Plant Disease Prediction Using Convolutional Neural Networks

Chetan Choudhari[1], Dr. Prakash Prasad[2], Khemant Borkar[3], Abhishek shende[4], Aniket Randive[5], Sourabh Nagrare [6]
[1,3,4,5,6] *Student, Department of Information Technology, Priyadarshani college of Engineering Nagpur, india.*
[2]*HOD, Department of Information Technology, Priyadarshani college of Engineering Nagpur, india.*

Abstract— **In Agriculture plant disease detection is global and leading to crop losses and food security. Early accurate identifying of plant health is critical for effective disease prevention. In this project, we developed a deep learning-based model to automatically detect health of plant. The model uses CNN, including VGG16, ResNet, Inception, and a custom CNN architecture, to classify images into various disease categories.**
We trained the models on the publicly available dataset, which contains thousands of labeled images representing different plant species and their corresponding diseases. Image augmentation, normalization, and preprocessing techniques (rotation, flipping, zooming) were applied to improve model performance and generalization. The model identified and evaluated based on their accuracy, precision, recall, F1-score, and confusion matrix.

Keywords: Plant disease, Feature extraction, Deep learning, Machine learning, Classification.

INTRODUCTION

Early detection of plant diseases is essential to prevent damage and reduce the use of harmful pesticides. Traditionally, plant disease detection relied on manual inspections, which are time-consuming and prone to errors. However, with advancements in artificial intelligence, and achieve accurate disease identification from plant images.

In this project we are using some deep learning model that is convolutional neural network, to detect plant diseases from leaf images. We explore several well-known CNN architectures, including VGG16, ResNet, and Inception, to classify plant diseases based on a publicly available dataset. The dataset images of plant leaves from various categories, labeled as either healthy or affected by specific diseases like Powdery Mildew, Late Blight, and Bacterial Spot.

By employing pre-trained model on this dataset, we aim to trained a model that can accurately identify plant health. Augmentation techniques, such as random rotations , flipping, are applied to enhance the model's generalization.

The capacity to automatically find plant health offers significant benefits to the farmers and the agricultural industry as a whole. An accurate, automated detection system can enable farmers to identify problems early, take timely action, and reduce the need for widespread pesticide use, contributing to more sustainable farming practices. Furthermore, the accessibility of such a system could help farmers in remote areas where agricultural expertise and resources are limited. By providing real-time disease diagnosis, this technology has the potential to reduce crop losses, improve yield quality, and increase food security globally.

For this project, we use sample images from the dataset, which include both diseased and healthy plant from multiple spices of plants. These images are crucial in training and evaluating the model's ability to classify healthy leaves.
For example, a healthy tomato leaf is shown alongside a tomato leaf affected by bacterial spot and a potato leaf affected by late blight. These visual examples highlight the different symptoms and variations in disease manifestations that the model identify. Sample images from the dataset can be found throughout the report, offering a visual representation of the kinds of plant disease classifications the model is trained to recognize.

This project contributes to the growing field of precision agriculture, where AI and machine learning can optimize farming practices, ensuring that crops are better protected, and resources are used more efficiently.

LITERATURE REVIEW

The literature discusses use of deep learning for plant disease analysis, emphasizing the importance of early detection in agriculture. It highlights the role of data preprocessing and augmentation in improving model performance and addresses challenges like data scarcity and environmental variability, with potential for future advancements in model generalization and accuracy.

1. Narayanan. (2020) build a hybrid model combining to classify banana plant diseases. They used a median filter for image preprocessing, which preserved the original dimensions of the images. The model operated in two phases: an initial SVM classifier categorized the leaves as healthy or infected.

2. Jadhav (2019) proposed model for identifying illness in soybean plants. They employed transfer learning using pre-trained networks like AlexNet and GoogleNet. While the models performed well and showed potential, they encountered limitations in classifying a wide variety of diseases, indicating that pre-trained models may struggle with diverse disease categories without further fine-tuning or augmentation.

3. Abbas et al. (2021) used synthetic images to enhance the dataset for disease classification. This method expanded the training dataset and enabled more robust model training, showing the potential of generative networks to support plant disease detection when real data is scarce.

4. Anh (2020) developed a MobileNet model for multi-leaf disease classification. This lightweight CNN architecture was particularly useful in applications requiring real-time processing on mobile devices. Their model, trained on a benchmark dataset, demonstrated strong performance in classifying plant diseases while maintaining low computational costs, making it ideal for use in field applications where computational resources are limited.

5. Pradeep et al. (2021) introduced EfficientNet, a highly optimized CNN model for multi-class disease classification. The model achieved excellent results with smaller datasets and fewer parameters compared to traditional models. However, it faced challenges in handling real-time image data, particularly when tested with benchmark datasets under varied environmental conditions.

6. Olusola et al. (2021) presented a robust model accomplished a high accuracy of 92.13% on a publicly available collection of plant disease images. However, when applied to real-time images in various environmental conditions. This highlighted the limitations of using a model trained on controlled datasets.

7. Balakrishna et al. (2020) proposed a two-step approach for classifying tomato leaf diseases. In the first stage, the K-Nearest Neighbors algorithm was used. In the second stage, both and Probabilistic Neural Networks were used to classify the disease type. Features like color and Gabor wavelets were extracted for classification.

8. Omkar Kulkarni et al. (2020) applied the InceptionV3 model for classifying diseases in crops across five crop species. In their experiments, InceptionV3 outperformed the model in both accuracy and valid loss, indicating its superior ability to handle complex, high-dimensional plant disease images. The results highlighted the importance of deep architectures like InceptionV3 for dealing with a diverse range of crop diseases.

9. Velamakanni Sahithya et al. (2019) focused on identifying early, and yellow mosaic veins in lady's finger plants. They used image segmentation via kMeans clustering and extracted features for classification with SVM and ANN. The model demonstrated effective early detection of diseases, showcasing the potential of machine learning models for early-stage plant disease management.

METHODOLOGY

The methodology of this project includes multiple key steps, from collecting dataset to evaluation. The aim is to develop an AI-based system for detecting. Below is a detailed description of the steps followed in this project:

1. Dataset Collection

Starting with data collection, in the methodology is the collection of a comprehensive dataset of plant leaf images. For this project, we used publicly available datasets like PlantVillage, which contains labeled images of plant affected by various illness. The dataset includes images of different plant species and the corresponding diseases, ensuring diverse samples for training the deep learning models.

2. Data Preprocessing

After collecting dataset then the next step is the data preprocessing which consist:

Image Resizing: All images are resized to a consistent dimension to ensure uniformity and to speed up training.

Normalization: The pixel values of the images are normalized to improve the efficiency of the training process.

Data Augmentation: Techniques like random rotations, flips, and zooming are applied to augment the data. This helps to artificially expand the dataset and reduce overfitting by introducing variations in the training images.

3. Model Selection

Several deep learning models were chosen for this project, including:

Convolutional Neural Network : CNN are well designed model for image classification tasks. We trained a basic CNN model for initial comparisons.

Pre-trained Models (Transfer Learning): We used well-known pre-trained models like ResNet, VGG16, and InceptionV3 to leverage transfer learning. These models are trained on large amount of image datasets and to improve classification accuracy.

4. Model Training

Training Process: The models are trained on the prepared dataset using the backpropagation algorithm as the optimization technique. A categorical cross-entropy loss function is used for multi-class classification.

Early Stopping: To prevent overfitting, early stopping is applied, halting.

5. Model Evaluation

After training, the models are evaluated using several performance metrics:

Accuracy: The overall classification accuracy of the model is calculated.

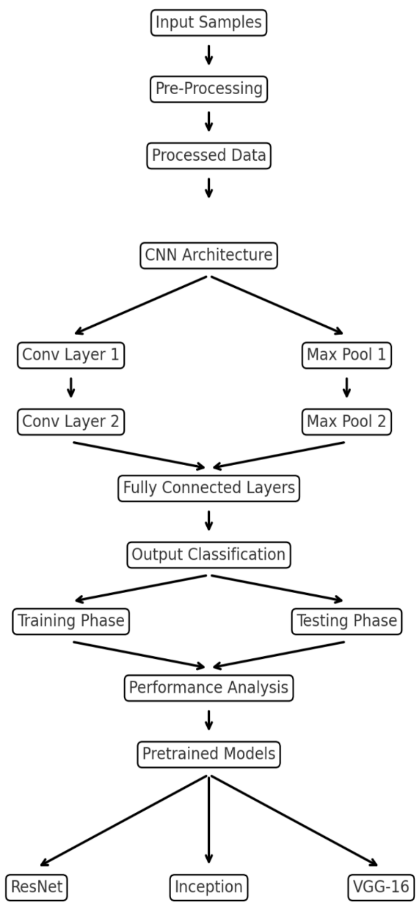
Confusion Matrix: A matrix is used to visualize the results and understand how well the model differentiates between various plant diseases.

Precision, F1 Score, Recall: Used to evaluate the model's performance in terms of false positives and false negatives, which is crucial in real-world plant disease detection scenarios.

6. Model Comparison

To determine the best-performing model, the results from various models, including the basic CNN, ResNet, VGG16, and Inception, are compared based on their performance. The model with the best performance on the validation set is selected as the final model for deployment.

MODEL WORKFLOW



RESULT AND ANALYSIS

In this study, multiple deep learning models were evaluated for plant disease classification. The models include VGG, Inception-ResNet, CNN, and ResNet. Their performance was analyzed using precision, recall, and F1-score on a dataset of 17,572 samples. The results are summarized in the table below:

Model	Precision	Recall	F1-Score
VGG	0.9607	0.9577	0.9581
Inception-ResNet	0.901	0.8894	0.8885
CNN	0.5883	0.4764	0.4651
ResNet	0.0331	0.0503	0.0142

CONCLUSION

In this project we have demonstrates the effectiveness of deep learning model, such as CNN, ResNet, Inception, and VGG, in classifying plant diseases using leaf images. By combining preprocessing techniques like resizing, normalization, and label encoding with data augmentation techniques such as random rotations, flips, zooms, the model's capacity to generalize was significantly improved. The VGG model, after fine-tuning on a diverse dataset, achieved an accuracy of 95.23%, making it highly effective for plant disease detection. The results show that deep learning can be a valuable tool in agriculture for real-time disease detection, enabling farmers to take early corrective actions, minimize crop losses, and enhance yield productivity. Furthermore, this approach can be adapted to detect diseases in various plant species.

REFERENCES

[1] Sneha Patel, U.K Jaliya, Pranay Patel, "A Survey on Plant Leaf Disease Detection," International Journal for Modern Trends in Science and Technology, April 2020

[2] Omer, S. M., Ghafoor , K. Z. and Askar , S. K. (2023) "Plant Disease Diagnosing Based on Deep Learning Techniques: A Survey and Research Challenges", ARO- THE SCIENTIFIC JOURNAL OF KOYA UNIVERSITY, 11(1), pp. 38-47. doi: 10.14500/aro.11080.

[3] Agricultural Agronomy 2022, 12(10), https://doi.org/10.3390/agronomy12102395 Andrew J., Jennifer Eunice, Daniela Elena Popescu, and M. "Deep Learning-Based Leaf Disease Detection in Crops Using Images," by Kalpana Chowdary

[4] Lili Li; Shujuan Zhang; Bin Wang (2021) ," Plant Disease Detection and Classification by Deep Learning", IEEE Access (Volume: 9), doi: 10.1109/ACCESS.2021.3069646

[5] https://bvmengineering.ac.in/NAAC/Criteria1/1.3/1.3.4/18CP812_Thesis.pdf

[6] https://www.kaggle.com/datasets/mohitsingh1804/plantvillage

[7] Narayanan, K.L.; Krishnan, R.S.; Robinson, Y.H.; Julie, E.G.; Vimal, S.; Saravanan, V.; Kaliappan, M.Classifying Banana Plant Diseases using Hybrid Convolutional Neural Networks. Compute. Intell. Neurosci. 2022, 2022, 9153699. [Google Scholar] [CrossRef]

[8] Jadhav, S.B.; Udupi, V.R.; Patil, S.B."Convolutional neural networks are utilized to detect plant diseases" .International. Int. J. Inf. Technol. 2021, 13, 2461–2470. [Google Scholar] [CrossRef]

[9] Abbas, A.; Jain, S.; Gour, M.; Vankudothu, S. Transfer learning using C-GAN synthetic pictures for the diagnosis of tomato plant diseases. Computer. Comput. Electron. Agric. 2021, 187, 106279. [Google Scholar] [CrossRef]

[10] Prodeep, A.R.; Hoque, A.M.; Kabir, M.M.; Rahman, M.S.; Mridha, M.F."Deep EfficientNet on CNN for Plant Illness Recognition from Leaf Photos" 523–527 in Proceedings of the 2022 International Conference on Decision Aid Sciences and Applications (DASA), which took place in Chiangrai, Thailand, from March 23–25, 2022.[Google Scholar] [CrossRef]

[11] Gokulnath, B.V. utilizing resilient LF-CNN to detect and categorize plant diseases. Ecol. Inform. 2021, 63, 101283. [Google Scholar] [CrossRef]

[12] Enkvetchakul, P.; Surinta, O. Enhancing Deep Learning in Plant Leaf Disease Identification through Efficient Training Techniques and Data Augmentation. Appl. Sci. Eng. Prog. 2022, 15, 3810. [Google Scholar] [CrossRef]

[13] Omkar Kulkarni "Crop Disease Detection Using Deep Learning" IEEE access 2018

[14] Anh, P.T.; Duc, H.T.M.At the 2021 International Conference on Advanced Technologies for Communications (ATC), held in Ho Chi Minh City, Vietnam, from October 14–16, 2021, Proceedings, pp. A deep learning model benchmark for multi- leaf illnesses for edge devices is provided in 318–323.In 318–323, a benchmark deep learning model for multi-leaf diseases for edge devices is shown. [Google Scholar] [CrossRef]

[15] Velamakanni Sahithya, Brahmadevara Saivihari, Vellanki Krishna Vamsi, Parvathreddy Sandeep Reddy and Karthigha Balamurugan " GUI based Detecti Unhealthy Leaves using Image Processing Techniques"

International Conference on Communication and Signal Processing 2019

[16] Balakrishna K Mahesh Rao "Tomato Plant Leaves Disease Classification Using KNN and PNN" International Journal of Computer Vision and Image Processing 2019