Results

Plagiarism   1.22%

Search settings

Only latin characters	❌
Exclude references	❌
Exclude in-text citations	❌
Search on the web	✅
Search in my storage	❌
Search for AI text	❌

Sources (10)

#	Source	
1	nature.com https://www.nature.com/articles/s41598-025-19303-0	0.46%
2	pmc.ncbi.nlm.nih.gov https://pmc.ncbi.nlm.nih.gov/articles/PMC10662291/	0.46%
3	nature.com https://www.nature.com/articles/s41598-025-14169-8	0.42%
4	learnopencv.com https://learnopencv.com/mean-average-precision-map-object-detection-model-evaluation-metric/	0.33%
5	peerj.com https://peerj.com/articles/cs-3154.pdf	0.28%
6	medrxiv.org https://www.medrxiv.org/content/10.1101/2025.05.26.25328342v1.full-text	0.28%
7	nature.com https://www.nature.com/articles/s41598-025-03402-z	0.28%
8	blog.roboflow.com https://blog.roboflow.com/mean-average-precision/	0.2%
9	francis-press.com https://francis-press.com/papers/10907	0.15%
10	collegedunia.com https://collegedunia.com/college/24339-panimalar-engineering-college-chennai	0.14%

🔴 Plagiarism 🟡 Similarities 🟣 AI probability 🟢 Citations 🔵 References 🔵 Character replacement

Lightweight CNN for plant disease diagnosis: Providing an instantaneous solution to various plant diseases through the enhanced plantVillage dataset

Dr Sangeetha K, Computer Science Engineering, Panimalar Engineering College.

Mahadevi.M

Computer Science Engineering

Panimalar Engineering College (Anna University)

Chennai, India

mahadevimaharajan@gmail.com

Kavishri. N

Computer Science Engineering

Panimalar Engineering College

(Anna University)

Chennai, India.

kavishrinatarajan@gmail.com

Meenakshi.M

Computer Science Engineering

Panimalar Engineering College

(Anna University)

Chennai, India.

meenakshideiveegan294h@gmail

ABSTRACT: The adoption of lightweight CNN (Convolutional Neural Network) for detecting plant disease provides an accurate solution for the various diseases instantaneously using machine learning and deep learning. Due to plant diseases, a large number of crops are lost annually, which is one of the major reasons for poverty. Traditional methods, such as inspecting the conditions of the plant by the agriculture officers,

Plagiarism   Similarities   AI probability   Citations   References   Character replacement

were time-consuming, expensive, and often inaccurate. A robust model was proposed that utilises Machine learning (ML) and Deep learning (DL) architecture instantaneously to analyse images of plant leaves and detect diseases. The model overcomes the generalisation error and predicts the disease more accurately. The accuracy was 98.7, and the F1-score was 98.5. The Mean Squared Error was very low, approximately equivalent to 0.8.

Keywords: Machine learning, Deep learning, Convolutional Neural Network (CNN), F1-score, robust, Generalization.

1. INTRODUCTION

The population of India has been increasing enormously, and the need to protect the crops is becoming a more crucial task than before. It is estimated that plant diseases pose a large yield loss annually, affecting up to 40% of the economy. If the world population increases in the same way, the need for food will also increase. Therefore, the protection of crops from various diseases becomes a basic need that ensures a continuous supply of food to all people and eradicates hunger and poverty.

Long years back, the plant diseases were detected through manual visual inspectors manually. Visual Inspection is effective only on a small scale, but it is time-consuming and expensive when used on a large scale. So, this app is designed with several features such as voice assistance and multilingual support, which enable even uneducated people to identify what disease has affected their crops and the corresponding treatments to be done. Along with deep learning, the use of Convolutional Neural Network (CNN) classifies the images with high accuracy. It predicts the disease and provides appropriate treatment for the disease. This enables the application to perform much faster than human experts.

This paper presents a plant detection application utilizing deep learning, which enables the automatic detection and classification of diseases, providing an end-to-end solution for various diseases. The usage of TensorFlow Lite compressed the model's size and made it efficient while deploying. The use of Google Colab, from downloading the dataset to preprocessing it (removing null values) and training the model, made the entire workflow transparent. The primary objective of this application is to empower farmers by providing solutions to various diseases and promoting the Sustainable Development Goals.

2. LITERATURE REVIEW

Traditional Methods: These traditional methods mainly depend on human visual inspection by agricultural

experts. They try to identify and predict diseases based on symptoms such as discoloration(change of color or degradation of color) and spots. It often shows errors and is inefficient for large-scale farms. More advanced traditional methods are accurate but time-consuming and cost-inefficient, which is not suitable for most farmers

Modern approaches: The modern approach mainly focuses on computer vision, Machine learning, and deep learning. These methods automatically analyze images of plant leaves or other parts to identify and classify diseases.

Dataset: PlantVillage dataset is the most commonly used dataset (consists of nearly 54000 images across 38 classes in many distributions. The dataset most commonly contains controlled background, zoomed leaf photos. Field imagery from UAVs, multispectral datasets, and other locally collected small datasets.

Evaluation Metrics: The performance can be analyzed based on accuracy and recall..

Imaging: RGB imaging is the cheapest, easiest, and most widely used for visual symptom detection. Multispectral/hyperspectral sensors capture wavelengths beyond visibility and enable early detection of diseases before visible symptoms appear, and can help to differentiate various types of pathogens. These sensors are very helpful in identifying diseases. Early work used color/texture features (e.g., color histograms) + classifiers like SVM, Random Forest, k-NN. These methods can be interpretable but require careful feature design and do poorly under large visual variability (lighting, viewpoint, background.Mohanty et al reported very high accuracy on the controlled images and established the feasibility of deep learning for leaf-based disease classification.

Image Acquisition and processing: Images are captured using mobiles. This allows for rapid data collection over large areas. The model even supports voice assistance, which helps even the uneducated farmers to identify what disease has affected their crops and the corresponding treatment to be taken. It also supports multilingualism, which enables the farmers to convey their problems in their mother tongue. Raw images of leaves are pre-processed to enhance the quality of the image. This may include noise reduction, resizing, and cropping of unwanted background images. The most important step is image segmentation, which isolates the infected leaf or infected part of the leaf for the prediction of diseases.

Feature Extraction: After segmentation, features of the diseased area, such as color and texture, are extracted. For example, a disease might be characterized by a specific shade of brown and a rough surface (texture).

Classification: The extracted features are fed into Machine learning and Deep learning for the classification

of diseased plants

3. OBJECTIVES

The main objective of this project is

•   To empower farmers by predicting various diseases with the implementation of offline prediction and classification of diseases.

•   To provide multilingual support and voice assistance.

•   To provide specific information about the disease and the accurate treatment.

4. PROPOSED METHODOLOGY

The first step is to collect the dataset of plant images. The plant images should include both healthy and diseased leaves from different plant species. The dataset also consists of images from various conditions, such as different lighting and different angles, so that the learning algorithm learns all possibilities and makes the model much effective. The dataset is preprocessed by cropping the unwanted background images and focusing only on the diseased part of the leaves. Increasing the brightness and scaling of existing images to provide high clarity.CNN can automatically learn from the images of the plant and predict the diseases without any manual intervention. After the validation set. The parameters are fine-tuned, and the model is generalized to perform well on the testing set. The trained model is deployed as a user-friendly application, such as a mobile app. This allows the farmers to take a picture of a plant leaf and receive instant health status of the plant. The app provides information such as the name of the disease, the severity of the disease, and the corresponding treatment that should be undertaken. The model also supports voice assistance, multilingual and offline support, to make it much easier for the farmers.

 Figure1-Methodology

5. PROPOSED ARCHITECTURE

The system's proposed architecture is a much efficient framework designed for plant disease detection. It processes images of the plant through a series of layers to extract and learn features. The architecture includes an Input Layer, which receives a preprocessed image. Convolutional Layers, where the network learns to identify and analyze the features by applying various kinds of filters. The CNN model is converted into a lightweight TensorFlow Lite model to make it suitable for mobile applications.It also includes a voice

assistant, which handles voice input from the user, translating it into actions. The app also includes Camera Integration to capture images of diseased plant leaves for analysis and to display the predicted disease and treatment information.

5.1 ARCHITECTURAL DIAGRAM

Figure2-Architectural diagram

6. RELATED WORK

PLANT DISEASE DETECTION DATASET DESCRIPTION

Dataset Parameters	The value of the parameters
Dataset Name	PlantVillage Dataset
Source	Kaggle.com (Originally from PlantVillage)
Total Number of Images	~54,300
Number of Classes	38 (includes both healthy and diseased plant leaves)
Plant Species Covered	14 species, including apple, cherry, corn, grape, peach, potato, strawberry, and tomato
Image Type	RGB (color) images of plant leaves
Image Resolution	256x256 pixels
Data Collection Environment	Controlled conditions with a plain background
Data Split	80% training set, 10% validation set, and 10% testing set
Data Augmentation	Common augmentation techniques like rotation, zoom, and horizontal/vertical flips are applied during training.

TABLE 1-Description of the plantVillage dataset

Figure 3-Sample image from PlantVillage Dataset[9]

7. WORKFLOW

7.1 DATA PREPROCESSING

 In machine learning for plant disease detection, data preprocessing is the most crucial step that converts raw, unorganized image data into a clean, standardized format suitable for a deep learning model. Collection of images is the initial step.The images affected by diseases can be captured from mobiles or the dataset can be taken from websites like Kaggle, but they must include images of healthy and diseased leaves. Once the image is collected, the image must be reduced to a particular dimension to improve the model's prediction. The image may include irrelevant backgrounds such as soil and other plants. Image segmentation is a process that is used to separate the leaf from other background images by cropping, allowing the model to focus only on the relevant features of the.This helps the model to be more generalized to the unseen data. Geometric Transformations, such as rotating, zooming, and flipping of images, and Color Augmentation, such as adjusting the brightness and contrast, are applied to make the model efficient. The pixel values are normalized. This involves fixing the pixel values of the images in a standardized range such as[0,1] or [-1,1].

7.2 FEATURE SELECTION

In feature selection, the main difference is that instead of using the manual step, the CNN automatically learns during the training process.As the data passes through deeper layers, these simple features are integrated to get more complex views, such as the texture of a fungal growth. The various algorithms used help to highlight the most predictive of the disease class. In this way, the feature selection using CNNs is highly effective to extract complex features that would be much difficult for humans to define and extract.

7.3 ALGORITHM FOR CONVOLUTIONAL NEURAL NETWORK

Input and convolution: The model starts with the preprocessed image of a plant leaf. The model uses a set of filters and analyzes visual features such as edges, textures, and color patterns.

Activation Function: The model uses a nonlinear function on features.This helps the model to observe and analyze the important patterns using ReLU after identifying the various features using convolution

Pooling: The pooling layer summarizes. It reduces the size by keeping the most significant information.

Flattening: The flattening layer summarizes a single long line. The 2D grid of features is converted into a 1D vector, which is a required set of layers to process all the information at once.

Fully Connected Layers: The vector is passed through the connected layers. Fully connected layers integrate

the features to get the final classification.

Output and Prediction: The output is provided as a probability distribution for each class. A softmax function is used to normalize(standardized form)these probabilities into a single value, and the class with the highest probability is considered the final prediction.

8. DEEP LEARNING APPROACH

It leverages the power of CNN to automatically learn from complex features and help to predict the disease accurately.

1. Data collection and preprocessing: The collection of data includes images from various repositories, such as PlantVillage.Images were chosen from various conditions to improve the model accuracy. Image Resizing, Normalization, and Image Augmentation are done to improve the model's prediction.

2. Model development and training: The model is trained on a large dataset of healthy and diseased plant leaf images. TensorFlow Lite has been included to convert the CNN model to a lightweight model. This app allows for making predictions without an internet connection. The voice assistant is included, which uses voice commands to provide the name of the disease and the treatment required

3. Prediction: The model predicts the disease, its severity, and the treatment required within an instant of time.

4. Evaluation of Model: The model's performance on the testing data is used for evaluation. Accuracy, Prediction, and Recall are used to evaluate a model's efficiency

9. EXPERIMENTAL RESULTS

The model's performance is evaluated based on a large dataset, which includes a large collection of healthy and diseased plant leaf images.

The plant disease dataset used in the learning algorithm is a comprehensive collection of images from various repositories such as PlantVillage. The dataset includes details on plant species, disease types, and a variety of leaf images(healthy and diseased) captured under different environmental conditions. For uniformity, all images are preprocessed by standardization. The dataset consists of a large number of samples split into the training and testing sets. The training set is used to train the CNN for the extraction of features and the prediction of disease, while the testing set is utilized to evaluate the model's performance and generalization.

10.EVALUATION METRICS

There are several methods to evaluate the performance of the proposed approach,

• Accuracy

• Precision

• Recall

• F1-Score

• Mean Average Precision (mAP)

11. IMPLEMENTATION DETAILS

The implementation of plant disease detection includes several stages, from model training to the development of the mobile application. The learning rate is fixed at a small value for stable convergence of model weights. After training, the CNN model is converted into TensorFlow Lite. The converted model, along with class labels, is integrated into the mobile app. The app's functionality includes :

Camera API integration: This is used to capture images with high resolution.

User Interface: It is designed to display the predicted disease and the treatment recommendations

Voice Recognition: Integration of a voice assistant API to interact with the user.

Figure 4 User Interaction Flow

12. BASELINE METHODS

To illustrate the effectiveness and efficiency of the model using a deep learning approach, the model's performance is compared against traditional image processing and machine learning methods.

1) Support Vector Machine (SVM): This method involves the extraction of images manually, such as color(RGB), texture(rough or smooth), and shape(regular or irregular).This approach is less intensive than deep learning as it cannot predict with high accuracy, as in CNN.

2)K-Nearest Neighbors (K-NN): The K-nearest neighbor does not produce a model with the dataset.

3)Traditional Image Processing: This approach depends on a series of rules and various filters to identify diseased areas. It is effective for simple cases but varies highly when there is poor lighting. It lacks generalization when it predicts the unseen data.

Crop Management using CNN

By providing accurate diagnoses, this system helps farmers and agricultural professionals to take immediate

actions to prevent yield loss and optimize the available resources. A mobile application can analyze the images of the crop. The CNN running on the device can analyze the image based on early signs of disease. Once the disease is identified, farmers can use fungicides and the required treatments to a specific location instead of applying the fungicides to the entire crop field. All the above information is crucial for farm planning, adjustments in planting strategies, and ensures efficient crop yield

.

13. PERFORMANCE ANALYSIS

RETRIEVAL ACCURACY

The retrieval accuracy can be measured by using prediction, accuracy, recall, and F1-score. The following table provides performance metrics for a deep learning model (such as a CNN) on a large dataset like PlantVillage, when compared to other traditional models.

Model	Accuracy	Precision	Recall	F1-score
Custom CNN	98.7	98.6	98.6	98.5
VGG16	95.2	94.8	95.1	94.9
ResNet50	97.5	97.1	97.4	97.2

TABLE 2-Performance metrics for a deep learning model (such as a CNN) on a large dataset like PlantVillage, when compared to other traditional models.

.

Figure 5-Retrieval accuracy comparison of Custom CNN, VGG16, and ResNet50 models for plant disease detection.

Results

It showed that our proposed approach performs much better than the baseline methods in terms of precision, accuracy, recall, and F1-score.

RETRIEVAL EFFICIENCY

The retrieval efficiency is the measure of the average retrieval time per query image.

Model	Retrieval Time(s)	F1-score
Custom CNN	1.0	98.5
ResNet50	1.5	97.2
VGG16	2.0	94.7

TABLE 3-Representing the retrieval time and F1-score of various models

Figure 6-Retrieval efficiency comparison of Custom CNN, VGG16, and ResNet50 models for plant disease detection.

Results

The results demonstrate that our approach retrieves much faster when compared to baseline methods.

MEAN AVERAGE PRECISION

To evaluate the overall performance, we calculate Mean Average Precision(mAP).

Model	Average Precision-Early Blight	Average precision-Late Blight	Mean Average Precision(mAP)
Custom CNN	0.96	0.94	0.95
Resnet50	0.95	0.92	0.935
VGG16	0.88	0.85	0.865

TABLE 4-Represents the mAP of various models.

Figure 7: Bar chart representing the mean Average Precision of various models.

Results

- Custom CNN-0.95

- ResNet50-0.935

- VGG6-0.865

The Custom CNN has the Highest Mean Average Prediction.

MEAN SQUARE ERROR

Image	Actual % affected	Custom CNN Predicted %	ResNet50 Predicted%	VGG16 Predicted %
1	15%	15%	13%	10%
2	40%	41%	38%	45%
3	5%	4%	6%	7%
4	60%	59%	55%	62%
5	25%	24%	23%	20%

TABLE 5-Actual affected percentages vs predicted percentages of various models.

Figure 8-Comparing the actual percentage affected with the predictions from the Custom CNN, ResNet50, and VGG16 for each image

Figure 9-Model comparison based on MSE

14. CONCLUSION

Plant disease detection using AI and deep learning has proved to be an effective approach for improving productivity in agriculture and reducing crop losses.The ability of each of the models was tested to classify plant diseases accurately. Among them, the Custom CNN has the highest accuracy of 98.7%, high precision, recall, and F1-score. The results proved that deep learning models can significantly perform better than traditional methods. Early identification of diseases allows farmers to take immediate actions to prevent crop loss and to improve crop yield. The work of deploying such models in real-world applications, such as mobile apps, makes plant disease detection accessible even in rural areas.

○ Plagiarism ○ Similarities ○ AI probability ○ Citations ○ References ○ Character replacement