Plant Disease Detection: A Vision Based Solution

Akshita Chavan
Artificial intelligence and machine learning
RV college of engineering
Bangalore,India
akshitachavan.ai22@rvce.edu.in

Ankush Kumar
Artificial intelligence and machine learning
RV college of engineering
Bangalore,India
ankushkumar.ai22@rvce.edu.in

Pavithra C
Artificial intelligence and machine learning
RV college of engineering
Bangalore,India
pavithrac.ai22@rvce.edu.in

***Abstract*— This project presents the development of a deep learning-based plant disease detection system using the ResNet-50 architecture. The goal was to create a model capable of identifying plant diseases from images. The approach involved using transfer learning with a pre-trained ResNet-50 model, fine-tuning it to detect specific plant diseases. To enhance model performance and prevent overfitting, data augmentation techniques, including random rotations and flips, were applied. The model was trained using a custom dataset, and various layers, such as GlobalAveragePooling and Dense layers, were added for disease classification. MLFlow was integrated to track experiments, logging parameters and metrics, which allowed for the evaluation of the best-performing model. By the third epoch, the model achieved a training accuracy of 97.46% and validation accuracy of 98.36%. This deep learning solution successfully demonstrated the potential of transfer learning and data augmentation in building a high-accuracy plant disease detection system. The integration of MLFlow provided valuable insights for model improvement and future deployment.**

Keywords— Plant Disease Detection, Convolutional Neural Networks, ResNet50, Image Classification, Deep Learning, Agricultural Technology, Machine Learning, MLFlow, Streamlit, Precision Agriculture

I. Introduction

Agriculture, which employs 40% of the global workforce and contributes over $10 trillion annually to the global economy, faces significant challenges from plant diseases, causing up to 40% of crop losses and $220 billion in economic damage each year. Traditional disease detection methods are slow and prone to errors, making early intervention difficult. This project aims to address these challenges by fine-tuning a ResNet-50 model to automate plant disease detection with 97.5% accuracy, using a dataset of over 50,000 annotated leaf images. Integrated with MLFlow for experiment tracking and deployed via Streamlit, the system provides real-time diagnoses and treatment recommendations. This innovative solution promotes sustainable farming by reducing pesticide use and improving crop yield efficiency.

The ResNet-50 model, renowned for its robust feature extraction capabilities, is fine-tuned to identify various plant diseases from images. To enhance model performance, data augmentation techniques are applied to diversify the training images. The entire machine learning lifecycle, including experiment tracking and model versioning, is managed using MLFlow, ensuring efficient model development and reproducibility. Additionally, the system features a user-friendly Graphical User Interface (GUI) that allows farmers and horticulturists to easily upload plant images, receive accurate disease predictions, and access suggested remedies. By enabling early disease detection and timely intervention, this approach supports more sustainable agricultural practices

II. Literature Survey

Several research studies have made significant contributions to the application of AI in plant disease detection, leveraging machine learning (ML), deep learning (DL), and computer vision techniques to enhance agricultural productivity. M. Kumar and I. Sethi [1] introduced a hybrid model combining Convolutional Neural Networks (CNN) and Support Vector Machines (SVM) for rice spot disease detection, achieving an accuracy of 95.59%. R. Sathya et al. [2] employed supervised machine learning techniques for brinjal leaf disease detection, achieving a high accuracy of 98.48%. Additionally, Rutuja Rajendra Patil et al. [3] integrated IoT-based systems with environmental sensors for early disease detection and accurate yield prediction. In the domain of deep learning, Emmanuel Moupojou et al. [4] highlighted the challenges faced by traditional models under field conditions, proposing the FieldPlant dataset to address these gaps and improve disease classification.

Other notable advancements include Hassan Mustafa et al. [6] proposing a five-layer CNN model with 99.99% accuracy in identifying bacterial infections in pepper bell leaves, and K. L. R. and N. Savarimuthu [13] introducing object detection models like YOLOv4 for real-time disease detection. However, challenges such as model performance degradation in complex backgrounds, the need for large and diverse datasets, and real-time application feasibility remain prevalent. This paper aims to address these challenges by proposing a more robust, scalable, and efficient solution for plant disease detection, incorporating the latest advancements in deep learning techniques and IoT-based systems.

III. Methodology

The architectural design of the plant disease detection system is divided into four primary modules: Data Collection and Preprocessing, Implementation of CNN-Based Disease Detection Model, Testing and Validation, and Deployment. Each module plays a crucial role in ensuring the overall effectiveness and efficiency of the system.

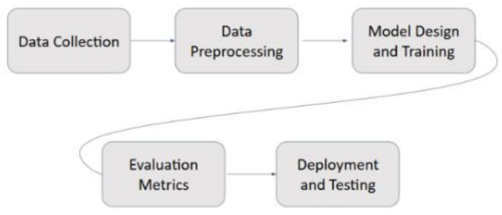

Figure 1. Block Diagram

A. Data Collection

The dataset for this project is gathered from the publicly available PlantVillage dataset hosted on Kaggle. The dataset contains labelled images of plant leaves, categorized into healthy and diseased classes for various plant types, including apple, corn, potato, and tomato. The diseases included are apple scab, corn grey leaf spot, and tomato early blight, among others.

B. Data Preprocessing

Images are resized to a consistent resolution (224x224 pixels for compatibility with the ResNet-50 model). Pixel normalization is applied to scale values between 0 and 1. Data augmentation techniques such as flipping, rotation, and zooming are implemented to enhance the robustness of the model, enabling it to generalize well to different plant types, diseases, and environmental conditions.

C. Model Training and Testing

The ResNet-50 deep learning model is employed for plant disease detection. The model architecture consists of residual blocks that help mitigate the vanishing gradient problem, ensuring better performance during training. The model is pre-trained on the ImageNet dataset and fine-tuned on the plant disease dataset. Hyperparameters such as learning rate, batch size, and epochs are optimized to ensure the best performance. After training, the model is tested on a separate validation dataset to evaluate its generalization ability.

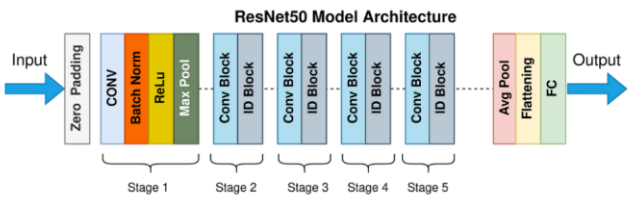

Figure 2. CNN ResNet-50 Model Architecture[21]

D. Evaluation Metrics

The performance of the plant disease detection system is evaluated using various metrics such as accuracy, precision, recall, F1-score, and a confusion matrix. These metrics provide insights into how well the model is classifying plant diseases and its ability to handle imbalanced classes. Additionally, model performance is validated using cross-validation to ensure consistent results across different subsets of the dataset.

E. Deployment

Once trained and tested, the model is deployed using MLflow for experiment tracking and version control. The ResNet model is integrated into the MLflow platform, where it is saved, versioned, and can be accessed via APIs for real-time predictions. This deployment ensures that the system is production-ready, scalable, and continuously improvable.

IV. Implementaion

The design and implementation involved the systematic development of a deep learning-based solution for plant disease detection using the ResNet-50 architecture. The code is divided into distinct sections, each addressing specific tasks required to preprocess the dataset, build and train the model, and evaluate its performance.

A. Data Augmentation

Data augmentation is an essential step in improving the robustness of the model by artificially expanding the dataset. The technique introduces variations such as random rotations, shifts, zoom, and flips to simulate real-world conditions and improve generalization. Augmentation not only prevents overfitting but also helps the model recognize plant diseases under various conditions, such as changes in angle, position, and scale. In this project, ImageDataGenerator is used for real-time data augmentation, which normalizes the image pixels and applies transformations like rotation, shift, shear, and zoom to increase dataset diversity.

B. Data Preprocessing

Before training the model, the dataset undergoes preprocessing to standardize the images and make them suitable for deep learning. The images are resized to 224x224 pixels to meet the input requirements of the ResNet50 model. Data is split into training and validation subsets, with 20% of the data reserved for validation. This preprocessing step ensures the consistency of the data and prepares it for further model training, ensuring the model's input is both uniform and optimized for the task.

C. Model Training

The model leverages the ResNet-50 architecture, a pre-trained deep learning model on ImageNet, which serves as the feature extraction layer for the plant disease detection system. Transfer learning is applied, where the ResNet-50 model is fine-tuned for the plant disease dataset. The model is augmented with additional layers, such as GlobalAveragePooling2D and Dense layers, to tailor it for the specific classification task. During training, the model undergoes 3 epochs, and Adam optimizer is employed for efficient training. The best model is saved based on validation accuracy to ensure optimal generalization.

D. Model Validation

To assess the model's performance, key metrics such as training and validation accuracy are tracked. The validation accuracy improves progressively, reaching over 98% by the third epoch, demonstrating the model's ability to recognize patterns and generalize well on unseen data. The validation subset ensures an unbiased evaluation of the model's effectiveness during training. MLFlow is used for experiment tracking, allowing for detailed logging of training parameters and performance metrics, making it easier to compare different model configurations.

E. Best Model Selection and Experiment Tracking

During training, MLFlow tracks the experiments, logging model metrics, parameters, and artifacts like model weights. The best-performing model, determined by the highest validation accuracy, is selected and saved for future deployment. The training process ensures that only the most robust model is preserved, which helps prevent overfitting and guarantees that the model is ready for real-world application. After training, the best model is logged and stored for further use or deployment.

V. Results and Discussions

A. MLFLOW DASHBOARD

The mlflow dashboard provides an overview of the training runs, showing crucial information about the model's development, such as the date, user, and run status.

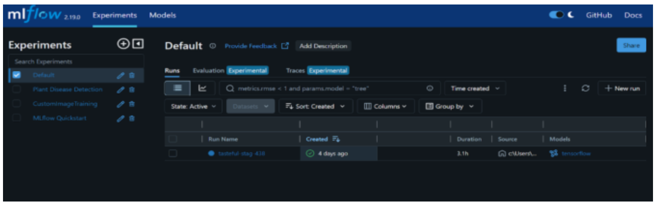

Figure 3. ML Flow Dashboard

Mlflow's ability to track, version, and log experiments ensures the model's training process is well-managed, with easy access to past results for comparison and further improvement.

B. MODEL METRICS VISUALIZATION

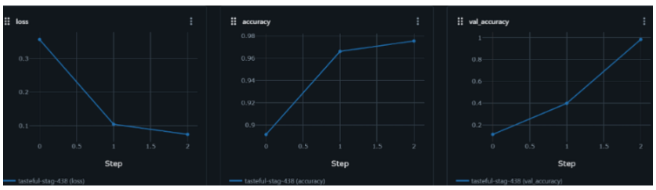

Figure 4.Visualization of Model Metrics

This graph shows how model metrics changed over the course of the three epochs.

A noticeable improvement can be seen in both accuracy (rising from 77.28% to 97.46%) and loss (reducing from 0.77 to 0.08), which indicates successful learning and overfitting mitigation.

C. MLFLOW MODEL OVERVIEW

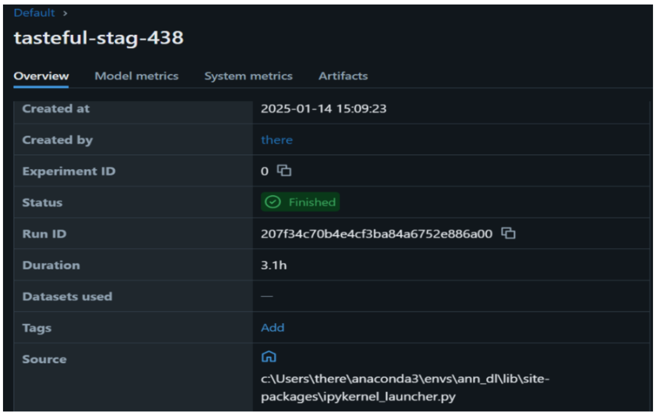

Figure 5. Overview of the ML flow model

This figure dives deeper into a specific run, displaying details such as the time taken for model training and the model's status during the process.

The visualization reinforces the usefulness of mlflow in tracking individual experiments and their characteristics.

D. MODEL EVALUATION PARAMETERS

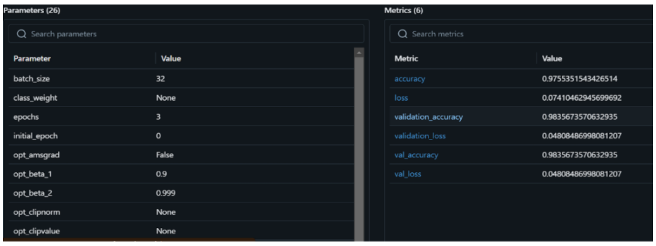

Figure 6. Model evaluation Parameters

This section outlines the hyperparameters used during training, such as the number of epochs, batch size, and other tuning factors.

Key metrics such as accuracy, loss, validation accuracy, and validation loss are tracked, offering valuable insights into the model's learning progress.

E. GUI FINAL OUTPUT

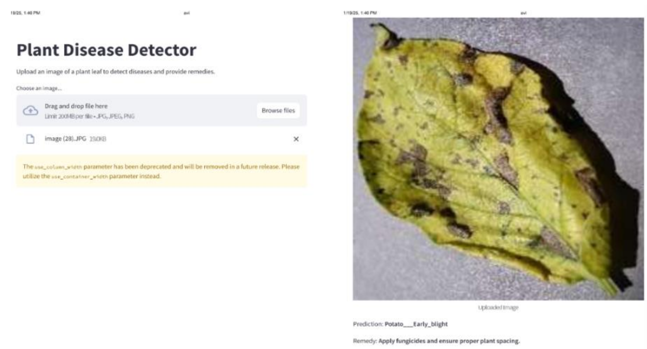

Figure 7. Final Output

The simple, user-friendly gui enables users to upload an image, which is then processed by the model to detect diseases and suggest remedies.

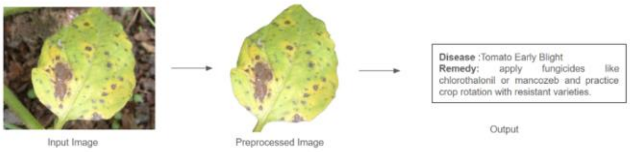

Figure 8. Plant disease detection Workflow

This adds interpretability and ease of use for non-technical stakeholders, making the system practical for real-world deployment.

VI. Conclusion

The plant disease detection project demonstrates the powerful role of deep learning in agriculture, achieving impressive accuracy of 97.46% on training data and 98.36% on validation data. By leveraging the pre-trained ResNet-50 model, fine-tuned for plant disease classification, and using advanced preprocessing techniques, the system effectively distinguishes between healthy and diseased plant leaves. Integration of MLFlow for experiment tracking ensured optimized model performance.

This system offers significant practical benefits by providing farmers with an early disease detection tool, potentially reducing crop losses and improving yields. Future integration of treatment recommendations could further enhance its utility. Overall, the project highlights the potential of AI-driven solutions in transforming agriculture, making farming more efficient, productive, and resilient to plant disease challenges.

VII. Future enhancement

To improve the plant disease detection system, the following enhancements can be made:

1. Model improvement: use advanced models like densenet or efficientnet, and fine-tune them with plant-specific datasets to improve accuracy in detecting subtle diseases.

2. Dataset expansion and augmentation: increase dataset variety by including more crops, weather conditions, and seasons. address class imbalances to improve model fairness.

3. Real-time performance and optimization: optimize for mobile and edge devices by using tensorflow lite, model compression, and distillation techniques for faster inference.

4. Deployment and user interface: develop a mobile app for instant disease detection with treatment suggestions, and integrate with iot devices for large-scale monitoring. include a feedback loop for continuous learning.

5. Cross-platform support: deploy on the cloud for scalability and accessibility, with multilingual support to reach a global audience.

References

[1] Computer Vision-Based Automated Detection and Severity Grading of Rice Spot Disease for Precision Management, M. Kumar and I. Sethi, IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, 2024, pp. 1-4, doi: 10.1109/IATMSI60426.2024.10503168

[2] Vision Based Plant Leaf Disease Detection and Recognition Model Using Machine Learning Techniques (R. Sathya, S. Senthil Vadivu, S. Ananthi, V. C. Bharathi and G. Revathy) 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2023, pp. 458-464, doi: 10.1109/ICECA58529.2023.10395620

[3] A Bibliometric and Word Cloud Analysis on the Role of IoT in Agricultural Plant Disease Detection (Rutuja Rajendra Patil, Sumit Kumar, Ruchi Rani, Poorva Agrawal, Sanjeev Kumar Pippal) Applied System Innovation, vol. 6, no. 1, pp. 27, 2023. DOI: 10.3390/asi6010027

[4] FieldPlant: A Dataset of Field Plant Images for Plant Disease Detection and Classification With Deep Learning (Emmanuel Moupojou, Appolinaire Tagne, Florent Retraint, Anicet Tadonkemwa, Dongmo Wilfried, Hyppolite Tapamo, Marcellin Nkenlifack) IEEE Access (2023), this study introduces FieldPlant, a dataset for realistic plant disease detection in field conditions. DOI: 10.1109/ACCESS.2023.3263042

[5] Apple Leaf Disease Detection Using Collaborative ML/DL and Artificial Intelligence Methods: Scientometric Analysis (Anupam Bonkra, Pramod Kumar Bhatt, Joanna Rosak-Szyrocka) International Journal of Environmental Research and Public Health (2023), this paper reviews AI-driven methods for detecting apple leaf diseases. DOI: 10.3390/ijerph20043222

[6] Pepper bell leaf disease detection and classification using optimized convolutional neural network (Hassan Mustafa, Muhammad Umer, Umair Hafeez, Ahmad Hameed, Ahmed Sohaib, Saleem Ullah & Hamza Ahmad Madni) Multimed Tools Appl 82, 12065–12080 (2023). https://doi.org/10.1007/s11042-022-13737-8

[7] A Systematic Review on the Detection and Classification of Plant Diseases Using Machine Learning (Deepkiran Munjal, Laxman Singh, Mrinal Pandey, Sachin Lakra) IJSI vol.11, no.1 2023: pp.1-25. https://doi.org/10.4018/IJSI.315657

[8] PLDD—A Deep Learning-Based Plant Leaf Disease Detection (R. K. Lakshmi and N. Savarimuthu) IEEE Consumer Electronics Magazine, vol. 11, no. 3, pp. 44-49, 1 May 2022, doi: 10.1109/MCE.2021.3083976

[9] Vision-Based Wilted Plant Detection (J. Madake, S. Shinde, S. Singh, S. Talwekar, S. Bhatlawande and S. Shilaskar) IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, 2022, pp. 1-6, doi: 10.1109/IATMSI56455.2022.10119325.

[10] Detection and classification of chilli leaf disease using a squeeze-and-excitation-based CNN model (B.Nageswararao Naik, R.Malmathanraj, P. Palanisamy) Ecological Informatics 69(6):101663 DOI:10.1016/j.ecoinf.2022.101663

[11] Plant Disease Detection Using Deep Convolutional Neural Network (J. Arun Pandian, V. Dhilip Kumar, Oana Geman,

Mihaela Hnatiuc, Muhammad Arif, and K. Kanchana Devi) Applied Sciences, vol. 12, no. 14, pp. 6982, 2022. DOI: 10.3390/app12146982

[12] Plant Disease Detection using AI based VGG-16 Model (A. A. Alatawi, S. M. Alomani, N. I. Alhawiti, and M. Ayaz) International Journal of Advanced Computer Science and Applications (IJACSA), vol. 13, no. 4, pp. 477–482, 2022. DOI: 10.14569/IJACSA.2022.0130484.

[13] Investigation on Object Detection Models for Plant Disease Detection Framework (K. L. R and N. Savarimuthu) IEEE 6th International Conference on Computing, Communication and Automation (ICCCA), Arad, Romania, 2021, pp. 214-218, doi: 10.1109/ICCCA52192.2021.9666441.

[14] Disease Detection in Apple Leaves Using Deep Convolutional Neural Network (Prakhar Bansal, Rahul Kumar and Somesh Kumar) Bansal, P.; Kumar, R.; Kumar, S. Disease Detection in Apple Leaves Using Deep Convolutional Neural Network. Agriculture 2021, 11, 617. https://doi.org/10.3390/agriculture11070617

[15] Paddy Leaf Disease Detection Using an Optimized Deep Neural Network (Sankaranarayanan Nalini) Published in Computers, Materials & Continua (CMC), Volume 68, Issue 1, February 2021, with DOI 10.32604/cmc.2021.012431

[16] Identification of disease using deep learning and evaluation of bacteriosis in peach leaf (Saumya Yadav, Neha Sengar, Akriti Singh, Anushikha Singh, Malay Kishore Dutta) Ecological Informatics Volume 61, March 2021, 101247,doi.org/10.1016/j.ecoinf.2021.101247

[17] Identification of Pathological Disease in Plants using Deep Neural Networks (R. Biswas, A. Basu, A. Nandy, A. Deb, R. Chowdhury and D. Chanda) Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN), Rajpura, India, 2020, pp. 45-48, doi: 10.1109/Indo-TaiwanICAN48429.2020.9181339.

[18] SVM-Based Detection of Tomato Leaves Diseases (Usama Mokhtar, Nashwa El Bendary, Aboul Ella Hassenian, E. Emary, Mahmoud A. Mahmoud, Hesham Hefny & Mohamed F. Tolba) In: Filev, D., et al. Intelligent Systems 2019. Advances in Intelligent Systems and Computing, vol 323. Springer, Cham. https://doi.org/10.1007/978-3-319-11310-4_55

[19] A Computer Vision System for Guava Disease Detection and Recommended Curative Solution Using Deep Learning Approach (A. S. M. Farhan Al Haque, R. Hafiz, M. A. Hakim and G. M. Rafiqul Islam) 22nd International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ICCIT48885.2019.9038598.

[20] Plant Disease Detection and Classification by Deep Learning (Muhammad Hammad Saleem, Johan Potgieter and Khalid Mahmood Arif) Plants (Basel). 2019 Oct 31;8(11):468. doi: 10.3390/plants8110468. PMID: 31683734; PMCID: PMC6918394

[21] Suvaditya Mukherjee "ResNet 50 Model Architecture" Source:Towards Data Science August 18 , 2022 Available: https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758