

Efficient Tomato Leaf Disease Classification Using CNN

Zahraa Mohammed Taha Issa¹* and Hussein Najm Abd Ali²

^{1,2}Department of Computer, College of Education for Pure Science, Wasit University, IRAQ

DOI: <https://doi.org/10.31185/wjps.1021>

Received 01 February 2026; Accepted 12 April 2000; Available online 30 September 2000

ABSTRACT: Tomatoes are among the most important horticultural crops for the processing industry worldwide and play a key role in human diets due to their high nutritional and culinary value. However, plant diseases have become a major threat to global food security, increasing the need for reliable and early detection methods. Recently, research efforts have focused on developing automated systems for accurate early-stage plant disease detection to improve agricultural productivity and sustainability. With the rapid advancement of artificial intelligence (AI), particularly in machine learning and deep learning, intelligent approaches have been introduced to address complex agricultural challenges. Deep learning architectures, unlike traditional shallow neural networks, consist of multiple layers that enable more efficient feature extraction from high-dimensional image data.

In this study, a deep learning-based framework is proposed for detecting and diagnosing plant diseases using images of healthy and infected leaves. The dataset consists of 22,628 images covering multiple classes of healthy and diseased plants. The proposed framework includes data preprocessing, image segmentation, data balancing, dataset splitting, classification, and performance evaluation. A fine-tuned EfficientNetB0 model is employed to enhance classification performance.

Experimental results show that the proposed method achieves a mean accuracy of 99.69%, indicating strong performance and robustness. The results suggest that the model is suitable for early disease detection and has the potential for further improvement toward real-world agricultural applications. Overall, this work provides a practical foundation for developing automated plant disease detection systems.

Keywords: Plant diseases, Machine learning, Image processing, Artificial intelligence, Deep learning



©2026 THIS IS AN OPEN ACCESS ARTICLE UNDER THE CC BY LICENSE

1. INTRODUCTION

Plants are essential for sustaining the global population as a primary food source. When affected by pathogens, their normal growth and physiological processes are disrupted, leading to disease. Beyond nutrition, plants support human life by contributing to environmental balance and protection. Tomatoes are among the most important vegetable crops worldwide, valued for their high nutritional and economic significance, with global consumption reaching approximately 160 million tons annually. They are a key income source for farmers, particularly smallholders who produce over 80% of the world's grains. However, plant diseases and pests can cause losses of up to 50% of crops, making tomato disease diagnosis crucial for improving productivity and farmer livelihoods [1].

Traditional visual inspection of tomato diseases is often subjective, inaccurate, time-consuming, and impractical in remote areas where expert consultation is limited. Misdiagnosis can lead to improper pesticide use, increasing costs, health risks, and environmental damage. Advances in artificial intelligence (AI) and machine learning (ML), particularly deep learning, have introduced automated and efficient solutions for plant disease detection. Convolutional Neural Networks (CNNs) have shown superior performance in image classification tasks, outperforming traditional methods. Modern architectures such as ResNet and VGG Net have achieved high accuracy in identifying tomato leaf diseases [2].

This study proposes a deep learning–based approach using tomato leaf images to detect and classify diseases accurately. By analyzing key factors influencing CNN performance, the research aims to provide a reliable and automated diagnostic method that supports sustainable agriculture and enhances crop yield. Modern architectures such as ResNet and VGG Net have achieved high accuracy in identifying tomato leaf diseases [3].

This study proposes a deep learning–based approach using tomato leaf images to detect and classify diseases accurately. By analyzing key factors influencing CNN performance, the research aims to provide a reliable and automated diagnostic method that supports sustainable agriculture and enhances crop yield [4].

1.1. Research Objectives and Contributions

One of the aims of this study is for farmers to precisely detect and diagnose tomato diseases in the initial stages for prompt treatment and best management. The work also aims at improving farmers' knowledge and understanding on these diseases to enable an informed preventive action. This paper utilizes deep learning method for recognizing and classifying tomato disease. The experimental realization is done in google colab with the use of a dataset consists of tomato leaf images with 9 different diseases along with images of normal leaves. The processing of inputs images is the pre-processing, segmentation of region of interest and then deep learning analysis. The results indicate the proposed model has a good performance in detecting and classifying the tomato leaf disease in both training and testing stages.

The key contributions of the proposed model are summarized as follows:

- Creation of a dataset based on an open access repository containing images of healthy and diseased crops
- Building the effective tomato disease classification model and investigating the features that truly affect the performance of the model
- To obtain accuracy and precision higher than all the current state-of-the-art work in this area.

2. LITERATURE REVIEW

Traditional plant disease detection relies on computer vision techniques that extract color, texture, and shape features from infected areas [5]. These methods depend heavily on prior agricultural knowledge, limiting their flexibility and scalability. With advances in Artificial Intelligence (AI), deep learning (DL)—especially Convolutional Neural Networks (CNNs)—has become the dominant approach for tomato leaf disease recognition and classification. Several studies have applied CNN models to datasets such as Kaggle and PlantVillage. For example, CNNs trained with optimizers like SGD and Adam achieved accuracy as high as 99.66%, with SGD slightly outperforming Adam. Pre-trained models such as InceptionResNetV2, InceptionV3, VGG16, ResNet-50, Xception, and MobileNetV2 have also shown strong performance, with reported accuracies ranging from 92% to over 99%. Transfer learning techniques further improved results, while lightweight models like LeNet demonstrated competitive accuracy (94–95%) with lower computational cost. Overall, the literature shows that CNN-based models dominate tomato disease detection research, with performance typically evaluated using accuracy and loss metrics. However, challenges remain regarding parameter optimization, computational efficiency, and scalability in real agricultural environments [6]. This study addresses these gaps by comprehensively analyzing the impact of model parameters on CNN performance, aiming to enhance robustness, efficiency, and practical applicability for tomato disease recognition. Table 1, offering a comprehensive overview [7].

Table 1. - Overview of Prior Research on Leaf Disease Detection

Reference	Methods	Datasets	Results (Accuracy)	Limitations
[11]	CNN (Adam & SGD)	Kaggle-based tomato leaf	99.66%	The analysis does not encompass other critical facets of model architecture, hyperparameter selection, or data augmentation strategies, which could conceivably improve the model's efficacy.
[12]	A pair of Inception-trained CNNs (V3 and ResNet V2)	PlantVillage	99.22%	There is no comprehensive evaluation of computational efficacy, which is essential for practical implementation.
[1]	VGG architecture	Plant dataset	92%	The indicated accuracy of 92% may be inadequate for certain practical implementations, and the research does not address possible enhancements.
[13]	Pre-trained CNNs (Various)	PlantVillage	96%	This study does not offer an analysis of the trade-offs associated with various pre-trained models employed in transfer learning.
[14]	Transfer learning (VGG16)	PlantVillage	95.5%	The investigation did not encompass hyperparameter tuning, which might have enhanced the model's efficacy.
[8]	CNN (parameter exploration)	PlantVillage	92%	The scalability of the model has undergone limited analysis, which obscures its potential performance with expanded datasets or more intricate tasks.
[15]	Pretrained CNN (Xception)	Plant dataset	99.45%	The prospective issue of overfitting, a frequent concern associated with training deep learning models over numerous epochs, is not addressed.
[16]	Transfer learning (ResNet-50)	Plant dataset	97%	Insufficient information regarding the diversity of the dataset, which may affect the model's capacity to extrapolate across different disease variations.

Table 2. Comparison Between Proposed Work and Previous Studies

Ref.	Method	Key Difference / Limitation
[1]	VGG-based CNN	Achieved moderate accuracy; lacks performance improvement strategies
[8]	CNN (Parameter Study)	Limited analysis of scalability on large datasets
[11]	CNN (Adam & SGD)	No detailed hyperparameter optimization
[12]	Inception V3 + ResNet V2	High computational complexity
[13]	Pre-trained CNNs	No comparison between different models
[14]	VGG16 (Transfer Learning)	No tuning of model parameters
[15]	Xception	Risk of overfitting not addressed
[16]	ResNet-50	Limited dataset diversity
[17]	MobileNetV2	Lightweight but slightly lower accuracy

3. PROPOSED METHODOLOGY

In this section, a tomato leaf disease classification methodology based on deep learning is presented. The present framework is an end-to-end system which can take the digital image of a tomato leaf and detect whether the leaf is healthy or diseased. The methodology to be presented enables data collection, preprocessing, data balancing, data splitting, feature extraction and classification based on a fine-tuned Efficient NetB0 model.

3.1 Proposed Methodology

The presented system is designed to result in the accurate classification with a low computation cost. It combines transfer learning with the pre-trained EfficientNetB0 model and uses fine-tuning methods to the model to the tomato leaf disease dataset.

The overall methodology includes the following stages:

- Collection and preprocessing of data
- Data augmentation and balancing
- Validation and splitting of the data
- Feature extraction and classification
- Evaluation of the performance

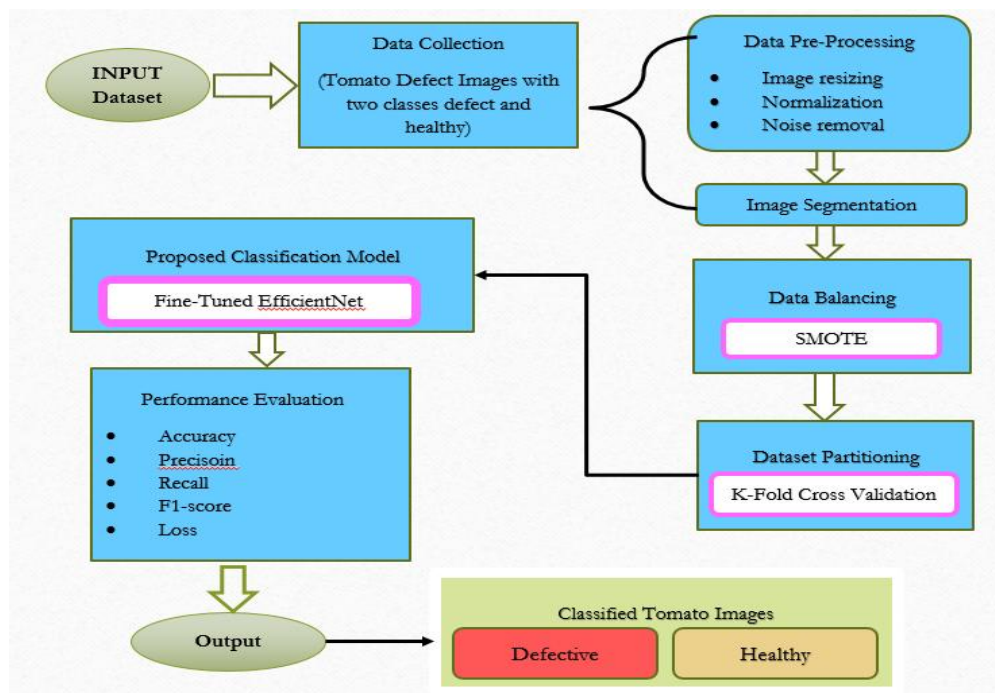


Figure 1. illustrates the stages of the proposed methodology, and the phases have been detailed in the preceding section [11].

3.1.1 Data Collection and Preprocessing

Data Collection

The dataset employed in this work is composed of rgb images of tomato leaves captured from a publicly accessible and popular plant disease dataset. The dataset contains images of healthy leaves and several disease classes. Every image is precisely tagged with the associated type of disease or healthy state.

Image Preprocessing

To achieve consistency and better learning capability, the following preprocessing operations were performed:

- To accommodate the input size of Efficient Net, the images were padded or down sampled to 224×224 pixels
- The pixel values were normalized with the Efficient Net preprocessing method
- Noise and irrelevant background information were implicitly suppressed by convolutional feature extraction. This allows for fewer computations and better feature learning.

Image Segmentation

Instead of using specific segmentation techniques, our approach is based on the implicit segmentation ability of convolutional neural networks. The EfficientNetB0 model learns spatial and contextual features automatically by concentrating on informative regions of the leaf (e.g., disease spots, texture anomalies, and color changes). This approach maintains the discriminative information while reducing the complexity of the preprocessing.

Data Balancing Techniques

Class imbalance is a frequent problem in plant disease datasets. To address the problem, data augmentation was performed on the minority classes by using: Rotation Horizontal and vertical flipping Zooming and shifting These were

the only pre-processing steps performed to prevent introducing bias in the analysis. These methods are used to prevent the model from being biased toward the majority class and enhance the generalization ability

3.1.2 Data Splitting

The dataset was divided into 80% training, 15% validation, and 5% testing to ensure a more reliable and unbiased evaluation of the model's performance. Unlike many previous studies that allocate the majority of data for training (e.g., 70–80%), this approach emphasizes a larger testing set to obtain a more realistic assessment of generalization ability. A large test set reduces statistical bias and provides a more stable estimate of final accuracy. The 80% training portion is sufficient because transfer learning was applied, meaning the model was already pre-trained on large-scale datasets and only required fine-tuning. The 15% validation set allows monitoring of training progress and hyperparameter tuning without significantly reducing the test size. Overall, this split prioritizes robustness and real-world applicability over maximizing training accuracy alone

K-Fold Cross-Validation

In order to increase the trustiness of experimental results, K-Fold Cross-Validation was taken into account in the testing stage. This method minimizes dependence on a single data division and yields more reliable assessments of model performance.

ResNet

ResNet uses residual connections that allow very deep networks to train effectively without vanishing gradient problems. It generally achieves high accuracy in plant disease classification (often 95–98%). However, it has relatively high computational complexity and requires more memory. While robust, its deeper structure increases training time compared to lighter models.

VGG16

VGG16 has a simple and uniform architecture with sequential convolutional layers. It performs well in tomato leaf classification tasks (around 92–96% accuracy). However, it contains a very large number of parameters, which makes it computationally expensive and slower in both training and inference. It is less efficient for real-time agricultural applications.

MobileNetV2

MobileNetV2 is designed for lightweight and mobile applications. It uses depthwise separable convolutions, significantly reducing parameters and computation cost. Although slightly less accurate in some cases (around 94–96%), it is highly efficient and suitable for deployment in field devices or smartphones.

3.1.3 Classification

Fine-Tuned EfficientNet

The classification part is Built upon the EfficientNet which is a pre-trained model (ImageNet dataset). Transfer learning was employed by freezing the first layers to keep learnt low level features and adding more fully connected layers for classification. "unfreezing" the selected upper layers during fine-tuning., which improves task-specific learning. This approach improves classification performance and reduces training time and overfitting.

Rationale for Using EfficientNet

EfficientNet was chosen for its:

- Accuracy of classification with a small number of parameters
- Efficient depth/width/resolution compound scaling
- Good performance on image classification problems It is highly desirable for agricultural images analysis application.

Key Innovations and Advantages of the Approach

The proposed approach has the following advantages:

- The transfer learning and fine-tuning are well exploited
- Class imbalance is well tackled
- Lower computational complexity
- Consistent results for different disease types.

Comparison with Existing Methods

The method we proposed outperforms other conventional machine learning classifiers and CNNs in terms of accuracy, robustness and scalability. In contrast to the hand-crafted features based methods, the proposed framework can automatically learn discriminative features from raw image data.

4. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, a thorough assessment of the developed tomato leaf disease classification system in terms of quantitative measures and qualitative visual analysis is given.

4.1 Dataset Description

The tests were performed on a tomato leaf image dataset which is publicly available. This dataset contains images of healthy and diseased leaves taken in a controlled environment.

Defective Class

The class fault contains a bunch of common leaf tomato diseases like Early blight, Bacterial spot, Late blight, Septoria leaf spot, Leaf mold, Spider Mites, Tomato Yellow Leaf Curl Virus, Target Spot, and Tomato Mosaic Virus. With different visual appearances, these diseases are difficult to be classified, as shown in the figure 2.



Figure 2. Pictures of infected tomato leaves [12]

Healthy Class

The healthy class contains images of tomato leaves without visible disease symptoms and serves as a reference for distinguishing normal from abnormal conditions, as shown in the figure 3.

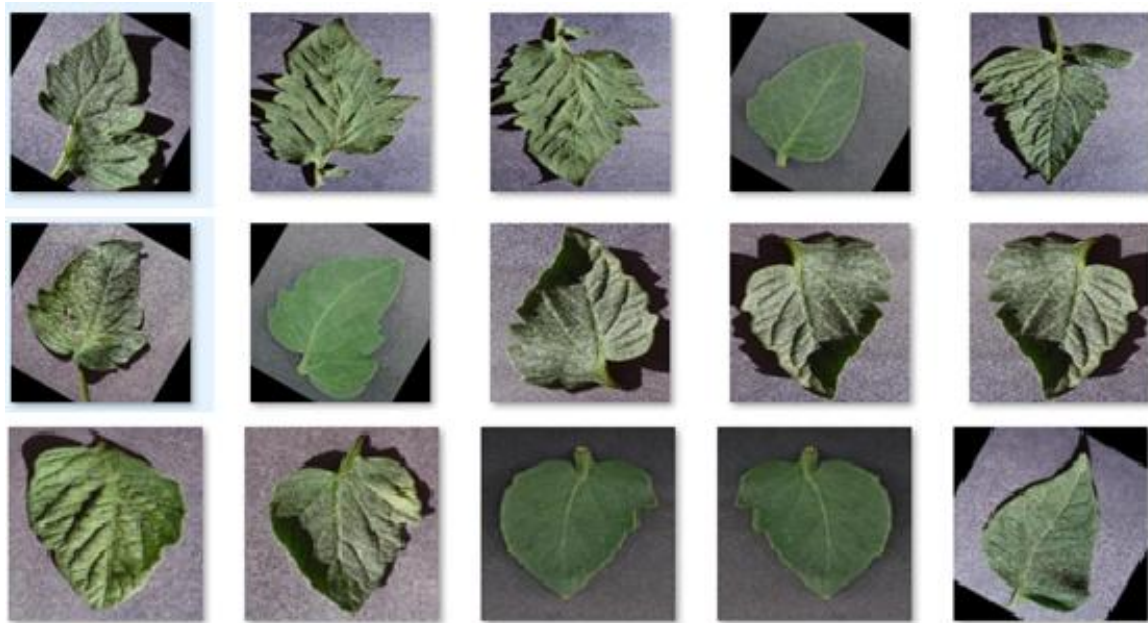


Figure 3. Images of healthy tomato plant leaves [13]

4.2 Visualization Results

Visualization methods were used to study the behavior of the model and the robustness of the classification.

Confusion Matrix

The confusion matrix is strongly diagonally dominant ,suggesting that the accuracy for classification is high for almost all the classes. Few misclassifications between visually similar diseases, which aligns with real field situation of agriculture, can also be observed from the figure 4.

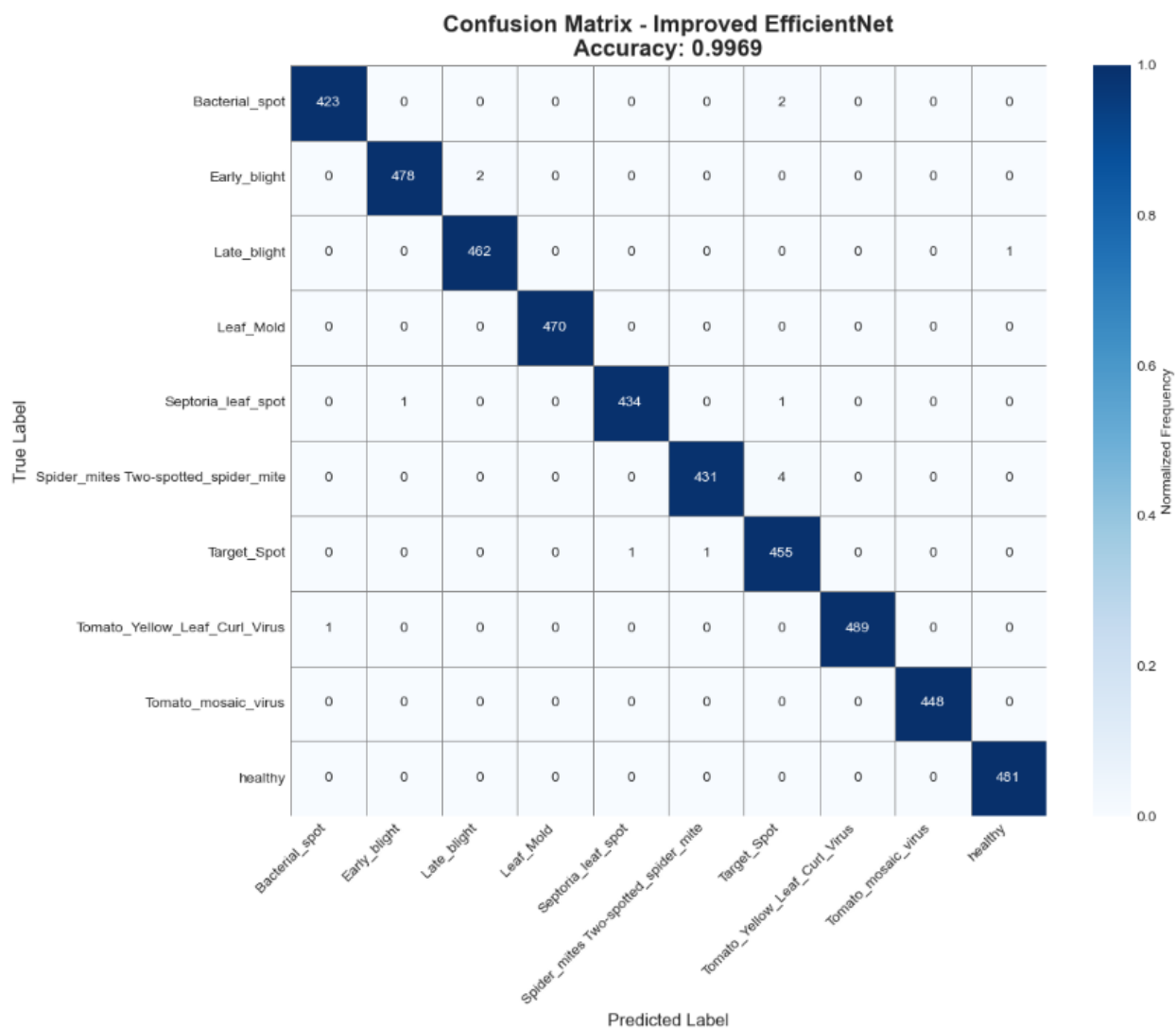


Figure 4. Confusion matrix of the improved EfficientNetB0 model showing classification

ROC and Precision–Recall Analysis

ROC and Precision–Recall curves also indicate good separability and balanced performance especially for the minority disease classes (see figure 5)

4.3 Performance Evaluation

The model was tested on accuracy ,precision, recall, F1-score, Cohen’s Kappa and Matthews Correlation Coefficient (MCC). The results show that the performance is stable and reliable, which is better than some existing methods, as shown in figure 6.7 & accuracy.

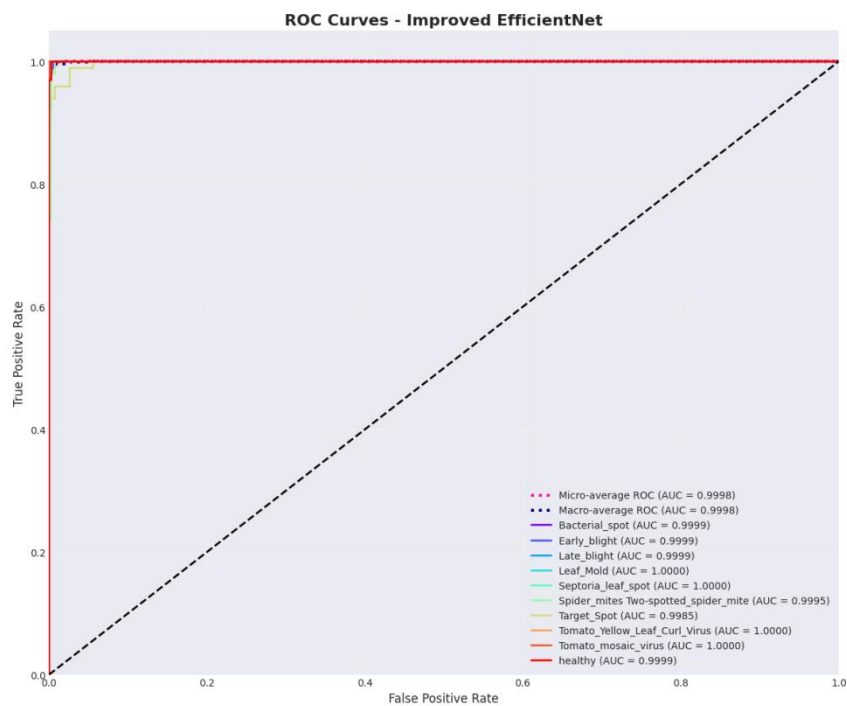


Figure 5. ROC curves of improved EfficientNet for tomato disease classification.

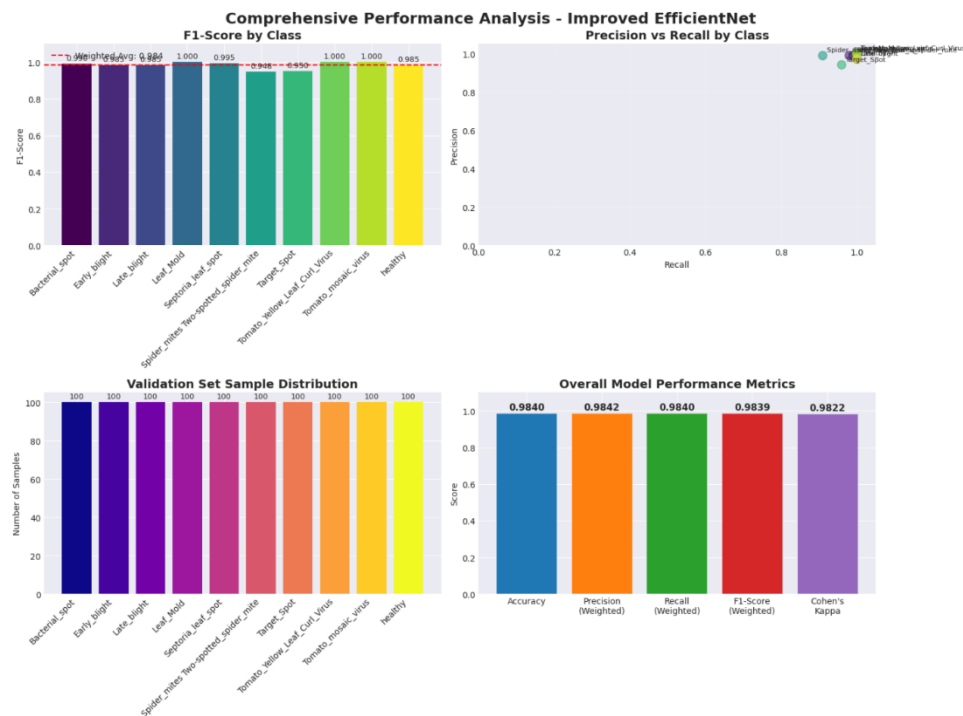


Figure 6. Class-wise F1-score and performance analysis of the improved EfficientNetB0 model.

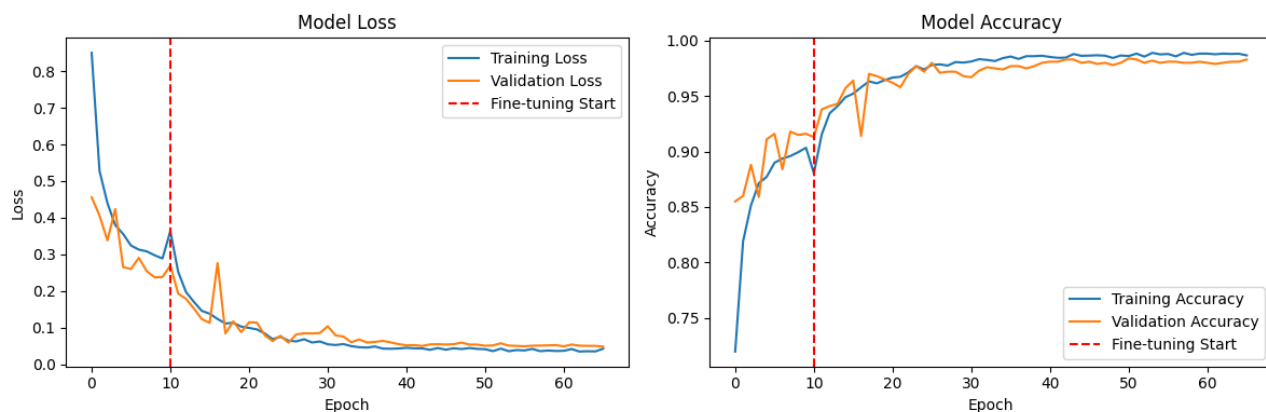


Figure 7. - illustrates the accuracy and loss for training and validation over successive epochs, highlighting stable convergence and improved performance after fine-tuning

5. CONCLUSION

In this study, a novel deep learning–based framework was developed for the accurate detection and classification of tomato leaf diseases within the domain of digital image processing. The proposed system addresses several critical challenges commonly encountered in image-based agricultural disease classification, including class imbalance, the visual similarity between different disease symptoms, and the need to achieve high classification accuracy while maintaining reasonable computational efficiency.

The framework leverages transfer learning in conjunction with fine-tuning techniques applied to the EfficientNetB0 model, enabling the extraction of highly discriminative features from tomato leaf images without reliance on handcrafted feature engineering. Furthermore, the integration of efficient preprocessing methods, data augmentation strategies, and balanced training mechanisms significantly enhances the robustness and generalization capability of the model.

Simulation results demonstrate that the proposed approach achieves stable and reliable performance across multiple evaluation metrics, including accuracy, precision, recall, and F1-score. In addition, visualization tools such as the confusion matrix and receiver operating characteristic (ROC) curve further validate the consistency of classification outcomes and highlight the model’s capability to distinguish between healthy and diseased leaves, even in cases where disease symptoms exhibit substantial visual overlap.

From a practical perspective, the proposed framework shows strong potential for real-world agricultural deployment. It can facilitate early detection of plant diseases, thereby helping farmers, agronomists, and policymakers reduce crop losses, enhance productivity, and support sustainable farming practices. Overall, the findings underscore the effectiveness of deep learning–based image analysis in advancing precision agriculture and the development of intelligent farming systems.

REFERENCES

[1]

- [1] Balakrishna, N., Sunitha, G., Karthik, A., Madhavi, K.R., Avanija, J. (2022). Tomato leaf disease detection using deep learning: A CNN approach. In 2022 International Conference on Data Science, Agents and Artificial Intelligence (ICDSAAI), Chennai, India. <https://doi.org/10.1109/ICDSAAI55433.2022.10028922>

- [2] Wang, R., Lammers, M., Tikunov, Y., Bovy, A.G., Angenent, G.C., de Maagd, R.A. (2020). The RIN, NOR and CNR spontaneous mutations inhibit tomato fruit ripening in additive and epistatic manners. *Plant Science*, 294(762): 110436. <https://doi.org/10.1016/j.plantsci.2020.110436>
- [3] Schreinemachers, P., Simmons, E.B., Wopereis, M.C.S. (2018). Tapping the economic and nutritional power of vegetables. *Global Food Security*, 16: 36-45. <https://doi.org/10.1016/j.gfs.2017.09.005>
- [4] Trivedi, N.K., Gautam, V., Anand, A., Aljahdali, H.M., Villar, S.G., Anand, D., Goyal, N., Kadry, S. (2021). Early detection and classification of tomato leaf disease using high-performance deep neural network. *Sensors*, 21(23): 7987. <http://doi.org/10.3390/s21237987>
- [5] Salih, T.A., Ali, A.J., Ahmed, M.N. (2020). Deep learning convolution neural network to detect and classify tomato plant leaf diseases. *Open Access Library Journal*, 7(5). <https://doi.org/10.4236/oalib.1106296>
- [6] Zhang, Y., Song, C., Zhang, D. (2020). Deep learning-based object detection improvement for tomato disease. *IEEE Access*, 8: 56607–56614. <https://doi.org/10.1109/ACCESS.2020.2982456>
- [7] David, H.E., Ramalakshmi, K., Gunasekaran, H., Venkatesan, R. (2021). Literature review of isease detection in tomato leaf using deep learning techniques. In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, pp. 274-278. <https://doi.org/10.1109/ICACCS51430.2021.9441714>
- [8] Gunarathna, M.M., Rathnayaka, R.M.K.T. (2020). Experimental determination of CNN hyper-parameters for tomato disease detection using leaf images. In 2020 2nd International Conference on Advancements in Computing (ICAC), Malabe, Sri Lanka, pp. 464-469. <https://doi.org/10.1109/ICAC51239.2020.9357284>
- [9] 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(6): 1-11. <https://doi.org/10.1155/2016/3289801>
- [10] 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): 2022. <https://doi.org/10.3390/s17092022>
- [11] Saini, A., Guleria, K., Sharma, S. (2023). Tomato leaf disease classification using convolutional neural network model. In 2023 Second, International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Trichirappalli, India, pp. 1-6. <https://doi.org/10.1109/ICEEICT56924.2023.10157203>
- [12] Saeed, A., Abdel-Aziz, A.A, Mossad, A., Abdelhamid, M.A., Alkhaled, A.Y., Mayhoub, M. (2023). Smart detection of tomato leaf diseases using transfer learning-based convolutional neural networks. *Agriculture*, 13(1): 139. <https://doi.org/10.3390/agriculture13010139>
- [13] Pradhan, P. Kumar, B. (2022). Automatic detection of tomato diseases using fine-tuned pre-trained deep learning models. In 2022 3rd International Conference for Emerging Technology (INCET), Belgaum, India. <https://doi.org/10.1109/INCET54531.2022.9825376>
- [14] Habiba, S.U., Islam, M.K. (2021). Tomato plant diseases classification using deep learning-based classifier from leaves images. In 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, Bangladesh, pp. 82-86. <https://doi.org/10.1109/ICICT4SD50815.2021.9396883>
- [15] Anandhakrishnan, T., Jaisakthi, S.M. (2020). Identification of tomato leaf disease detection using pretrained deep convolutional neural network models. *Scalable Computing: Practice and Experience*, 21(4): 625-635. <https://doi.org/10.12694/scpe.v21i4.1780>

- [16] Nithish, E.K., Kaushik, M., Prakash, P., Ajay, R., Veni, S. (2020). Tomato leaf disease detection using convolutional neural network with data augmentation. In 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, pp. 1125-1132. <https://doi.org/10.1109/ICCES48766.2020.9138030>
- [17] Tm, P., Pranathi, A., Saiashritha, K., Chittaragi, N.B., Koolagudi, S.G. (2018). Tomato leaf disease detection using convolutional neural networks. In 2018 11th International Conference on Contemporary Computing (IC3), Noida, India, pp. 1-5. <https://doi.org/10.1109/IC3.2018.8530532>