and plant debris, spreading via wind, rain, and irrigation water.

Spider mite infestations cause stippling, yellowing, and premature leaf drop, with fine webbing on leaf undersides. The two-spotted spider mite (Tetranychus urticae) thrives in hot, dry conditions and reproduces rapidly. Infestations reduce photosynthesis, stunting growth and lowering yields, and severe cases can defoliate plants and kill seedlings. **Powdery Mildew** is a white, powdery fungal growth appears on upper leaf surfaces, stems, and flowers, causing leaves to curl, yellow, and drop prematurely. Primarily caused by *Leveillula taurica*, it spreads through wind-borne spores. This disease reduces photosynthesis, vigor, and yield, with severe infections leading to significant defoliation and reduced fruit quality.

Bacterial Spot (*Xanthomonas campestris pv. vesicatoria*) causes small water-soaked spots on leaves that enlarge and become necrotic with a chlorotic halo, affecting stems and fruits. **Target Spot** (*Corynespora cassiicola*) results in small, dark lesions with a yellow halo on leaves, stems, and fruits, which can merge into extensive necrosis.

These two diseases spread through wind, rain, and contaminated equipment, decreasing fruit production and ripening, causing substantial yield losses.

Plants infected with **Tomato Mosaic Virus** show mosaic mottling, chlorosis, leaf distortion, stunted growth, and reduced fruit production, with fruits developing necrotic spots and uneven ripening. It spreads through handling, contaminated tools, and insects, persisting in plant debris and seeds. Yellow spots on the upper leaves and velvety olive-green mold on the undersides are caused by the fungus *Passalora fulva*. Leaves infected with **Leaf Mold** may curl, wither, and drop prematurely. Thriving in high humidity and poor ventilation, it spreads via airborne spores, reducing photosynthesis and plant vigor. Proper greenhouse management techniques and resistant cultivars are essential for control.

Understanding the symptoms, causes, and impacts of these diseases is essential for effective management and control. Through diligent monitoring and appropriate interventions, the adverse effects of these diseases on tomato crops can be minimized, ensuring healthy plants and optimal yields.

Although advancements have been made, several significant gaps remain in current research. Many studies have achieved high accuracy rates using advanced models such as DenseNet, ResNet, and Inception [2][3]. However, these models often demand substantial computational resources, which may not be accessible in all agricultural settings, particularly in developing regions [4]. Furthermore, most studies focus on a limited number of diseases or specific crops, resulting in models that lack generalizability across different plant species and disease types. Recent studies have also explored the use of UAV technology and transfer learning to enhance disease detection capabilities [5][6]. However, the dependency on UAV technology and sophisticated hardware can limit practical application in less accessible regions. Moreover, there is a growing need for models capable of handling and adapting toward robust disease classification [7][8].The primary objectives of this research are multi-faceted. Our primary goal is to create a highly efficient and precise model for classifying ten tomato plant diseases and distinguishing healthy plants. Secondly, we will evaluate the performance of DenseNet201 and InceptionV3 models to identify their respective strengths and limitations for this classification task. Next, we will merge the strengths of these models through feature extraction and fusion to create a hybrid model that addresses their limitations and boosts performance. Further, we will fine-tune the hyperparameters of this integrated model to enhance its classification accuracy and generalizability. Finally, we will validate the model's effectiveness using a comprehensive dataset of tomato leaf images. This research makes several important contributions to the field of tomato plant disease classification. They are as follows:

- A significant gap in the existing literature is addressed by providing a comprehensive comparison of several models, including DenseNet, ResNet, VGG, and Inception. By evaluating these models against various metrics, their strengths and limitations are highlighted specifically in the context of tomato leaf disease classification.
- A novel approach is introduced by combining DenseNet201 and InceptionV3 through transfer feature fusion and fine-tuning their hyperparameters. This hybrid model demonstrates enhanced performance and adaptability, showcasing the potential benefits of integrating advanced deep learning techniques for agricultural applications.
- To employ rigorous validation methods to ensure the robustness and reliability of our model. This thorough assessment confirms the model's accuracy and effectiveness in real-world scenarios.

Finally, discuss the practical implications of our research, emphasizing the model's potential for deployment in real life agricultural settings. The findings provide valuable insights for farmers and researchers, indicating that this approach could lead to substantial improvements in disease management and crop yield.

2 Literature Review

Plant disease identification and management are crucial for ensuring global food security and agricultural sustainability. This literature survey provides a comprehensive review and critical analysis of recent works in plant disease identification using deep learning methods. By examining a wide range of studies, this survey aims to elucidate the current landscape of research in this domain and identify key challenges and opportunities for further exploration.

Arsenovic et al. achieved notable results in deep learning-based segmentation and classification of plant diseases, underscoring the potential of deep learning for precise disease identification [1]. Brahimi et al. demonstrated effective leaf disease classification using deep neural networks, emphasizing the importance of deep learning in achieving high classification accuracy [2]. Khalid and Karan explored deep learning for plant disease detection, showcasing the benefits of convolutional neural networks and MobileNet architectures in improving early disease detection [12]. Ferentinos focused on the automatic detection of tomato diseases using deep learning, highlighting the advantages of automated systems in large-scale agricultural monitoring [4]. Huang et al. utilized YOLOv5n for maize leaf disease identification, showcasing the efficiency of this model in achieving high accuracy with optimized computational requirements [5]. Koklu et al. achieved a 97.6% success rate in classifying grape leaf diseases using deep learning, emphasizing the significance of feature extraction and reduction techniques [6].

Liu et al. explored plant disease detection using deep learning and data augmentation, demonstrating the potential of augmented datasets to enhance model performance [7]. Mohanty et al. investigated image-based plant disease diagnosis using deep learning, confirming the practicality and high accuracy of these methods in various agricultural contexts [8]. Ramcharan et al. applied MobileNetV2 for tomato plant disease detection, achieving high accuracy rates and highlighting the efficiency of lightweight models in resource-constrained environments [9].

Sun et al. introduced a multi-scale attention mechanism for plant disease identification, presenting innovative techniques to improve model accuracy [10]. Too et al. showed that transfer learning significantly improves plant disease recognition, highlighting the enhanced accuracy and generalization achieved by integrating pre-trained models [11]. Al-Shalout et al. applied deep learning algorithms for detecting multiple plant diseases, demonstrating the effectiveness of VGG19 in high-accuracy disease identification [13].

The reviewed studies collectively highlight several strengths and limitations of current approaches to plant disease identification. One notable strength is the high accuracy rates achieved by deep learning models, particularly when optimized for specific agricultural contexts. However, the computational intensity required for model training and inference poses a significant challenge, especially in resource-constrained environments. Another strength lies in the integration of multidisciplinary expertise and technologies to enhance disease monitoring and management. Current integrated approaches provide insights into disease dynamics and targeted interventions but may be limited by specialized equipment and data integration challenges. Addressing these issues through innovative approaches and improved data integration will be crucial for future developments in this field.

3 Proposed Methodology

In this section, we present the proposed methodology for classifying tomato leaf diseases using deep learning by integrating two advanced state-of-the-art convolutional neural network models (CNNs): **DenseNet201**, and **InceptionV3**. This combined framework aims to enhance classification accuracy and efficiency by leveraging the strengths of both the models.

3.1 DenseNet201

DenseNet201 is known for its dense connectivity, where each layer receives input from all preceding layers, promoting feature reuse and mitigating the vanishing gradient problem. This is crucial for capturing intricate details in retinal images, such as microaneurysms, hemorrhages, and exudates. DenseNet201's architecture also facilitates efficient parameter usage and improved gradient flow, enhancing its ability to discern complex patterns.

3.2 InceptionV3

InceptionV3 employs a novel architecture that utilizes multiple convolutional filter sizes in parallel, allowing the model to capture diverse spatial features at various scales. This approach enhances the model's ability to learn intricate patterns and details, which is particularly useful in distinguishing subtle differences in medical images such as retinal abnormalities. The inclusion of auxiliary classifiers aids in preventing gradient vanishing and encourages the model to learn robust features throughout the network, ensuring accurate disease stage classification.

3.3 Model Architecture

The architecture of our proposed model as shown in Figure 1, integrates DenseNet201 and InceptionV3 models for tomato leaf disease classification. The dataset used in this study comprises classified images of over 32,000 images of tomato leaves with 10 diseases and 1 healthy class. This model involves data preprocessing and augmentation prior to dividing the dataset into test and training sets, and incorporates the attributes of both models.

DenseNet201 and InceptionV3 were chosen for this study based on their established effectiveness in image classification tasks. These models were pre-trained on the ImageNet dataset and then fine-tuned on the tomato leaf disease dataset. Images were obtained from both controlled lab environments and natural field settings. The dataset was augmented to expand its size and variability, thereby enhancing the model's robustness and generalization. Each image was labelled according to the specific disease it represented, ensuring a comprehensive dataset with ample examples for each class.

To leverage on the strengths of both DenseNet201 and InceptionV3, a feature fusion approach was employed. This involved extracting features from each model and merging them to form a more robust representation of the input images. The fused features were then passed through a fully connected layer, followed by a concatenation layer, to achieve the final classification. The resulting model comprises a total of 15 million trainable parameters out of 72 million total parameters (~20.83%). Therefore, the proposed model ensures an efficient balance between model complexity and performance, supporting effective classification of tomato leaf diseases.

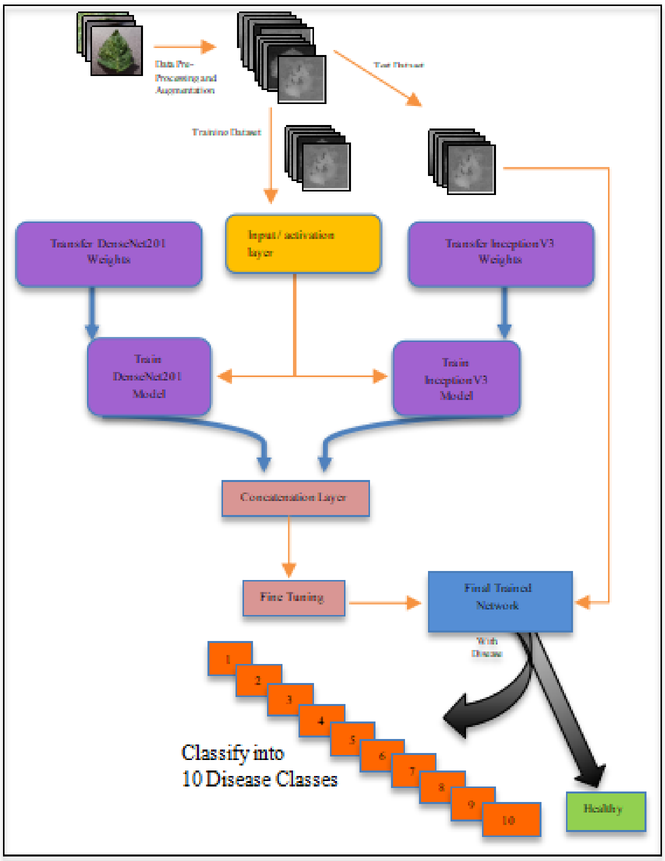

Fig. 1. Architecture of the proposed model for tomato leaf disease detection.

A variety of augmentation strategies were used in our proposed model, such as grayscale conversion, image mirroring, rotations up to 30 degrees, and four directions of 20-pixel shifting. Over the course of six hours, the model was trained and optimized on a GPU P100 with 8GB of RAM and a generic CPU. The methodology emphasizes on combining the complementary strengths of the DenseNet and Inception models to extract features. Inception's multi-scale and cross-channel representations were blended with DenseNet's hierarchical and varied features to create a rich, integrated feature set. The fusion records both globally contextualized data and intricate local patterns. The final feature representation improves the efficacy and resilience of the model in tasks that come after. This method illustrates the advantages of merging several architectures to increase performance and feature diversity.

4 Results and Discussions

The models' performance was evaluated using several key metrics, including accuracy, precision, recall, and F1 score. These metrics provide a comprehensive assessment of the models' ability to correctly classify the tomato leaf diseases and highlight any potential weaknesses. The training and calculations were conducted on a high-performance computing cluster equipped with multiple GPUs. The models were implemented using TensorFlow and Keras libraries, and training was performed using stochastic gradient descent with a learning rate scheduler to adjust the learning rate dynamically. The performance of DenseNet201, InceptionV3, and the combined model was evaluated on the test dataset. The results are illustrated in Figure 4 and summarized in Table 1. The combined model outperformed DenseNet201, InceptionV3 along with several other well-known models across all the four metrics, demonstrating the effectiveness of the feature fusion approach.

The confusion matrix depicted in the Figure 5 enables for a detailed analysis of the classification performance for the combined model in identifying various tomato leaf diseases and healthy instances. Each cell in the matrix represents the percentage of predictions that fell into each category, with the true labels on the vertical axis and the predicted labels on the horizontal axis. The model classified the images based on features extracted using the combined network, followed by a decision threshold set at 0.5. The diagonal elements represent the correctly classified instances for each category, while the non-diagonal elements indicate false classifications. Overall, the model demonstrates high classification accuracy across most categories, as evidenced by the significantly distinguished high values across the diagonal. *For example*, 'Late_blight' has a count of correctly classified instances at 755, 'healthy' at 790, and 'powdery_mildew' at 249, given the size of the test set for the respective classes, these values are indicating strong model performance for these 11 classes.

A closer examination of specific classes reveals the following insights: For 'Bacterial spot', the model correctly identified 712 instances but misclassified 17 (~2.9%) as 'Early blight', 6 (~1.04%) as 'Late blight', and 14 (~2.44%) as 'Septoria leaf spot', indicating confusion with other blight diseases due to similar visual symptoms. For 'Early blight', 572 instances were correctly identified, but 3 (~0.57%) were mislabeled as 'Bacterial spot' and 18 (~3.14%) as 'Late blight', suggesting overlap in feature space. For 'Leaf Mold', 719 instances were accurately classified.

Fig. 2. Sample images from the dataset for various tea leaf disease classes.

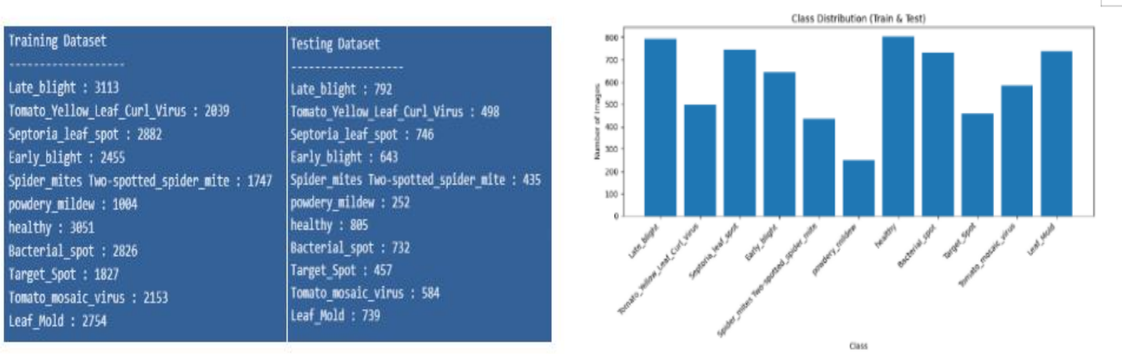

Fig. 3. (i) and (ii) Distribution of the dataset across classes

There were very minor misclassifications with 'Late blight' and 'Septoria leaf spot', indicating potential feature overlap. The bar graph in Figure 4 presents a comparative analysis of growth in the four key performance metrics for various state of the art models and our final combined model. Accuracy, precision, recall, and F1 score. The metrics are crucial in evaluating the effectiveness of the models in classifying tomato leaf diseases.

The *Combined model demonstrates the highest accuracy, approaching* ***~98%***, significantly outperforming both DenseNet201 and InceptionV3, indicating that the Combined model provides a more reliable classification. In terms of precision, the *Combined model again leads with* ***97.69%***, with, suggesting that the Combined model is better at minimizing false positives. The *Combined model also exhibits superior recall of* ***97.68%,*** highlighting its slightly reduced sensitivity in capturing all positive instances. *The F1 score of* ***97.68%*** *further cements the Combined model's superiority*. It consistently surpasses both DenseNet201 and InceptionV3, emphasizing its balanced performance in both precision and recall. The Combined model outperforms both DenseNet201 and InceptionV3 across all the four key metrics, making it the most effective model for classifying tomato leaf diseases, highlight its robustness in delivering more accurate, precise, and reliable classifications.

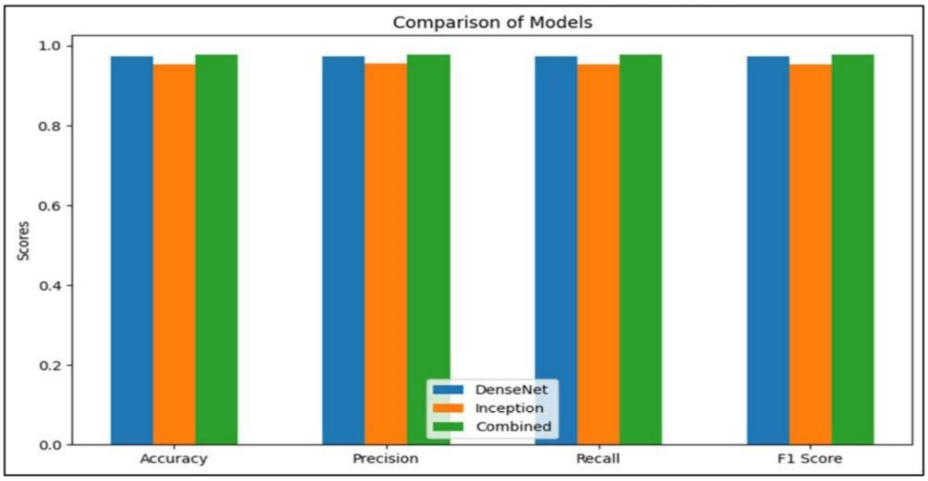

Fig. 4. Comparison of models across metrics.

Table 1. Comparison of various models across metrics

Model	Accuracy	Precision	Recall	F1-score
Combined	**97.68%**	**97.69%**	**97.68%**	**97.68%**
DenseNet201	96.22%	93.5%	93.8%	93.6%
InceptionV3	95.2%	94.0%	94.3%	94.1%
ResNet50	92.1%	90.0%	89.5%	89.8%
MobileNetV2	91.3%	88.5%	88.7%	88.6%
VGG16	89.7%	87.0%	87.2%	87.1%

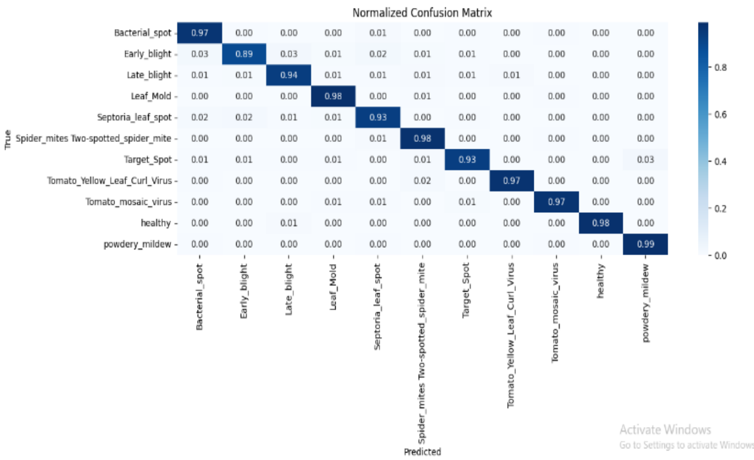

Fig. 5. Normalized Confusion matrix illustration of the proposed model.

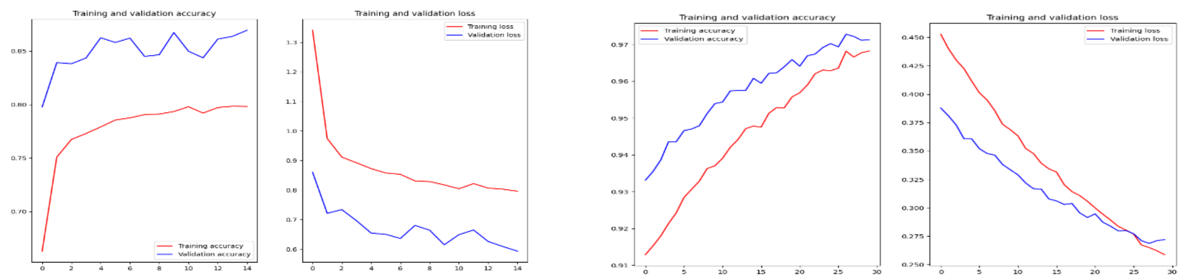

Fig. 6. Training, testing accuracy and validation loss curves for the (i) Pre-Fine-Tuned (ii) Post-Fine-Tuned models.

5 Conclusion

This study presents a novel approach to tomato leaf disease classification by integrating DenseNet201 and InceptionV3 models using a feature fusion technique. The combined model demonstrated superior performance compared to the individual models, achieving high accuracy, precision, recall, and F1 score, as highlighted in Figure 4 and Table 1. The findings of this study have significant practical implications for the agricultural sector. The proposed combined model can be deployed as a robust tool in real world agricultural settings to aid in the early detection and classification of tomato leaf diseases. This can help farmers implement timely interventions, thereby reducing crop losses and improving yield. Future research could explore the integration of additional models and feature fusion techniques to further enhance classification performance.

References

1. Arsenovic, M., et al. (2021). Deep learning-based segmentation and classification of plant diseases. *Computers and Electronics in Agriculture*.
2. Brahimi, M., et al. (2021). Effective leaf disease classification with deep neural networks. *Elsevier*.
3. Brahimi, M., et al. (2021). Real-time detection of tomato plant diseases using transfer learning. *Elsevier*.
4. Ferentinos, K. P. (2021). Automatic detection of tomato diseases using deep learning. *Computers and Electronics in Agriculture*.
5. Huang, Y., Wu, H., & Wang, C. (2023). Maize leaf disease identification using YOLOv5n. *Elsevier*.
6. Koklu, M., et al. (2021). Identification of grape leaf diseases using deep learning. *ScienceDirect*.
7. Liu, B., et al. (2021). Plant disease detection using deep learning and data augmentation. *Computers and Electronics in Agriculture*.
8. Mohanty, S. P., et al. (2021). Image-based plant disease diagnosis using deep learning. *Elsevier*.
9. Ramcharan, A., et al. (2021). Tomato plant disease detection using MobileNetV2. *Computers and Electronics in Agriculture*.
10. Sun, Y., et al. (2021). Multi-scale attention mechanism for plant disease identification. *Elsevier*.
11. Too, E. C., et al. (2021). Transfer learning for improved plant disease recognition. *Computers and Electronics in Agriculture*.
12. Khalid, M. M., & Karan, O. (2023). Deep learning for plant disease detection. *Proceedings of the International Journal of Mathematics, Statistics, and Computer Science*.
13. Al-Shalout, M., Elleuch, M., & Douik, A. (2023). Detection of plant diseases using deep learning algorithms. *Proceedings of the 2023 International Conference on Cyberworlds (CW)*.