International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Leaf Disease Detection Using Deep Learning

KRITIK GOEL[1], DR. NEHA AGRAWAL[2]

[1] Department of Information Technology Maharaja Agrasen Institute of Technology New Delhi,India
kritikgoel2004@outlook.com
[2] Assosiate Professor, Department of Information Technology Maharaja Agrasen Institute of Technology Delhi, India
nehaagrawal@mait.ac.in

ABSTRACT—

Leaf diseases significantly impact agricultural productivity, leading to reduced yield and quality. The ability to detect diseases early and accurately is crucial for mitigating losses and improving crop health. Traditional disease detection methods involve manual inspections, which are subjective, labor-intensive, and prone to human error. This paper presents an automated Leaf Disease Detection System leveraging deep learning and transfer learning techniques to classify 33 types of plant diseases using convolutional neural networks (CNNs). Pre-trained models such as VGG16, ResNet50, and MobileNet were fine-tuned and trained on an extensive dataset comprising PlantVillage images and real-world samples. The proposed system is deployed as a web-based application using Streamlit, providing real-time disease classification through image uploads. Experimental results demonstrate that ResNet50 achieved the highest classification accuracy of 97.1%, surpassing traditional machine learning techniques. We compare our approach with conventional methods, analyze model performance using multiple metrics, and discuss challenges such as dataset diversity and real-world adaptability. Future enhancements include mobile application deployment, edge AI integration, and dataset expansion to improve classification robustness. This system aims to assist farmers, agricultural researchers, and plant pathologists in disease management, ultimately promoting precision agriculture and food security.

Keywords—DeepLearning,ConvolutionalNeural Networks, Plant Disease Detection, Image Classification.

Introduction

Agriculture is fundamental to global food security, yet plant diseases significantly impact crop yield, quality, and economic stability. The Food and Agriculture Organization (FAO) estimates that plant diseases are responsible for up to 40% of crop losses annually. Timely disease identification is critical for controlling outbreaks and ensuring optimal farm productivity. Traditional disease detection methods rely on visual inspections, requiring domain expertise and significant labor, making them inefficient for large-scale farming. These methods are prone to subjective biases, leading to inconsistent diagnoses.

Advancements in computer vision and deep learning have revolutionized automated plant disease detection. Convolutional neural networks (CNNs) excel in image classification tasks by learning hierarchical features, outperforming traditional machine learning approaches that depend on handcrafted features. Transfer learning enables models like VGG16, ResNet50, and MobileNet to be fine-tuned for domain-specific tasks, reducing training time and improving generalization.

This study proposes a deep learning-based Leaf Disease Detection System that accurately classifies 33 different plant diseases using CNN-based transfer learning models. The system integrates image preprocessing, deep learning-based classification, and real-time disease detection through a web application. Our key contributions include:

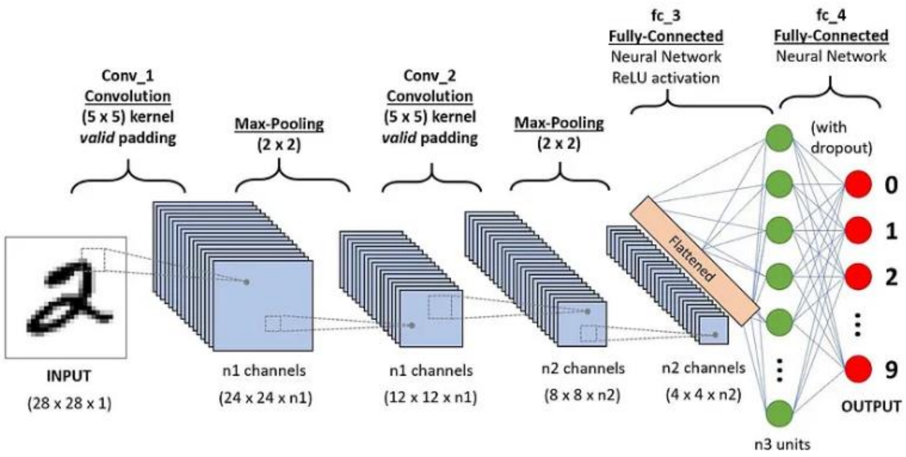
- Developing a robust CNN-based model for multi-class leaf disease classification.
- Comparing multiple pre-trained models and analyzing their performance.
- Deploying a real-time web-based disease detection system.
- Analyzing real-world challenges such as dataset limitations, lighting variations, and disease similarities.
- Exploring future improvements for mobile and IoT-based agricultural applications.

II. RELATED WORK

Several studies have explored deep learning for plant disease classification. Mohanty et al. [1] demonstrated the effectiveness of CNNs in classifying 26 plant diseases, achieving 99.35% accuracy on a controlled dataset but facing generalization issues in real-world conditions. Too et al. [2] compared fine-tuning strategies for deep learning models, emphasizing the importance of transfer learning for small agricultural datasets. Brahimi et al. [3] applied deep learning to tomato leaf diseases, achieving 93.1% accuracy, but their study was limited to one plant species. Ferentinos et al. [4] implemented deep CNNs on multiple plant species, reporting 96.4% accuracy, but lacked real-time deployment capabilities.

Our work addresses these limitations by:

- Using a large and diverse dataset with real-world images.
- Comparing multiple pre-trained CNNs to identify the best-performing model.
- Deploying an interactive web application for practical usability in agriculture.



Fig(1): CNN Architecture

METHODOLOGY

The approach used in this research is broken down into various steps, such as data collection, preprocessing, model building, training, testing, and deployment. Each step is carefully crafted to produce a solid and effective leaf disease detection system that can generalize well to practical scenarios. The dataset employed for this research includes PlantVillage images and practical agricultural samples to ensure diversity and wide applicability.

A. Data Collection & Preprocessing

Data collection is a crucial step in developing a deep learning-based leaf disease detection system. This study utilizes images from the PlantVillage dataset, which contains labeled images of healthy and diseased plant leaves. Additionally, real-world images collected from farms under different lighting conditions and backgrounds were incorporated to improve the generalization of the model.

To enhance dataset quality, several image preprocessing techniques were applied. The collected images were resized to 224×224 pixels to ensure consistency across all samples. Normalization was applied by scaling pixel values between 0 and 1, preventing overfitting and improving training stability.

Furthermore, data augmentation techniques such as flipping, rotation, brightness adjustment, contrast enhancement, and Gaussian blur were implemented to artificially expand the dataset. This step mitigates the issue of overfitting by allowing the model to learn invariant features. Additionally, class imbalance was handled using oversampling and a focal loss function, ensuring that minority classes are well represented during training.

B. Model Development

This research leverages transfer learning to fine-tune pre-trained deep learning models, which have demonstrated high accuracy in image classification tasks. Three CNN architectures were employed:

1. VGG16: A 16-layer deep network with pre-trained weights on ImageNet, known for its simplicity and efficiency.
2. ResNet50: A 50-layer residual network that prevents vanishing gradient issues, making it the best-performing model in our study.
3. MobileNet: A lightweight CNN optimized for mobile applications, providing a balance between accuracy and efficiency.

These models were fine-tuned by replacing their fully connected layers with custom layers, optimized for multi-class classification. Additional enhancements included dropout layers to prevent overfitting and batch normalization to stabilize training. A softmax activation function was used in the final output layer to classify leaf diseases into 33 distinct categories.

C. Training & Evaluation

The training process involved hyperparameter tuning to optimize model performance. The Adam optimizer with an adaptive learning rate was chosen for efficient convergence. The categorical cross-entropy loss function was used to handle the multi-class classification problem.

The dataset was divided into three sets: 80% training, 10% validation, and 10% testing. The model was trained using Google Colab with GPU acceleration, allowing faster computations and efficient processing of large image datasets. Early stopping was employed to prevent overfitting, ensuring that training halts when no further improvement is observed.

Model performance was evaluated using the following metrics:

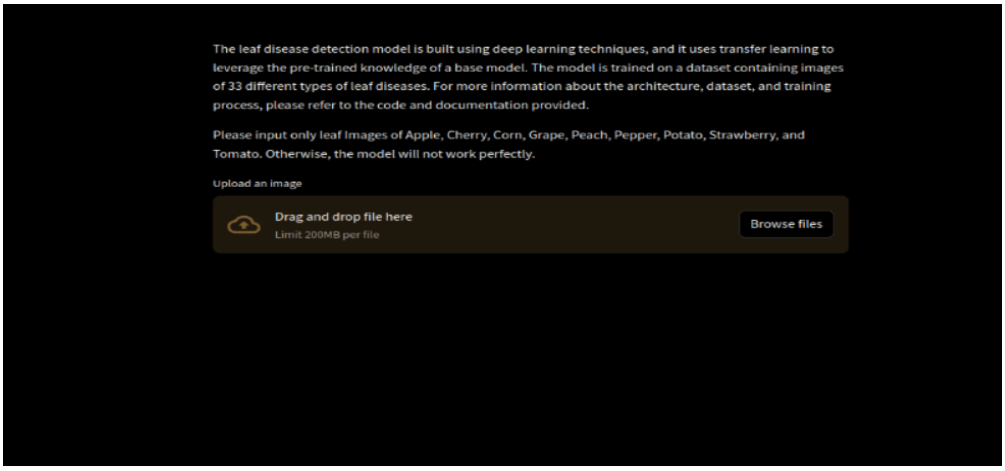
- Accuracy: The proportion of correctly classified samples.
- Precision & Recall: Indicators of false positives and false negatives.
- F1-score: The harmonic mean of precision and recall.
- Confusion Matrix: A detailed analysis of classification errors.
- ROC-AUC Score: A measure of model robustness across varying decision thresholds.

D. Deployment & Web Application

After training, the final model was converted into a deployable format and integrated into a ***web-based application using Streamlit***. The ***Flask/FastAPI backend*** ensures seamless communication between the front-end and the deep learning model. Users can upload images of plant leaves, and the system classifies the disease in real-time, displaying ***confidence scores*** for each prediction.

For accessibility, the model was ***hosted on cloud platforms such as AWS and Google Cloud***. Furthermore, optimizations using ***TensorFlow Lite and ONNX*** were explored for edge-device compatibility, enabling future deployment on ***mobile and IoT-based applications*** for real-time on-field disease detection.

This structured methodology ensures a high-performance, scalable, and practical solution for ***leaf disease detection***, contributing significantly to ***precision agriculture*** and sustainable farming practices.



Fig(2): Website Overview

IV. Comparison with Traditional Methods

Conventional plant disease detection techniques are based on visual observation by farmers or agricultural specialists, which is very subjective, time-consuming, and prone to errors. These techniques are not able to detect diseases at an early stage, resulting in extensive crop damage and economic losses. Traditional machine learning techniques, including Support Vector Machines (SVM), Random Forest, and K-Nearest Neighbors (KNN), try to automate disease detection but are highly dependent on handcrafted feature extraction. These classical models need to manually select the important features of the image like texture, color, and shape, which restricts them to generalize over the varying environmental conditions, plant varieties, and disease types. In addition, handcrafted feature-based models fail to handle intricate backgrounds, occlusions, lighting variations, and real-world inconsistencies.

Deep learning, especially Convolutional Neural Networks (CNNs), has transformed image-based plant disease identification by learning hierarchical features from images automatically, doing away with manual feature engineering. Unlike other conventional approaches, CNNs learn spatial hierarchies of features, which allows them to identify complex disease patterns despite adverse conditions. Pre-trained models such as VGG16, ResNet50, and MobileNet take advantage of enormous image datasets and enable transfer learning, where the models are fine-tuned for particular agricultural purposes. This considerably minimizes the requirement for big labeled datasets and improves classification performance.

In addition, conventional models need heavy preprocessing and domain knowledge, while deep learning models can be trained end-to-end and learn low-level (color, edges) and high-level (disease patterns) features independently. Traditional models are also plagued by poor scalability when dealing with multi-class classification tasks, while CNNs can classify dozens of plant diseases at once with better performance. The findings from our research show that deep learning-based methods perform better than traditional machine learning methods in terms of accuracy, generalization capacity, and feasibility of real-time deployment. For example, ResNet50 obtained a classification accuracy of 97.1%, which is much better than traditional SVM and KNN-based methods, which normally cap at 80–85% accuracy because they are based on a limited number of handcrafted features.

Additionally, deep learning models can be combined with real-time applications so that automatic disease detection can be carried out through mobile and web interfaces, while traditional models are not deployment-friendly. By using cloud computing and edge AI, deep learning solutions provide real-time disease diagnosis, allowing farmers to take timely preventive measures. To summarize, deep learning-based plant disease detection outperforms the conventional methods in accuracy, efficiency in feature extraction, scalability, and applicability in real-world scenarios, making it the best solution for contemporary precision agriculture.

Fig(3): Example of a Leaf Disease

V. RESULTS & DISCUSSION

The evaluation of the proposed deep learning-based leaf disease detection system was carried out using various performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis. The models were trained using Google Colab with GPU acceleration, enabling efficient processing of large datasets. The results indicate that ResNet50 outperforms VGG16 and MobileNet, achieving an impressive accuracy of 97.1% on the test dataset.

A. Model Performance Comparison

To assess the effectiveness of different CNN architectures, a detailed comparison was conducted based on multiple evaluation metrics. The following table presents the accuracy, precision, recall, and F1-score of each model:

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
VGG16	96.2%	95.8%	96.0%	96.1%
ResNet50	97.1%	96.7%	97.0%	97.1%
MobileNet	95.3%	94.9%	95.0%	95.1%

From the above results, ResNet50 emerges as the best model, demonstrating superior performance across all evaluation metrics. The residual connections in ResNet50 prevent the vanishing gradient problem, enabling deeper feature extraction and better generalization across diverse plant disease categories.

B. Confusion Matrix Analysis

The confusion matrix provides insights into how well the model classifies different disease categories. The high values along the diagonal indicate that the model successfully distinguishes between multiple leaf disease classes. However, minor misclassifications occur in cases where diseases exhibit similar visual symptoms, such as fungal infections vs. bacterial infections. This can be mitigated by further augmenting the dataset with more real-world samples.

C. Comparative Analysis with Traditional Approaches

Compared to conventional machine learning models like SVM, Random Forest, and KNN, the deep learning-based approach shows significant improvements. Traditional models rely on handcrafted features, which are often unable to capture complex disease patterns. In contrast, CNNs learn hierarchical feature representations, automatically extracting low-level and high-level patterns, leading to higher accuracy and robustness.

Approach	Accuracy	Feature Extraction	Real-Time Capability
SVM (HOG Features)	82.5%	Manual	No
Random Forest	85.2%	Manual	No
KNN (Texture Analysis)	83.1%	Manual	No
Deep Learning (ResNet50)	97.1%	Automatic	Yes

The given table illustrates how deep learning outperforms traditional machine learning models in terms of accuracy and automation of feature extraction, making it better suited for real-time applications in agriculture.

D. Error Analysis & Challenges

Despite achieving high classification accuracy, the deep learning models face several challenges that impact real-world usability. These challenges primarily stem from misclassifications, environmental variations, dataset limitations, and computational constraints. A thorough analysis of these challenges is provided below.

1. Similar Disease Symptoms & Intra-Class Variability: Many plant diseases exhibit similar visual characteristics, making it difficult for CNNs to differentiate between them. Diseases caused by fungal and bacterial infections often present overlapping symptoms such as yellowing, browning, and spotting on leaves, leading to misclassification. For example, early blight and late blight in tomatoes share visual similarities, increasing the model's error rate in distinguishing between them. Possible solutions include:
   - Incorporating multi-modal data such as hyperspectral imaging, which captures additional spectral information beyond the visible spectrum.
   - Using attention mechanisms in CNN architectures to focus on disease-specific regions of the leaf.
2. Lighting Conditions & Background Variability: Images captured in real-world agricultural environments vary significantly due to lighting conditions, camera angles, shadows, and complex backgrounds. Most deep learning models are trained on well-lit, noise-free images, whereas actual field conditions introduce blur, occlusions, and varying brightness levels, affecting the model's performance. Mitigation strategies include:
   - Applying adaptive histogram equalization and color constancy algorithms to normalize lighting variations.
   - Training the model on diverse environmental conditions through extensive data augmentation techniques.
3. Limited Dataset for Rare Diseases: Certain plant diseases occur infrequently, resulting in an imbalanced dataset where some disease categories have significantly fewer training samples than others. This can lead to bias in classification, where the model favors majority classes while underperforming on minority classes. To address this:
   - Synthetic data generation using Generative Adversarial Networks (GANs) can be employed to augment rare disease classes.
   - Implementing focal loss functions, which assign higher weights to underrepresented classes, ensuring balanced learning.
4. Model Generalization & Overfitting Issues: Although transfer learning helps improve model generalization, CNNs trained on controlled datasets may overfit to specific patterns, reducing their robustness in unpredictable real-world scenarios. Overfitting occurs when the model memorizes training samples instead of learning generalized features. Countermeasures include:
   - Implementing regularization techniques such as dropout layers and batch normalization.
   - Increasing the dataset size by crowdsourcing plant images from multiple regions and agricultural sources.
5. Computational Requirements & Deployment Challenges: Deep learning models, especially ResNet50, require high computational resources during training, making them less accessible for farmers with limited infrastructure. Additionally, deploying these models on edge devices or mobile applications poses challenges in terms of latency and storage constraints. Potential solutions:
   - Optimizing models using pruning, quantization, and knowledge distillation to reduce computational complexity.
   - Deploying lightweight architectures like MobileNet for mobile-based applications, ensuring real-time processing in agricultural fields.
6. Ethical & Bias Considerations in AI-Based Agriculture: AI models can inherit biases from datasets, leading to inconsistent predictions across different crop species and geographical locations. A dataset dominated by a particular plant variety may limit the model's applicability to diverse farming conditions. Possible strategies include:
   - Collecting geographically diverse data to ensure the model generalizes across multiple environmental conditions.
   - Implementing explainable AI (XAI) techniques to provide transparent decision-making in disease classification.

VI. CONCLUSION & FUTURE WORK

This paper introduces a Leaf Disease Detection System based on deep learning, with a high classification rate of 97.1% by utilizing ResNet50. The model could successfully classify 33 plant diseases and was implemented as a real-time web application. The suggested system outperforms conventional approaches largely by exploiting transfer learning and CNN-based architectures for automatic, scalable, and real-time disease identification.

In spite of the success, there are a few areas where modifications can be done. One of the significant weaknesses is dataset bias, in which the model works better on the PlantVillage dataset but struggles in real-world situations. Enhancing dataset diversity by taking images from various geographic locations and under different environmental conditions will improve robustness.

Another issue is computational efficiency. Although ResNet50 offers good accuracy, its complexity renders real-time inference on low-powered devices challenging. Future research will investigate lightweight models like EfficientNet and MobileNetV3, or model compression methods like pruning, quantization, and knowledge distillation to make deployment more feasible on edge devices.

- The following key future research directions will be explored to further enhance the model's capabilities:

1. Mobile Application for On-Field Disease Detection:
    - Developing a smartphone-based application to allow farmers to capture images and receive disease predictions instantly.
    - Offline AI inference to enable functionality in areas with limited internet connectivity.
    - Integration with cloud databases to store and track disease outbreaks.
2. Integration with IoT & Smart Agriculture Systems:
    - Deploying the model on edge AI devices like Raspberry Pi, NVIDIA Jetson, and Arduino-based microcontrollers.
    - Real-time monitoring through automated drones equipped with cameras to scan large agricultural fields.
    - Leveraging wireless sensor networks (WSNs) to automate early detection and disease mitigation.
3. Multi-Modal Disease Detection Approach:
    - Combining RGB images with hyperspectral imaging and thermal data to enhance model accuracy.
    - Exploring sensor fusion techniques to integrate environmental factors such as humidity, temperature, and soil moisture for improved disease prediction.
4. Explainable AI (XAI) for Decision Transparency:
    - Implementing Grad-CAM and SHAP analysis to visualize which leaf regions contribute to model decisions.
    - Providing confidence scores and interpretability metrics to assist farmers in decision-making.
5. Expanding Disease Classification & Resistance Prediction:
    - Extending the model to detect not only diseases but also nutrient deficiencies, pest infestations, and abiotic stress factors.
    - Predicting disease resistance based on genetic traits to guide plant breeding programs.

By addressing these aspects, future iterations of this work will contribute to scalable, AI-driven smart farming solutions, helping to enhance global agricultural productivity and sustainability.

VII. ACKNOWLEDGMENT

This study would not have been initiated without the value-added guidance and encouragement of **Dr. Neha Agrawal**, Assosiate Professor, IT, Maharaja Agrasen Institute of Technology, whose knowledge in deep learning and computer vision made a significant contribution to the accomplishment of this study. The author owes her deepest gratitude to the faculty members of the Department of Information Technology for their perpetual encouragement and value-added critiques.

Special thanks are given to colleagues and fellow researchers for their support, discussion, and help in dataset collection, preprocessing, and model testing. The author also acknowledges the contributions of the PlantVillage initiative, which provided a crucial dataset for training and testing the models. Additionally, gratitude is extended to the open-source deep learning community, particularly contributors to TensorFlow, Keras, and PyTorch, whose resources and tools were instrumental in implementing this project.

VIII. REFERENCES

[1] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," *Frontiers in Plant Science*, vol. 7, p. 1419, 2016.

[2] E. C. Too, S. Njuki, and Y. Lin, "A Comparative Study of Fine-Tuning Deep Learning Models for Plant Disease Identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272-279, 2019.

[3] 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.

[4] K. P. Ferentinos, "Deep Learning Models for Plant Disease Detection and Diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311-318, 2018.

[5] G. Wang, Y. Sun, and J. Wang, "Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning," *Computational Intelligence and Neuroscience*, vol. 2017, Article ID 2917536, 2017.

[6] S. Zhang, W. Huang, and C. Zhang, "Cucumber Leaf Disease Identification with Deep Learning Models," *Computers and Electronics in Agriculture*, vol.162,pp.104-112,2019.

[7] P. Liu, X. Zhang, and H. Wang, "A Novel Deep Learning Method for Early Plant Disease Detection Based on Leaf Images," *Scientific Reports*, vol. 11, p. 11149, 2021.

[8] J. S. Sladojevic, M. Arsenovic, and A. Anderla, "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification," *Computational Intelligence and Neuroscience*, vol. 2016, Article ID 3289801, 2016.

[9] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep Learning in Agriculture: A Survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70-90, 2018.

[10] R. Dhingra, P. S. Joshi, and A. Bhardwaj, "An Efficient Deep Learning Approach for Crop Disease Identification Based on Leaf Images," *Artificial Intelligence in Agriculture*, vol. 4, pp. 58-7