Leaf Disease Detection using Convolutional Neural Network

Krupa Prasad K R.[1], Sunil Kumar R.[2], Harshitha B R.[3],

Manoj Kumar M.[4], Raghavendra Rajesh Yaragatti[5]

[1]Associate professor, [2-5]UG Students, Department of Electronics and Communications Acharya Institute of Technology, Bengaluru, India

E-mail: [1]krupaprasadkr@acharya.ac.in

Abstract

For the past two decades, the imbalance between food supply and population growth has been a major concern. Agriculture plays an important role in human development, and technological improvements have significantly contributed to this process. In this study, Convolutional Neural Networks (CNNs) will be utilized to identify plant leaf diseases based on leaf images. The objective is to develop an application that accurately classifies plant images as healthy or diseased. This will be achieved by collecting and preprocessing a dataset of damaged and healthy plant images under varying watering conditions. Globally, essential agricultural commodities such as tomatoes, cotton, paddy, etc. often experience price fluctuations due to supply and demand issues. Additionally, many farmers lack access to agricultural specialists for diagnosing and treating leaf diseases. To solve this issue, a low-cost image processing technique is developed to detect leaf diseases in tomato plants, an essential ingredient in Indian kitchens. Using the CNN models, farmers can compare images of diseased leaves and detect infections early, enabling timely intervention. This approach benefits both farmers and consumers by stabilizing the prices of, tomatoes as it is rapid, cost-effective, and applicable year-round.

Keywords: Farming, Leaf Disease Detection, Convolution Neural Network, Tomatoes.

1. Introduction

Modern agriculture irrigation systems employ cutting-edge technology to increase agricultural output, minimize labor costs, and ensure the most efficient use of water resources. These systems usually employ complex control systems, sensors, and weather data to regulate the flow of water, which can be performed through drip irrigation, sprinkler systems, or other means. Because moisture spreads many plant disease, modern irrigation technology can aid in disease prevention by minimizing the quantity of water that comes into touch with crops. This is important to help improve the agricultural production. Convolutional neural networks (CNNs) and other machine learning methods could be useful in enhancing the decision-making for this process.

CNNs, a type of deep learning system, that can be trained to distinguish the various characteristics associated with specific diseases and then be tested to detect fresh images as they are entered. This is performed by training the algorithm using a huge dataset of healthy and diseased plant images. Sensors can be spread around the field to enable disease detection using a CNN based on leaf images. Then, using a trained CNN, these images may be quickly assessed for disease symptoms.

The condition of the leaf on which a crop is grown is an essential factor influencing both the quantity and quality of tomato yield. An image-processing-based approach detects plant leaf diseases by analyzing visual cues on the leaf's surface. Several systems for evaluating plant health have been developed, all employing cameras and computer vision algorithms. However, a comprehensive computer-based diagnostic system for such disorders was never widely developed or adopted due to factors such as the high cost of agricultural consultants, the complexity of development environments, and the lack of affordability for small-scale farmers. Unlike other fields where a lack of understanding may affect progress, the presence of specific undesirable characteristics or discoloration on plant leaves has a direct global impact on tomato production.

Objectives

- Empower farmers for independent disease identification.
- Improve profitability in tomato cultivation.
- Enhance quality and quantity of tomato in markets.

2. Literature Review

India's agriculture sector, which accounted for 13.7% of GDP in 2013, is facing a decline due to shifting priorities away from farmers. Agriculture provides jobs and essential products but also contributes significantly to global warming. The country's climate is changing, and its unpredictability is causing issues for agricultural crops. Researchers have introduced innovative ideas and technological advancements to mitigate these challenges. The use of machine learning and deep learning has further enhanced agricultural decision-making, ultimately improving crop yield [9].

This study discusses the use of machine learning techniques like Naive Bayes, Decision Tree, K-Nearest Neighbor, Support Vector Machine, and Random Forest (RF) for maize plant disease detection using plant images. The RF algorithm achieved the highest accuracy of 79.23%, making it a potential preventive measure for farmers [1].

Bangladesh's economy heavily relies on agriculture, and due to its small population and developing nature, crop diseases cause significant losses. To combat this issue, a model has been developed to classify leaf diseases in five major production crops using computer vision techniques and convolutional neural networks. The model has a validation accuracy of 90.38%, helping farmers identify and treat diseases early on. The model aims to minimize plant damage, increase production, and reduce costs by allowing farmers to cure their crops at the right time [2].

New technologies are gaining attraction in agriculture, as 42% of production is lost due to plant leaf diseases. A new technique, using image pre-processing, segmentation, and feature extraction, has shown 98.56% accuracy in predicting plant leaf diseases, providing information on affected area, disease name, total accuracy, sensitivity, and elapsed time [3].

Plant diseases impact species growth, and early identification is crucial. Machine Learning models, particularly Deep Learning (DL), have shown potential for increased accuracy. DL architectures and visualization techniques are used to detect and classify plant diseases, with performance metrics used for evaluation. Research gaps are identified for greater transparency in disease detection [5]. They are a major obstacle to agricultural production and food quality, and early identification is crucial for global health. Traditional methods involve on-site visits by pathologists, but manual examinations are limited due to low accuracy and human resource limitations. To address these issues, a robust plant disease

classification system is introduced using a Custom CenterNet framework with DenseNet-77 as a base network. The method involves annotations, deep feature extraction, and a one-stage detector. Performance analysis using the PlantVillage Kaggle database confirms the method's effectiveness in identifying and classifying plant diseases [4].

This research presents a deep learning system for classifying and detecting plant leaf diseases using images from the Plant Village dataset. The system uses convolutional neural networks to classify 15 classes, including 12 for diseases detected by bacteria and fungi and 3 for healthy leaves. The system achieved excellent accuracy in training and testing, with 98.29% for training and 98.029% for testing [6].

This work proposes a Densely Connected Convolutional Networks (DenseNet)-based transfer learning method for detecting plant diseases on edge servers with augmented computing resources. A lightweight Deep Neural Networks (DNN) approach is proposed for IoT devices with constrained resources. The models are trained with different image sizes and tested on real-world datasets, demonstrating accurate disease detection using low computational resources [7].

This study examines recent advancements in plant disease detection and classification systems using Machine Learning and Deep Learning models. It analyzes over 45 peer-reviewed papers published between 2017-2020, focusing on the identification, recognition, and classification of plant diseases using these algorithms. The study identifies popular ML and DL algorithms, including SVM, NN, KNN, Naïve Bayes, AlexNet, GoogLeNet, and VGGNet, and their applications in plant disease categorization. The findings could aid researchers in recognizing plant diseases and increasing agricultural productivity [8,10].

3. Methodology

The development of a tomato leaf disease classification system using deep learning involves several stages. Figure 1 below shows the stages involved in tomato leaf disease classification.
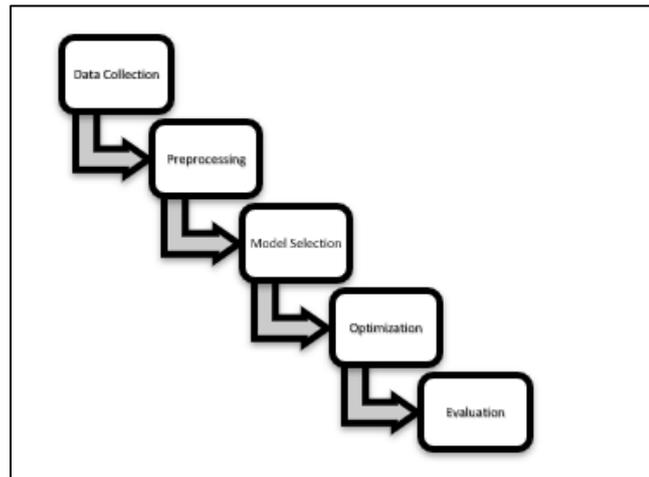
Figure 1. General Work Flow for Tomato Leaf Disease Classification

3.1 Dataset Description

The dataset used in this research is sourced from Kaggle (https://www.kaggle.com/datasets/kaustubhb999/tomatoleaf) and consists of 10 classes of tomato leaf diseases, including bacterial spot, early blight, late blight, leaf mold, septoria leaf spot, spider mites, target spot, tomato yellow leaf curl virus, tomato mosaic virus, and healthy leaves.

3.2 Preprocessing

The preprocessing phase for the "Tomato Leaf Disease Classification System" involves several essential steps to prepare the dataset for training CNN models. To ensure that the dataset images are compatible with the CNN models, VGG16, ResNet50, MobileNetV2, and EfficientNetB0 all images are resized to 224x224 pixels, the standard input size for models. This resizing ensures uniformity across the dataset and allows it to be fed into the models without any issues. To improve the model's generalization capabilities and reduce overfitting, data augmentation techniques are applied. These include random rotations up to 30°, width and height shifts up to 20%, shear transformations, zooming up to 20%, and horizontal flipping. In addition to augmentation, the images are normalized by rescaling pixel values to the range [0, 1]. This is done by dividing the pixel values by 255, ensuring numerical stability and faster convergence during model training. The ImageDataGenerator class from Keras is utilized for this purpose. It automatically handles image loading, augmentation, and rescaling in real-time during model training. Moreover, since this is a multi-class classification problem, the labels for each image are encoded using one-hot encoding, where

each class label is represented as a binary vector. These preprocessing steps ensure the dataset is well-suited for training the CNN models, leading to better performance and generalization.

3.3 Model used

The proposed work uses four different CNN models (VGG16, ResNet50, MobileNetV2, and EfficientNetB0) to detect the disease in tomato leaves. The VGG16 model, known for its simple yet deep architecture, incorporates convolutional layers followed by max-pooling layers, with a final fully connected layer of 512 neurons and a softmax activation function for multi-class classification. Similarly, ResNet50, a deep residual network, utilizes skip connections (residual learning) to mitigate the vanishing gradient problem, enhancing its ability to learn complex hierarchical features. The EfficientNetB0 model employs compound scaling techniques, balancing network depth, width, and resolution, resulting in a computationally efficient architecture. Lastly, MobileNetV2, designed for lightweight applications, incorporates depthwise separable convolutions to reduce computational complexity while maintaining high accuracy. The Table.1 below shows the hyperparameters used for training the models.

Table 1. Hyperparameter Used

Parameter	Value
Input Shape	(224, 224, 3)
Batch Size	32
Learning Rate	0.0001
Optimizer	Adam
Loss Function	Categorical Crossentropy
Dropout Rate	0.3-0.5
Epochs	25

4. Result and Discussion

The Tomato Leaf Disease Classification System utilizes Google Colab as the primary development environment, equipped with essential Python libraries such as TensorFlow with Keras for model development, Pandas for data manipulation, OpenCV for image preprocessing, and Scikit-learn for evaluation metrics. Pre-trained models like VGG16, ResNet50, MobileNetV2, and EfficientNetB0 from Keras applications are employed for transfer learning. The system leverages NVIDIA GPUs for accelerated training, with 16GB

RAM and SSD storage ensuring efficient processing. Hyperparameter tuning is performed using Keras Tuner. The Table .2 below shows the performance of the four models in leaf disease detection.

Table 2. Performance Comparison

Models	Precision	Recall	F1 Score	Accuracy
VGG 16	0.92	0.90	0.91	93.4%
ResNet50	0.94	0.92	0.93	95.5%
MobileNetV2	0.91	0.89	0.90	92.78%
EfficientNetB0	0.95	0.93	0.94	96.18%

The performance of the models was evaluated based on precision, recall, F1 score, and accuracy. VGG16 achieved a precision of 0.92, recall of 0.90, and an F1 score of 0.91, with an accuracy of 93.4%. This indicates a reasonably balanced performance with good precision and recall, though it was slightly outperformed by other models. ResNet50 showed superior results, with a precision of 0.94, recall of 0.92, and an F1 score of 0.93, achieving an accuracy of 95.5%. It demonstrated a strong capability to correctly identify both positive and negative classes, making it a robust choice for the task. MobileNetV2, being lightweight and efficient, delivered slightly lower performance, with a precision of 0.91, recall of 0.89, and an F1 score of 0.90, yielding an accuracy of 92.78%. Although it performed well, it did not match the higher metrics of ResNet50 or EfficientNetB0, possibly due to its smaller size and the trade-off between accuracy and computational efficiency. Finally, EfficientNetB0 emerged as the optimal performer, recording a precision of 0.95, recall of 0.93, and an F1 score of 0.94, with the highest accuracy of 96.18%. Its superior performance suggests it effectively balances precision and recall while achieving the highest classification accuracy, making it the most suitable model for this task. the graph in Figure .2 and 3 shows the comparison of the CNN models.
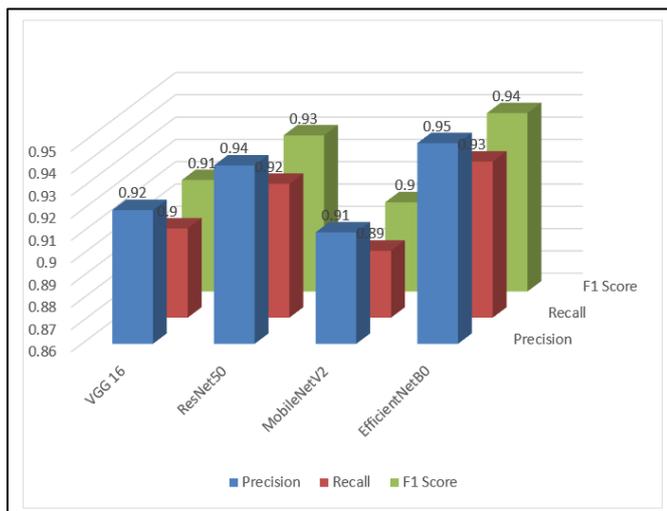
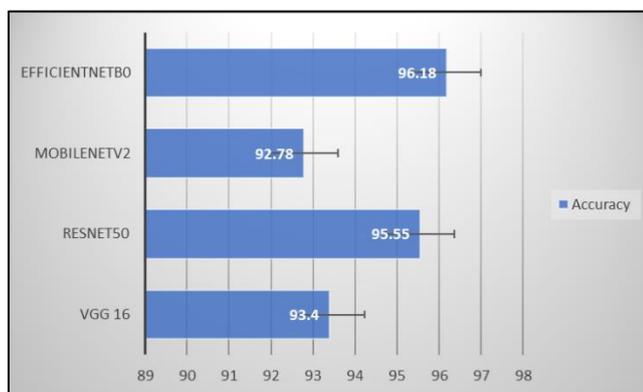**Figure 2.** Performance Comparison of Models

Figure 3. Accuracy of Models

4.1 Future Work

The future work aims to develop a user interface using the Flask application, so that the farmers could load the images of tomato leaves and know if the plant is healthy or not.

5. Conclusion

This study evaluated four deep learning models VGG16, ResNet50, MobileNetV2, and EfficientNetB0 for classifying tomato leaf diseases. Each model was assessed using precision, recall, F1 score, and accuracy. EfficientNetB0 emerged as the best performer, achieving the highest accuracy, precision, and recall. ResNet50 followed closely with strong performance, while VGG16 and MobileNetV2 delivered competitive results but were slightly outperformed. The findings highlight the trade-offs between model complexity,

computational efficiency, and classification accuracy, with EfficientNetB0 being the most suitable model for disease classification. Future work will focus on developing a user interface through a Flask application, enabling farmers to upload images of tomato leaves and receive real-time predictions on plant health, enhancing accessibility and practical use in agriculture.

References

[1] Panigrahi, Kshyanaprava Panda, Himansu Das, Abhaya Kumar Sahoo, and Suresh Chandra Moharana. "Maize leaf disease detection and classification using machine learning algorithms." In Progress in Computing, Analytics and Networking: Proceedings of ICCAN 2019, Springer Singapore, 2020. 659-669.

[2] Moyazzoma, Raida, Md Al Amin Hossain, Md Hasanuzzaman Anuz, and Abdus Sattar. "Transfer learning approach for plant leaf disease detection using CNN with pre-trained feature extraction method Mobilnetv2." In 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), IEEE, 2021. 526-529.

[3] Rakhra, Manik, Arun Singh, Gaurav Chawla, Jatin Rana, Aman Kumar, P. Teja, and Vikrant Sharma. "Plant Leaf Disease Detection Using Machine Learning." In 2023 1st International Conference on Circuits, Power and Intelligent Systems (CCPIS), IEEE, 2023. 1-6.

[4] Albattah, Waleed, Marriam Nawaz, Ali Javed, Momina Masood, and Saleh Albahli. "A novel deep learning method for detection and classification of plant diseases." Complex & Intelligent Systems (2022): 1-18.

[5] Saleem, Muhammad Hammad, Johan Potgieter, and Khalid Mahmood Arif. "Plant disease detection and classification by deep learning." Plants 8, no. 11 (2019): 468.

[6] Jasim, Marwan Adnan, and Jamal Mustafa Al-Tuwaijari. "Plant leaf diseases detection and classification using image processing and deep learning techniques." In 2020 International Conference on Computer Science and Software Engineering (CSASE), IEEE, 2020. 259-265.

[7] Ale, Laha, Alaa Sheta, Longzhuang Li, Ye Wang, and Ning Zhang. "Deep learning based plant disease detection for smart agriculture." In 2019 IEEE Globecom Workshops (GC Wkshps), IEEE, 2019. 1-6.

[8] Applalanaidu, Majji V., and G. Kumaravelan. "A review of machine learning approaches in plant leaf disease detection and classification." In 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), IEEE, 2021. 716-724.

[9] Jackulin, C., and S. Murugavalli. "A comprehensive review on detection of plant disease using machine learning and deep learning approaches." Measurement: Sensors 24 (2022): 100441.

[10] Ferentinos, Konstantinos P. "Deep learning models for plant disease detection and diagnosis." Computers and electronics in agriculture 145 (2018): 311-318.