GMR
The Original

UTLPD-NET: A UNIFIED TRANSFER LEARNING FRAMEWORK FOR AUTOMATED PLANT LEAF DISEASE DETECTION USING COMPARATIVE DEEP CONVOLUTIONAL NEURAL NETWORKS

M.Sampoorna[1], Dr.M.Chandran[2], Sri Ramakrishna[3]

[1]Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore, Tamil Nadu, India , Research Scholar, Department of Computer Science, m.sampoorna87@gmail.com,
[2]Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore, Tamil Nadu, India Associate Professor, Department of Computer Applications ,onchandran@gmail.com

ABSTRACT:
Plant diseases significantly reduce agricultural productivity and threaten global food security by affecting crop yield and quality. Early and accurate disease diagnosis is therefore essential for precision agriculture. Although transfer learning-based convolutional neural networks (CNNs) have achieved promising results in plant disease classification, existing comparative studies frequently employ different preprocessing techniques, augmentation strategies, and training protocols, making objective architectural comparison difficult. This paper proposes UTLPD-Net (Unified Transfer Learning Framework for Plant Disease Detection), a standardized benchmarking framework for evaluating pretrained CNN architectures under identical experimental conditions. The proposed framework integrates dataset preparation, image preprocessing, online data augmentation, adaptive transfer learning, fine-tuning, and comprehensive performance evaluation into a unified pipeline. Four pretrained CNN architectures—MobileNetV2, ResNet50, DenseNet201, and EfficientNetB3—were evaluated using the PlantVillage Dataset, comprising 87,867 RGB images across 38 healthy and diseased classes representing 14 crop species. Experimental results show that ResNet50 achieved the best overall performance with 97.68% accuracy, 97.79% precision, 97.68% recall, 97.68% F1-score, 0.9762 Cohen's Kappa, and 0.9762 MCC, while maintaining competitive computational efficiency. DenseNet201 produced comparable accuracy but required longer training time, whereas MobileNetV2 demonstrated the lowest computational cost, making it suitable for resource-constrained environments. The proposed UTLPD-Net provides a reproducible benchmarking framework for comparative evaluation of transfer learning models and establishes a reliable baseline for future intelligent plant disease diagnosis systems.

KEYWORDS: Plant Disease Detection, Transfer Learning, Deep Learning, Convolutional Neural Networks, PlantVillage Dataset, Precision Agriculture.

INTRODUCTION

Agriculture is fundamental to global food security and economic development. However, plant diseases continue to reduce crop productivity and quality, causing substantial economic losses worldwide. Traditional disease diagnosis relies on manual visual inspection by agricultural experts, which is often labor-intensive, subjective, time-consuming, and impractical for large-scale farming. Moreover, visually similar symptoms among different diseases make accurate diagnosis particularly challenging during the early stages of infection [1]–[4].

Recent advances in computer vision and deep learning have enabled automated plant disease recognition using leaf images. Conventional machine learning techniques based on handcrafted features achieved moderate success but were highly dependent on manual feature engineering and often failed to generalize under varying environmental conditions [5]–[8]. Deep Convolutional Neural Networks (CNNs) overcome these limitations by automatically learning hierarchical feature representations, resulting in significantly improved classification accuracy across a wide range of agricultural applications [9]–[12].

Transfer learning has further accelerated the adoption of CNNs for plant disease diagnosis by adapting models pretrained on ImageNet to agricultural datasets. This approach reduces computational cost and training time while improving classification performance. Popular architectures such as MobileNetV2, ResNet50, DenseNet201, and EfficientNetB3 have demonstrated excellent predictive capability for plant disease recognition [13]–[21]. However, existing comparative studies frequently employ different preprocessing methods, augmentation techniques, hyperparameter settings, and evaluation protocols, making direct comparison among architectures difficult and limiting experimental reproducibility [15], [16], [22]–[25].

To address these limitations, this paper proposes UTLPD-Net (Unified Transfer Learning Framework for Plant Disease Detection), a standardized benchmarking framework that evaluates multiple pretrained CNN architectures under identical experimental conditions. The proposed framework integrates standardized preprocessing, online data augmentation, adaptive fine-tuning, and comprehensive multi-metric evaluation into a unified experimental pipeline. Unlike previous comparative studies, all investigated architectures are trained using the same dataset partitions, optimization strategy, and hyperparameter configuration, enabling objective comparison while minimizing implementation bias [16], [18], [21], [26]–[30].

The framework is validated using a publicly available benchmark dataset comprising 87,867 RGB images representing 38 healthy and diseased classes across 14 crop species. Performance is evaluated using Accuracy, Precision, Recall, F1-score, Cohen's Kappa, Matthews Correlation Coefficient (MCC), Confusion Matrix, Receiver Operating Characteristic (ROC), Precision–Recall (PR) analysis, and computational training time [30]–[41].

The major contributions of this work are:

- Development of the UTLPD-Net unified transfer learning framework for standardized plant disease classification.
- Objective comparison of MobileNetV2, ResNet50, DenseNet201, and EfficientNetB3 under identical experimental conditions.
- Comprehensive evaluation using statistical, agreement-based, and threshold-independent performance metrics.
- Establishment of a reproducible benchmarking framework for future intelligent plant disease diagnosis research.

2. Related Work

Automated plant disease detection has evolved significantly with the advancement of computer vision and deep learning technologies. Early research relied on handcrafted feature extraction combined with conventional machine learning algorithms, whereas recent studies increasingly employ deep convolutional neural networks (CNNs) and transfer learning because of their superior feature learning capability and classification performance. This section reviews the progression of plant disease detection techniques and identifies the research gap addressed by the proposed framework [5]–[8].

2.1 Conventional Machine Learning Approaches

Early plant disease diagnosis systems extracted handcrafted visual features such as colour, texture, edge, and shape descriptors from leaf images. These features were classified using conventional machine learning algorithms including Support Vector Machines (SVM), Artificial Neural Networks (ANN), Decision Trees, Random Forests, and k-Nearest Neighbour (k-NN). Although these approaches produced acceptable results under controlled laboratory conditions, their performance depended heavily on manual feature engineering and expert knowledge. Furthermore, handcrafted features were sensitive to variations in illumination, background, image quality, and disease severity, limiting their applicability in real agricultural environments [5]–[8].

As agricultural image datasets increased in size and complexity, conventional machine learning approaches became less effective because manually designed features were unable to adequately capture complex disease characteristics. These limitations motivated the adoption of deep learning-based feature extraction methods [9], [10].

2.2 Deep Learning and Transfer Learning

Deep Convolutional Neural Networks (CNNs) have transformed image classification by automatically learning hierarchical feature representations directly from raw images. Unlike conventional methods, CNNs eliminate manual feature engineering and significantly improve classification performance across diverse computer vision applications, including agricultural disease diagnosis [9]–[12].

The PlantVillage dataset accelerated research in automated plant disease recognition. Mohanty et al. demonstrated that pretrained CNN models such as AlexNet and GoogLeNet achieved high classification accuracy using transfer learning on PlantVillage images [13]. Ferentinos further validated the effectiveness of deep CNNs for greenhouse crop disease diagnosis, reporting superior performance compared with traditional machine learning techniques [14]. Too et al. compared several pretrained CNN architectures and concluded that DenseNet-based models consistently outperformed many conventional architectures for plant disease classification [15].

Transfer learning has become the preferred approach because pretrained ImageNet models provide robust low-level feature representations that can be effectively adapted to agricultural datasets. Popular architectures including MobileNetV2, ResNet50, DenseNet201, and EfficientNetB3 provide different trade-offs between predictive accuracy and computational complexity [16]–[21].

2.3 Recent Advances and Research Gap

Recent investigations have explored attention mechanisms, lightweight CNNs, ensemble learning, explainable artificial intelligence (XAI), and Vision Transformers (ViTs) to further improve plant disease diagnosis [22]–[27]. Although these methods have achieved promising performance, most studies evaluate different CNN architectures using heterogeneous preprocessing methods, augmentation techniques, optimization algorithms, hardware configurations, and evaluation metrics. Consequently, differences in reported performance often arise from implementation variations rather than intrinsic architectural characteristics [22], [23].

In addition, many comparative studies focus primarily on classification accuracy while providing limited discussion regarding computational efficiency, Cohen's Kappa, Matthews Correlation Coefficient (MCC), Receiver Operating Characteristic (ROC) analysis, Precision–Recall (PR) analysis, and reproducibility. Such inconsistencies make it difficult to identify the most suitable CNN architecture for practical agricultural deployment [24]–[30].

To overcome these limitations, this study proposes UTLPD-Net, a unified transfer learning framework that standardizes preprocessing, augmentation, adaptive fine-tuning, and evaluation procedures across multiple pretrained CNN architectures. By maintaining identical experimental conditions for MobileNetV2, ResNet50, DenseNet201, and EfficientNetB3, the proposed framework enables objective architectural comparison while improving reproducibility and providing a reliable benchmarking methodology for future intelligent agricultural applications.

Table 1. Comparative Review of Representative Plant Disease Detection Studies

Ref.	Study	Dataset	Method	Major Contribution	Limitation
[13]	Mohanty et al.	PlantVillage	AlexNet, GoogLeNet	First large-scale transfer learning study	Controlled laboratory images
[14]	Ferentinos	Greenhouse Dataset	Deep CNN	High classification accuracy	Limited field validation
[15]	Too et al.	PlantVillage	DenseNet, ResNet, Inception	Comparative CNN evaluation	Different preprocessing strategies
[22]	Barbedo	Multiple datasets	CNN Survey	Comprehensive review	Dataset-dependent analysis
[23]	Ramcharan et al.	Cassava Dataset	CNN	Field disease classification	Single crop dataset
Proposed	UTLPD-Net	Benchmark Plant Leaf Image Dataset	Unified Transfer Learning Framework	Standardized benchmarking under identical conditions	Addresses reproducibility and fair comparison

The literature review highlights the following research gaps:

- Lack of a standardized benchmarking framework for comparing pretrained CNN architectures.
- Non-uniform preprocessing, augmentation, and fine-tuning strategies across previous studies.
- Limited use of comprehensive statistical metrics such as MCC, Cohen's Kappa, ROC, and Precision–Recall analysis.
- Insufficient emphasis on reproducibility and computational efficiency.

Accordingly, the proposed UTLPD-Net addresses these limitations by establishing a unified and reproducible experimental framework for comparative transfer learning-based plant disease classification.

3. Proposed UTLPD-Net Framework

This study proposes UTLPD-Net (Unified Transfer Learning Framework for Plant Disease Detection), a standardized transfer learning framework designed to evaluate multiple pretrained CNN architectures under identical experimental conditions. Unlike previous comparative studies that employ heterogeneous preprocessing techniques, augmentation strategies, and optimization settings, the proposed framework establishes a unified pipeline to ensure fair architectural comparison and reproducible experimental evaluation [16], [18], [21], [28].

The framework consists of six sequential stages: dataset preparation, image preprocessing, online data augmentation, transfer learning, adaptive fine-tuning, and performance evaluation. Each CNN architecture undergoes identical implementation procedures, ensuring that differences in performance are attributable to the network architecture rather than variations in experimental methodology.

3.1 Overall Framework

The architecture of the proposed UTLPD-Net is illustrated in Figure 1. Initially, the benchmark leaf image dataset is acquired and preprocessed. Online data augmentation is then performed during training to improve generalization capability. Subsequently, pretrained CNN architectures are adapted through transfer learning and optimized using adaptive fine-tuning. Finally, disease prediction is performed using a Softmax classifier, and model performance is evaluated using multiple statistical measures.

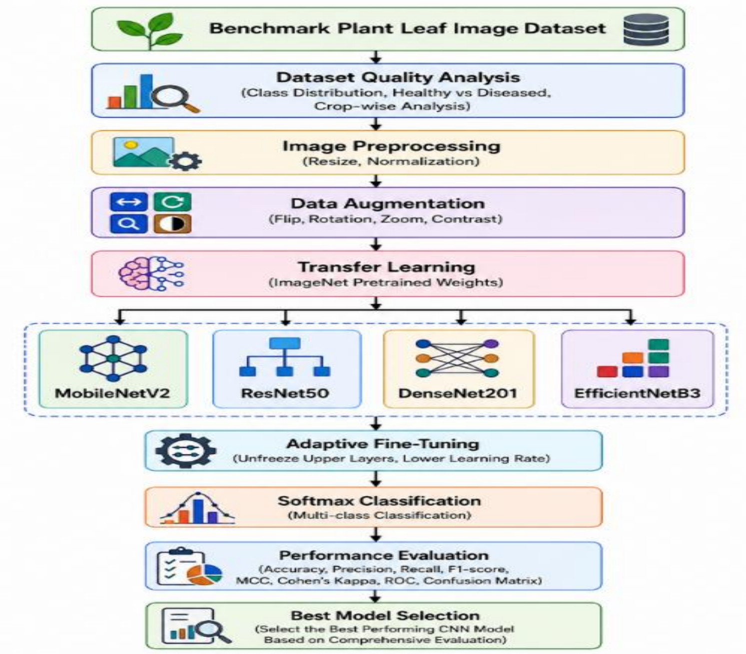


Figure 1. Overall Architecture of the Proposed UTLPD-Net Framework

The framework follows the workflow:

Dataset → Preprocessing → Data Augmentation → Transfer Learning → Adaptive Fine-Tuning → Softmax Classification → Performance Evaluation

3.2 Dataset Preparation

The proposed framework utilizes the Benchmark Dataset, which contains 87,867 RGB images representing 38 healthy and diseased plant leaf classes across 14 crop species [13], [15]. The dataset was divided into 70,295 training images and 17,572 validation images, ensuring identical data distribution for all investigated CNN architectures.

Table 2. Dataset Summary

Split	Images	Classes
Training	70,295	38
Validation	17,572	38
Total	87,867	38

3.3 Image Preprocessing and Data Augmentation

Before training, all RGB images were resized to 224 × 224 × 3 pixels and normalized using the architecture-specific preprocessing function provided by TensorFlow. Label vectors were transformed into one-hot encoded representations suitable for multi-class classification [18]–[21].

The preprocessing operation is expressed as

$$I_n = f(I)$$

where

- I denotes the input RGB image,
- $f(\cdot)$ represents the preprocessing function,
- I_n denotes the normalized image.

To improve model robustness and reduce overfitting, online data augmentation was applied during training. Four augmentation operations were adopted, as summarized in Table 3.

Table 3. Online Data Augmentation

Operation	Parameter
Horizontal Flip	Random
Rotation	±15°
Zoom	±15%
Contrast	±10%

Unlike offline augmentation, these transformations were generated dynamically during each training epoch, thereby increasing data diversity without expanding the physical dataset [27], [30].

3.4 Transfer Learning Strategy

Training deep CNN models from random initialization requires extensive computational resources and large annotated datasets. Therefore, pretrained ImageNet models were adopted to transfer generic visual representations to the plant disease classification task [16], [17].

The updated trainable parameters are expressed as

$W = W_p + \Delta W$

where

W_p denotes pretrained ImageNet weights,,
ΔW represents the parameters optimized during fine-tuning.
The proposed framework employs four pretrained CNN architectures:

- MobileNetV2
- ResNet50
- DenseNet201
- EfficientNetB3

All architectures were trained using identical preprocessing procedures, optimizer settings, augmentation strategies, and evaluation metrics to ensure objective comparison.

3.5 Adaptive Fine-Tuning

A two-stage adaptive optimization strategy was employed.

Stage 1: Feature Extraction

The pretrained convolutional backbone was frozen while only the newly added classification layers were trained.

Stage 2: Fine-Tuning

Selected upper convolutional layers were subsequently unfrozen and retrained using a reduced learning rate. This strategy preserves generic visual representations while enabling effective adaptation to disease-specific features [22], [28].

3.6 Mathematical Formulation

Let
$D = (x_i, y_i)_{i=1}^{N}$
represent the benchmark plant leaf image dataset.
The Softmax classifier estimates the probability of assigning an input image to class c as

$$P(y = c|x) = \frac{e^{z_c}}{\sum_{k=1}^{K} e^{z_k}}$$

where K = 38 denotes the total number of disease classes.
The network is optimized using the categorical cross-entropy loss function

$$L = -\sum_{i=1}^{N}\sum_{c=1}^{K} y_{ic}\log(\widehat{y_{ic}})$$

where y_{ic} and $\hat{y}_{ic}$ represent the ground-truth and predicted class probabilities, respectively.

3.7 Mathematical Workflow of the Proposed UTLPD-Net Framework

The proposed UTLPD-Net can be mathematically represented as a sequence of transformation functions that progressively convert the input dataset into the final disease prediction. Let

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

denote the benchmark plant leaf image dataset, where $\boldsymbol{x_i}$ represents the input RGB image and $\boldsymbol{y_i}$ denotes its corresponding disease label. The overall workflow of the proposed framework is expressed as

$$\boldsymbol{M^*} = \mathbf{arg}\max_{\boldsymbol{M_i}}\left[\boldsymbol{\Omega}\left(\boldsymbol{S}\left(\boldsymbol{F}\left(\boldsymbol{T}(\boldsymbol{P}(\boldsymbol{D}))\right)\right)\right)\right]$$

subject to

$$\boldsymbol{M_i} \in \{\boldsymbol{M_{MobileNetV2}}, \boldsymbol{M_{ResNet50}}, \boldsymbol{M_{DenseNet201}}, \boldsymbol{M_{EfficientNetB3}}\}$$

where

- $\boldsymbol{P}(.)$ denotes the image preprocessing function,
- $\boldsymbol{T}(.)$ represents online data augmentation,
- $\boldsymbol{F}(.)$ denotes transfer learning with adaptive fine-tuning,
- $\boldsymbol{S}(.)$ represents Softmax-based multi-class classification,
- $\boldsymbol{\Omega}(.)$ denotes the comprehensive performance evaluation function,
- $\boldsymbol{M^*}$ represents the optimal CNN architecture selected by the proposed framework.

Accordingly, the complete transformation pipeline of UTLPD-Net is represented as

$$\boldsymbol{D} \xrightarrow{\boldsymbol{P}} \boldsymbol{D_p} \xrightarrow{\boldsymbol{T}} \boldsymbol{D_a} \xrightarrow{\boldsymbol{F}} \boldsymbol{M_i} \xrightarrow{\boldsymbol{S}} \widehat{\boldsymbol{Y}} \xrightarrow{\boldsymbol{\Omega}} \boldsymbol{M^*}$$

where

- $\boldsymbol{D_p}$ denotes the preprocessed dataset,
- $\boldsymbol{D_a}$ denotes the augmented dataset,
- $\boldsymbol{M_i}$ denotes the pretrained CNN model,
- $\widehat{\boldsymbol{Y}}$ denotes the predicted disease labels,
- $\boldsymbol{M^*}$ denotes the best-performing CNN architecture.

The performance evaluation function is defined as

$$\boldsymbol{\Omega} = \{\boldsymbol{Acc}, \boldsymbol{Pre}, \boldsymbol{Rec}, \boldsymbol{F1}, \boldsymbol{\kappa}, \boldsymbol{MCC}, \boldsymbol{ROC}, \boldsymbol{PR}, \boldsymbol{CT}\}$$

where

- $\boldsymbol{Acc}$ – Accuracy,
- $\boldsymbol{Pre}$ – Precision,
- $\boldsymbol{Rec}$ – Recall,

- $\boldsymbol{F1}$– F1-score,
- $\boldsymbol{\kappa}$ – Cohen's Kappa,
- $\boldsymbol{MCC}$ – Matthews Correlation Coefficient,
- $\boldsymbol{ROC}$ – Receiver Operating Characteristic analysis,
- $\boldsymbol{PR}$ – Precision–Recall analysis,
- $\boldsymbol{CT}$ – Computational Training Time.

Finally, the optimal architecture is selected according to

$$\mathbf{M}^{*} = \mathbf{argmax}_{(\mathbf{M_i})}(\mathbf{Acc_i}, \mathbf{F1_i}, \boldsymbol{\kappa}_\mathbf{i}, \mathbf{MCC_i})$$

while simultaneously minimizing computational complexity

$$\mathbf{CT_i} = \mathbf{min}\big(\mathbf{CT}_{(MobileNetV2)}, \mathbf{CT}_{(ResNet50)}, \mathbf{CT}_{(DenseNet201)}, \mathbf{CT}_{(EfficientNetB3)}\big).$$

Therefore, the proposed UTLPD-Net framework performs standardized preprocessing, augmentation, transfer learning, classification, and evaluation under identical experimental conditions, enabling objective comparison of multiple pretrained CNN architectures while improving reproducibility.

3.8 Novel Contribution

The novelty of UTLPD-Net lies in establishing a standardized benchmarking framework rather than proposing a new CNN architecture. Unlike previous studies, all investigated architectures are evaluated using identical benchmark dataset partitions, preprocessing methods, augmentation strategies, optimizer settings, fine-tuning procedures, and statistical evaluation metrics. This standardized methodology enables fair architectural comparison, improves reproducibility, and provides a reliable baseline for future plant disease detection research involving attention mechanisms, Vision Transformers, explainable artificial intelligence, and real-time smart agriculture applications [24]–[30].

4. Experimental Setup

The proposed UTLPD-Net framework was implemented to evaluate four pretrained convolutional neural network architectures—MobileNetV2, ResNet50, DenseNet201, and EfficientNetB3—under identical experimental conditions. All models were trained using the same dataset partitions, preprocessing procedures, augmentation techniques, optimization strategy, and evaluation metrics to ensure objective comparison and experimental reproducibility [16], [22], [28].

4.1 Experimental Environment

The experiments were conducted using Python 3.11, TensorFlow 2.20, and the Keras deep learning framework. Image processing and data handling were performed using OpenCV, NumPy, and Pandas, while performance evaluation and visualization employed Scikit-learn and Matplotlib [31]–[34].

Model training was carried out on a workstation equipped with an Intel Xeon Silver 4214 processor, 128 GB RAM, and an NVIDIA RTX 2080 Ti GPU (11 GB) running Windows 10 (64-bit). CUDA and cuDNN acceleration were enabled to improve computational efficiency.

4.2 Training Configuration

To ensure fair comparison, all pretrained CNN architectures were trained using identical hyperparameter settings. Images were resized to 224 × 224 × 3, and the Adam optimizer was adopted due to its fast convergence and stable optimization characteristics [17], [18]. Training was performed in two stages: (i) feature extraction with frozen backbone layers and (ii) adaptive fine-tuning of the upper convolutional layers using a reduced learning rate.

Table 4. Hyperparameter Configuration

Parameter	Value
Input Size	224 × 224 × 3
Batch Size	32
Optimizer	Adam
Initial Learning Rate	1×10^{-4}
Fine-Tuning Learning Rate	1×10^{-5}
Initial Epochs	8
Fine-Tuning Epochs	5
Loss Function	Categorical Cross-Entropy
Output Layer	Softmax
Random Seed	42

4.3 Performance Evaluation

Model performance was evaluated using multiple statistical measures to provide a comprehensive assessment of classification effectiveness. In addition to Accuracy, Precision, Recall, and F1-score, agreement-based measures such as Cohen's Kappa (κ) and Matthews Correlation Coefficient (MCC) were employed to evaluate prediction reliability [29], [35], [36]. Furthermore, Confusion Matrix, Receiver Operating Characteristic (ROC) curves, Precision–Recall (PR) curves, and training time were used to analyze classification behaviour and computational efficiency.

4.4 Experimental Workflow

The experimental workflow followed the standardized UTLPD-Net pipeline illustrated in Figure 1. After preprocessing and online augmentation, each pretrained CNN architecture underwent transfer learning and adaptive fine-tuning under identical training conditions. Finally, the trained models were evaluated using the selected statistical measures, and comparative analysis was performed to identify the most suitable architecture for automated plant disease classification.

5. Results and Discussion

The proposed UTLPD-Net framework was experimentally evaluated using four pretrained convolutional neural network architectures—MobileNetV2, ResNet50, DenseNet201, and EfficientNetB3—under identical preprocessing, augmentation, optimization, and evaluation settings. Performance was assessed using Accuracy, Precision, Recall, F1-score, Cohen's Kappa (κ), Matthews Correlation Coefficient (MCC), training time, Confusion Matrix, Receiver Operating Characteristic (ROC), and Precision–Recall (PR) analysis. The standardized implementation ensures that performance differences primarily reflect architectural characteristics rather than experimental variability [15], [16], [22], [28].

5.1 Training Behaviour Analysis

The convergence behaviour of all investigated CNN architectures is illustrated in Figure 2 and Figure 3, which present the combined training–validation accuracy and loss curves, respectively.

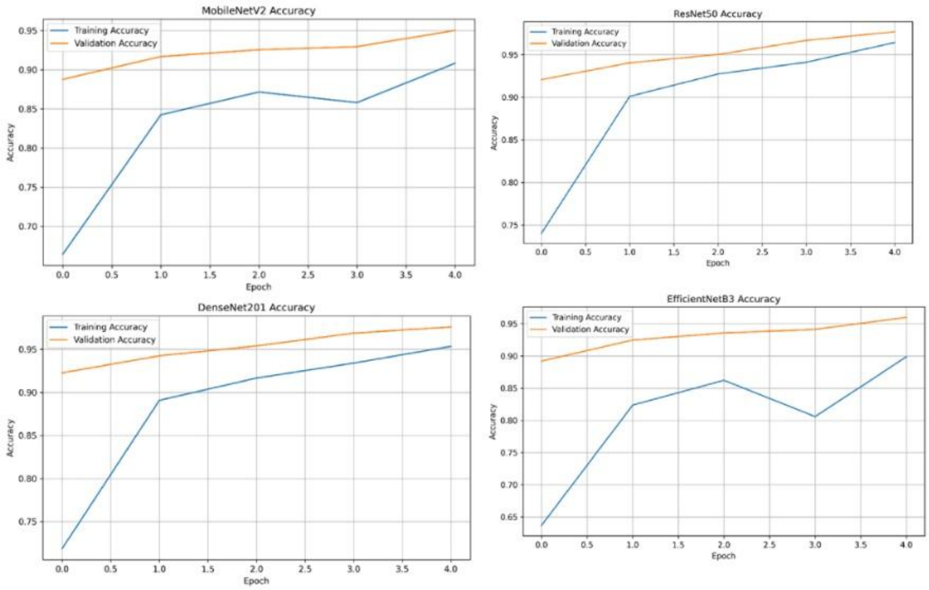


Figure 2. Combined Training and Validation Accuracy of MobileNetV2, ResNet50, DenseNet201, and EfficientNetB3

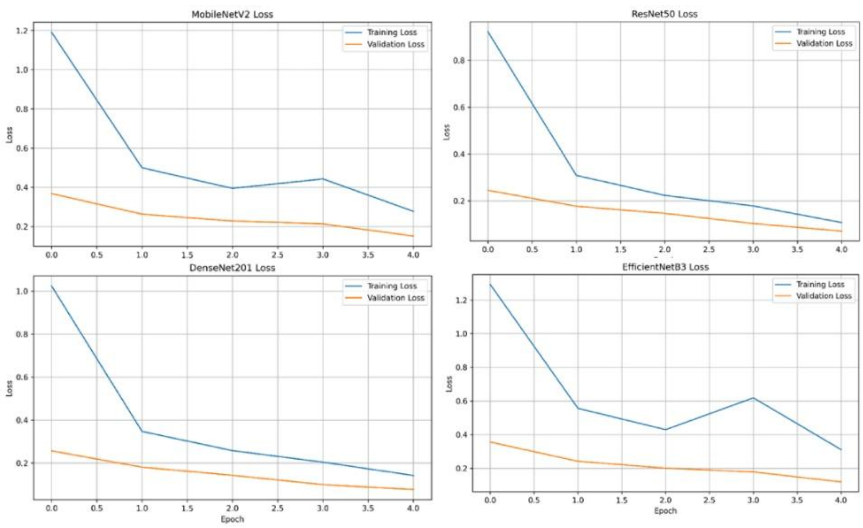


Figure 3. Combined Training and Validation Loss of MobileNetV2, ResNet50, DenseNet201, and EfficientNetB3

All models converged successfully within the specified training epochs. ResNet50 and DenseNet201 achieved faster convergence with higher validation accuracy, while MobileNetV2 converged rapidly with the shortest computational time. EfficientNetB3 exhibited stable optimization behaviour and provided a balanced trade-off between predictive performance and computational complexity. The relatively small gap between training and validation accuracy indicates that the proposed preprocessing and online augmentation strategy effectively reduced overfitting and improved generalization capability [18]–[21].

5.2 Comparative Performance Evaluation

The quantitative performance of all investigated CNN architectures is summarized in Table 5, while Figure 4 illustrates the comparative evaluation of the major statistical metrics.

Table 5. Comparative Performance of Pretrained CNN Architectures

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	κ	MCC	Training Time (min)
MobileNetV2	95.03	95.18	95.03	95.02	0.9466	0.9466	46.16
ResNet50	97.68	97.79	97.68	97.68	0.9762	0.9762	58.62
DenseNet201	97.59	97.69	97.59	97.60	0.9753	0.9753	82.24
EfficientNetB3	96.88	97.02	96.88	96.88	0.9679	0.9679	69.41

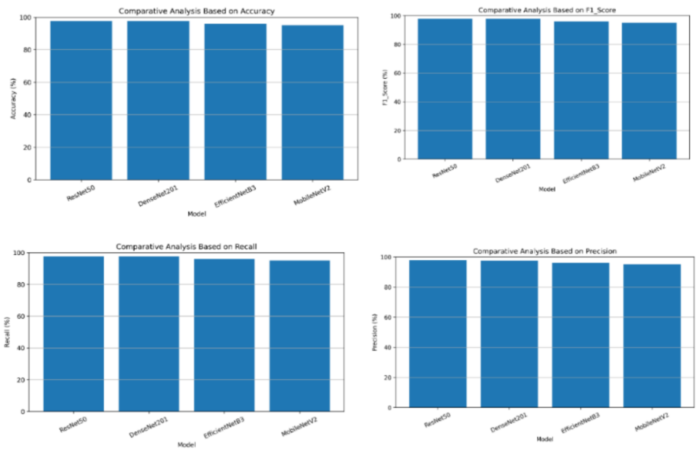


Figure 4. Comparative Performance of the Investigated CNN Architectures

The comparative evaluation demonstrates that ResNet50 achieved the highest overall predictive performance across all statistical measures. It recorded an overall classification accuracy of 97.68%, outperforming DenseNet201 by a small margin while requiring significantly less computational time. DenseNet201 produced comparable predictive capability but exhibited the highest computational cost owing to its dense connectivity. MobileNetV2 required the shortest training time because of its lightweight architecture, whereas EfficientNetB3 provided intermediate performance with balanced computational complexity.

The results indicate that residual learning enables ResNet50 to learn discriminative disease-specific visual features more effectively than the remaining architectures while maintaining computational efficiency suitable for practical deployment.

5.3 Confusion Matrix Analysis

To investigate class-wise prediction performance, the confusion matrix corresponding to the best-performing ResNet50 model is presented in Figure 5.

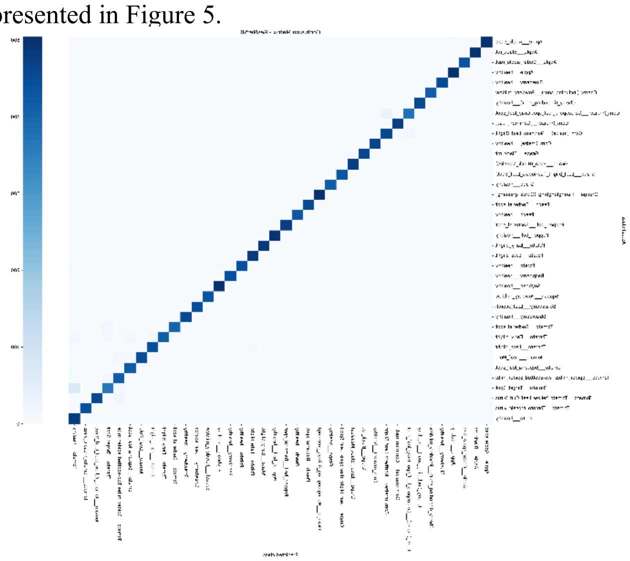
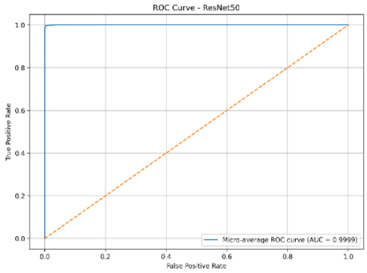


Figure 5. Confusion Matrix of ResNet50

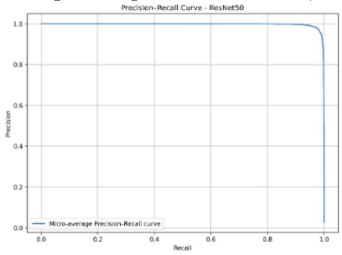
The confusion matrix exhibits strong diagonal dominance, indicating that the majority of leaf images were correctly assigned to their respective disease categories. Misclassifications occurred only among a limited number of visually similar diseases exhibiting comparable lesion colour and texture characteristics. The low frequency of off-diagonal elements demonstrates that the proposed preprocessing strategy, adaptive fine-tuning, and transfer learning framework effectively distinguish among the 38 healthy and diseased plant leaf classes [13]–[15].

5.4 ROC and Precision–Recall Analysis

The discriminative capability of the best-performing ResNet50 model was further evaluated using Receiver Operating Characteristic (ROC) and Precision–Recall (PR) analysis.



(a) Receiver Operating Characteristic (ROC) Curve



(b) Precision–Recall Curve

Figure 6. Performance Evaluation of the Proposed ResNet50 Model

The ROC curves remain close to the upper-left corner of the ROC space, indicating excellent discrimination capability across all disease classes. Similarly, the Precision–Recall curves remain near the upper-right corner, demonstrating consistently high precision and recall over different decision thresholds. The corresponding Area Under the Curve (AUC) and Average Precision (AP) values confirm that the proposed framework effectively minimizes both false-positive and false-negative predictions, thereby providing reliable multi-class disease classification [35], [37]–[41].

5.5 Comparative Discussion

The experimental results clearly demonstrate that all four pretrained CNN architectures successfully learned disease-specific visual representations from the benchmark dataset. Nevertheless, significant differences were observed in predictive accuracy, computational efficiency, and convergence behaviour.

Among the investigated architectures, ResNet50 consistently achieved the best overall performance. Its residual shortcut connections facilitate efficient gradient propagation during optimization, enabling deeper feature learning without suffering from degradation problems associated with conventional deep neural networks [19]. Consequently, ResNet50 produced the highest classification accuracy while maintaining moderate computational complexity.

Although DenseNet201 achieved comparable predictive accuracy, its dense connectivity considerably increased training time. In contrast, MobileNetV2 exhibited the fastest training speed because of its lightweight depthwise separable convolution, making it particularly suitable for deployment on mobile and embedded agricultural devices. EfficientNetB3 achieved a favourable balance between predictive performance and computational efficiency through compound scaling of network depth, width, and input resolution.

Compared with previous studies, the proposed UTLPD-Net differs by evaluating all pretrained architectures under identical preprocessing procedures, augmentation strategies, optimization parameters, and evaluation metrics. This standardized implementation improves reproducibility and enables objective comparison among CNN architectures, thereby establishing a reliable benchmarking framework for future intelligent plant disease diagnosis systems [15], [16], [22], [24]–[30].

6. CONCLUSION AND FUTURE WORK

This paper presented UTLPD-Net (Unified Transfer Learning Framework for Plant Disease Detection), a standardized benchmarking framework for comparative evaluation of pretrained convolutional neural network architectures. Unlike previous studies that employed heterogeneous preprocessing techniques, augmentation strategies, and training protocols, the proposed framework established a unified experimental pipeline that ensured fair architectural comparison and improved reproducibility. The framework integrates standardized image preprocessing, online data augmentation, adaptive transfer learning, and comprehensive statistical evaluation into a single workflow suitable for automated plant disease classification [15], [16], [22], [28].

The proposed framework was validated using the publicly available benchmark plant disease dataset, consisting of 87,867 RGB images representing 38 healthy and diseased plant leaf classes across 14 crop species. Four pretrained CNN architectures—MobileNetV2, ResNet50, DenseNet201, and EfficientNetB3—were evaluated under identical implementation conditions. Among the investigated models, ResNet50 achieved the best overall performance with 97.68% accuracy, 97.79% precision, 97.68% recall, 97.68% F1-score, 0.9762 Cohen's Kappa, and 0.9762 Matthews Correlation Coefficient, while maintaining competitive computational efficiency. These results demonstrate that residual learning effectively captures disease-specific visual features and provides an appropriate balance between predictive accuracy and computational cost [19], [35]–[39].

The standardized evaluation using Confusion Matrix, Receiver Operating Characteristic (ROC) analysis, Precision–Recall (PR) analysis, Cohen's Kappa, and Matthews Correlation Coefficient further confirmed the robustness of the proposed framework. By maintaining identical preprocessing, augmentation, optimization, and evaluation procedures across all investigated architectures, UTLPD-Net provides a reliable benchmarking methodology for comparative transfer learning research and establishes a reproducible baseline for intelligent agricultural disease diagnosis.

Although promising results were obtained, the present study is limited to the PlantVillage Augmented Dataset, which primarily contains laboratory-acquired images. Future research will extend the proposed framework using real-field datasets captured under varying environmental conditions. Additional investigations will incorporate attention mechanisms (CBAM and SE blocks), Vision Transformer (ViT)-based architectures, and Explainable Artificial Intelligence (XAI) techniques such as Grad-CAM, Grad-CAM++, and Score-CAM to improve both predictive performance and model interpretability. Furthermore, deployment of lightweight optimized models on smartphones, unmanned aerial vehicles (UAVs), and Internet of Things (IoT)-based precision agriculture platforms will be explored to support real-time crop disease monitoring and intelligent farming applications [24]–[30].

In summary, the proposed UTLPD-Net provides a robust, reproducible, and computationally efficient framework for comparative transfer learning-based plant disease detection. The standardized methodology presented in this work offers a strong foundation for future research and contributes toward the development of intelligent decision-support systems for sustainable precision agriculture.

Acknowledgement

The authors gratefully acknowledge the PlantVillage Project for providing the publicly available dataset used in this research. The authors also thank the developers of TensorFlow, Keras, and Scikit-learn for their open-source software libraries. The computational facilities provided by Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore, are sincerely acknowledged.

Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Data Availability Statement

The PlantVillage Augmented Dataset used in this study is publicly available through the official PlantVillage repository. Additional implementation details and trained models can be made available by the corresponding author upon reasonable request.

REFERENCES

[1] Food and Agriculture Organization of the United Nations (FAO), The State of Food and Agriculture 2023: Revealing the True Cost of Food to Transform Agrifood Systems. Rome, Italy: FAO, 2023.

[2] D. P. Hughes and M. Salathé, “An open access repository of images on plant health to enable the development of mobile disease diagnostics,” arXiv preprint arXiv:1511.08060, 2015.

[3] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, “MobileNetV2: Inverted residuals and linear bottlenecks,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, pp. 4510–4520.

[4] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778.

[5] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 4700–4708.

[6] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. 36th Int. Conf. Machine Learning (ICML), Long Beach, CA, USA, 2019, pp. 6105–6114.

[7] R. Pydipati, T. F. Burks, and W. S. Lee, “Identification of citrus disease using color texture features and discriminant analysis,” Computers and Electronics in Agriculture, vol. 52, no. 1–2, pp. 49–59, 2006.

[8] A. Camargo and J. S. Smith, “Image pattern classification for the identification of disease causing agents in plants,” Computers and Electronics in Agriculture, vol. 66, no. 2, pp. 121–125, 2009.

[9] S. Arivazhagan, R. N. Shebiah, S. V. Nidhyanandhan, and L. Ganesan, “Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features,” Agricultural Engineering International: CIGR Journal, vol. 15, no. 1, pp. 211–217, 2013.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.

[11] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” Frontiers in Plant Science, vol. 7, Art. no. 1419, 2016.

[12] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, “A comparative study of fine-tuning deep learning models for plant disease identification,” Computers and Electronics in Agriculture, vol. 161, pp. 272–279, 2019.
[13] K. P. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.
[14] A. G. Howard et al., “MobileNets: Efficient convolutional neural networks for mobile vision applications,” arXiv preprint arXiv:1704.04861, 2017.
[15] M. Brahimi, K. Boukhalfa, and A. Moussaoui, “Deep learning for tomato diseases: Classification and symptoms visualization,” Applied Artificial Intelligence, vol. 31, no. 4, pp. 299–315, 2017.
[16] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, “Deep neural networks based recognition of plant diseases by leaf image classification,” Computational Intelligence and Neuroscience, vol. 2016, Art. ID 3289801, 2016.
[17] M. S. Islam, A. Dinh, K. Wahid, and P. Bhowmik, “Detection of potato diseases using image segmentation and multiclass support vector machine,” International Conference on Instrumentation, Control and Information Technology (ICICT), 2017.
[18] J. Hu, L. Shen, and G. Sun, “Squeeze-and-Excitation Networks,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7132–7141.
[19] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional Block Attention Module,” in Proc. European Conf. Computer Vision (ECCV), 2018, pp. 3–19.
[20] A. Dosovitskiy et al., “An image is worth 16×16 words: Transformers for image recognition at scale,” in International Conference on Learning Representations (ICLR), 2021.
[21] Z. Liu et al., “Swin Transformer: Hierarchical vision transformer using shifted windows,” in Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), 2021, pp. 10012–10022.
[22] R. R. Selvaraju et al., “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in Proc. IEEE Int. Conf. Computer Vision (ICCV), 2017, pp. 618–626.
[23] J. Ristaino, P. K. Anderson, D. P. Bebber, et al., “The persistent threat of emerging plant disease pandemics to global food security,” Proceedings of the National Academy of Sciences, vol. 118, no. 23, 2021.
[24] S. Mahlein, “Plant disease detection by imaging sensors – Parallels and specific demands for precision agriculture and plant phenotyping,” Plant Disease, vol. 100, no. 2, pp. 241–251, 2016.
[25] TensorFlow Developers, TensorFlow 2.20 Documentation, Google, 2024.
[26] P. Micikevicius et al., “Mixed Precision Training,” International Conference on Learning Representations (ICLR), 2018.
[27] L. Perez and J. Wang, “The effectiveness of data augmentation in image classification using deep learning,” arXiv preprint arXiv:1712.04621, 2017.
[28] S. J. Pan and Q. Yang, “A survey on transfer learning,” IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, 2010.
[29] J. Cohen, “A coefficient of agreement for nominal scales,” Educational and Psychological Measurement, vol. 20, no. 1, pp. 37–46, 1960.
[30] G. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” Journal of Big Data, vol. 6, no. 60, 2019.
[31] A. Krizhevsky, “Learning multiple layers of features from tiny images,” University of Toronto, Technical Report, 2009.
[32] A. Howard et al., “Searching for MobileNetV3,” in Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), 2019.
[33] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016.
[34] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
[35] T.-Y. Lin et al., “Feature Pyramid Networks for Object Detection,” Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
[36] J. G. A. Barbedo, “Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification,” Computers and Electronics in Agriculture, vol. 153, pp. 46–53, 2018.
[37] A. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, and D. P. Hughes, “Deep learning for image-based cassava disease detection,” Frontiers in Plant Science, vol. 8, Art. no. 1852, 2017.
[38] A. Picon, A. Alvarez-Gila, M. Seitz, A. Ortiz-Barredo, J. Echazarra, and A. Johannes, “Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild,” Computers and Electronics in Agriculture, vol. 161, pp. 280–290, 2019.
[39] D. Singh, N. Jain, P. Jain, P. Kayal, S. Kumawat, and N. Batra, “PlantDoc: A dataset for visual plant disease detection,” in Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, 2020, pp. 249–253.
[40] T. Fawcett, “An introduction to ROC analysis,” Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006.
[41] J. Davis and M. Goadrich, “The relationship between Precision-Recall and ROC curves,” in Proceedings of the 23rd International Conference on Machine Learning, 2006, pp. 233–240.