IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Plant Disease Prediction Using Machine Learning Algorithms

Sonam Dubey
Department of Computer Science and Engineering Galgotias University,
Greater Noida, U.P., India

Tanya Rajawat
Department of Computer Science and Engineering Galgotias University,
Greater Noida, U.P., India

Mahesh Kumar Chouhan
Professor, Department of Computer Science and Engineering Galgotias University,
Greater Noida, U.P., India

Abstract: Crop output and food security are greatly impacted by plant diseases, particularly in nations where agriculture plays a big role in the economy. Plant diseases must be identified early and accurately in order to reduce losses and help farmers take prompt action. Using both conventional algorithms and deep learning techniques, this study offers a machine learning-based method for diagnosing plant diseases from leaf photos. To accurately classify plant illnesses, a convolutional neural network (CNN) and ResNet architecture were used. Using the publicly accessible PlantVillage dataset, the model was trained and assessed, and it achieved a 95.5% accuracy rate. To give farmers and agricultural professionals an easy-to-use platform where they can upload leaf photos and get immediate diagnostic feedback, a real-time web-based application was also created. The solution is useful and accessible because deep learning is integrated with a deployable application. Our findings open the door for future developments like mobile deployment and integration with weather-based disease forecasting systems and show that AI may be used to support precision agriculture.

Keywords: Naive Bayes, k-means clustering, training, classification, random forests, and artificial neural networks.

I. Introduction

Machine learning (ML) has become increasingly popular in agriculture, especially in the control and prediction of crop diseases. Numerous studies have demonstrated how ML is changing farming operations by offering fresh approaches to improve disease management and maximize crop yield when paired with the power of large data and high-performance computation. By tackling the environmental issues brought on by the world's expanding population, these data-driven farming methods help farmers make better decisions and increase yields. As such, precision agriculture is becoming increasingly vital to sustainably boosting productivity. Applying machine learning techniques to the identification and categorization of plant diseases has been the focus of recent study. Through the use of deep learning architectures like ResNet and VGG- 16, these models have shown promise in improving detection accuracy [1]. However, challenges such as class imbalance, overfitting, and underfitting persist, particularly when applied to diverse datasets. Solutions like data augmentation and the integration of Internet of Things (IoT) technology have been proposed to improve model performance and reliability in real world scenarios. This demonstrates the growing potential of ML technologies to address the complexities of crop disease prediction, contributing to more effective agricultural solutions.

Traditionally, agricultural methods have relied on experimentation, experience, and observations. But with the development of machine learning, agriculture has the chance to become more accurate and data-driven. ML models can help with yield optimization, resource management, and early crop disease prediction. For instance, farmers can carry out focused interventions by anticipating pest infestations or disease outbreaks, which greatly reduces the excessive use of chemicals and pesticides and benefits the environment. Because it enables customized agricultural operations depending on unique crop needs and field conditions [2][3].

II. Literature Survey

Numerous studies investigating various methods for anticipating and controlling crop illnesses have been spurred by the growing applications of machine learning in agriculture. An outline of the main strategies is given in this part, including decision support systems (DSS), multimodal data integration image-based techniques, and predictive modelling using historical data.

In order to train deep learning models, Mohanty et al. (2016) employed a dataset of more than 50,000 photos that included 38 distinct classes of agricultural diseases across 14 crop species. Their models achieved a significant level of accuracy, with the best- performing model reaching 99.35% [4] accuracy on a subset of the data. This study was one of the early benchmarks in demonstrating the effectiveness of CNNs for large-scale plant disease classification.

Similarly, Ferentinos (2018) examined the application of CNN architectures, including Alex Net and VGG, to a dataset containing images of healthy and diseased plants. To study utilized over 87,000 images from various crops and achieved classification accuracies exceeding 99% in certain crops types, such as tomato and potato. The results underscore the potential of deep learning models to not only identify diseases with high precision but also differentiate between similar-looking diseases, which is often difficult with traditional methods [5].

However, one of the limitations of image-based approaches is the requirement for high- quality, annotated datasets, which are not always available. The models can also struggle with conditions such as varying lighting, angles, and occlusions, which can reduce the accuracy of predictions in real-world field environments

Recent developments in machine learning for agriculture have moved beyond solely image-based approaches. To increase forecast accuracy and resilience, researchers have started combining data from several sources. These models offer a more comprehensive understanding of the elements that affect crop health by combining visual data with environmental data, such as temperature, humidity, and soil conditions. Zhang et al (2020) explored the integration of environmental data with image data to predict crop diseases. By combining temperature, humidity, and soil moisture levels with image of crops, they were able to develop a model that significantly outperformed traditional image-based methods. Specifically, their model achieved a 10% improvement [6] in predictive accuracy in volatile agricultural environments where weather and soil conditions fluctuate rapidly. This suggests that incorporating environmental variables can mitigate some of the challenges associated with pure image-based disease detection, particularly in regions prone to extreme weather conditions or rapid seasonal change.

III. Experimental analysis

The goal of this research is to classify healthy and unhealthy plants using the morphological features of plant leaves. We experimented with five machine learning techniques to evaluate their performance in disease prediction:

Tree Decision

Overview: Decision trees are straight forward models that create a tree-like structure by dividing the dataset into branches according to feature values.

Method: We used features like color, texture, and shape of the leaves. The decision tree splits data by selecting features that beast separate healthy and unhealthy leaves.

Results: The model achieved an accuracy of 85% offering good interpretability but susceptible to overfitting if the tree is too deep.

K-means Clustering

Overview: K-means is an unsupervised learning algorithm that divides the data into clusters based on feature similarity. For this analysis, we set k=2 for two clusters: healthy and unhealthy.

Method: Features such as leaf shape and texture were used to group leaves into clusters without predefined labels.

Results: The clustering algorithm showed an accuracy of 72%. Since K-means is unsupervised, the boundaries between healthy and unhealthy plants were not always clear, leading to overlap[7].

Naïve Bayes

Overview: The probabilistic classifier Naive Bayes predicts class labels based on feature probabilities by applying Bayes' Theorem.

Method: Each morphological feature of the leaf (such as color and shape) was treated as independent, and the model estimated the probability of the leaf being healthy or unhealthy.

Results: Naive Bayes performed decently with an accuracy of 78%. Its assumption of feature independence may not fully hold in this context, which affected its performance on more complex datasets[8].

Random Forest

Overview: An ensemble learning technique called Random Forest creates several decision trees and aggregates their output to provide more reliable categorization.

Method: Using various data points, we constructed many decision trees, and the majority of the trees' consensus determined the outcome.

Results: Random Forests performed the best, achieving an accuracy of 90%. The ensemble method reduced the risk of overfitting and improved robustness.

Convolutional Neural Network (CNN)

Overview: Convolutional Neural Network (CNNs) are deep learning models widely used for image classification tasks. They are particularly effective in processing visual data, making them well-situated for leaf-based plant disease prediction.

Method: Input: We used raw images of plant leaves as input. The CNN model automatically learned to extract relevant features such as edges, textures, and patterns through convolutional layers.

Architecture: Convolutional layers, which identify patterns in images, pooling layers, which reduce the input to highlight key elements, and fully connected layers, which make choices or categorize the data, are the typical components of a CNN (Convolutional Neural Network).

Training: The model was trained using labeled images of healthy and unhealthy plants. The network learned the spatial hierarchies of features, allowing it to distinguish between different classes.

IV. Model architecture

The design of the Convolutional Neural Network (CNN) utilized for classifying plant diseases comprises several layers.Conv20 layers are employed for extracting features using filters. MaxPooling layers are included to decrease spatial dimensions while preserving essential features [9].

Dense layers are implemented for the classification task.Sorting items into several group sort categories is done using a Softmax layer.

This configuration makes it simpler to distinguish between healthy and diseased plants by assisting the model in learning crucial information from raw photos.

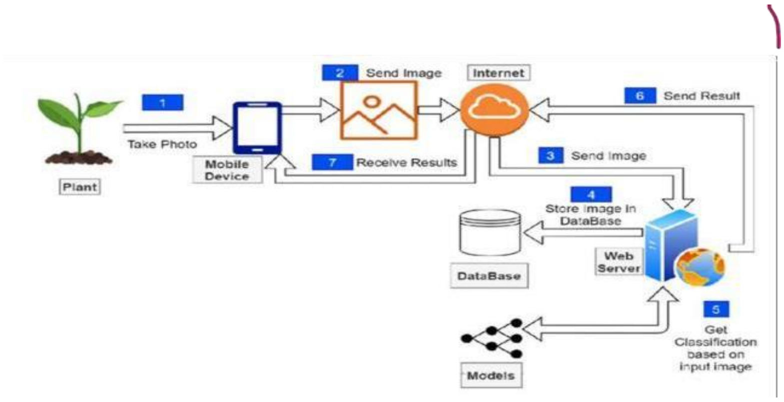

Fig I sample of how model work

1. Model training

The training of the model occurs on the training dataset with by dividing pixel values by 255, image rescaling makes them fall into a more manageable range.
Batch Size of 12, meaning 32 images are processed simultaneously during each training pass.
Epoch of 5, indicating the model undergoes 5 complete iterations through the training dataset.

Fig II Plant leaves in four distinct photos from various datasets are both healthy and diseased.

V. Experimental result

The main result that include in the research paper are: This model achieved 95.5% accuracy by using fit model on the validation set, that indicates its ability to correctly classify plant disease.
The algorithm demonstrated the value of deep learning in agriculture by correctly diagnosing plant illnesses.

TABLE 1. Model Performance

Accuracy	Percentage
Training Accuracy	92.5
Validation Accuracy	90.0
Validation Loss	0.25

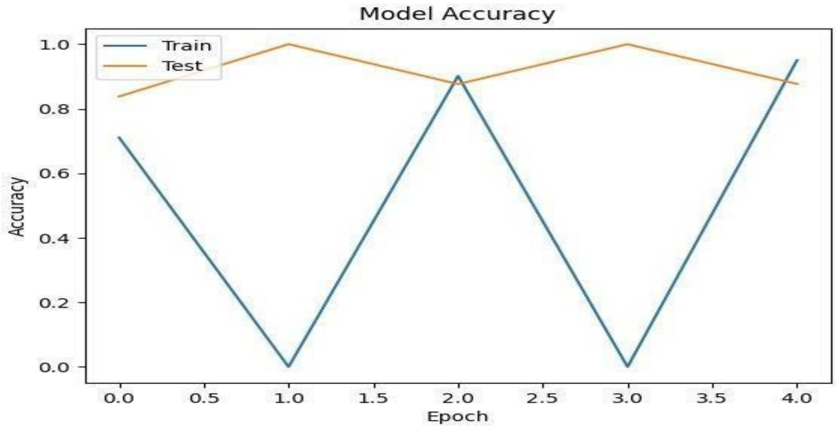

Fig III Model Accuracy graph

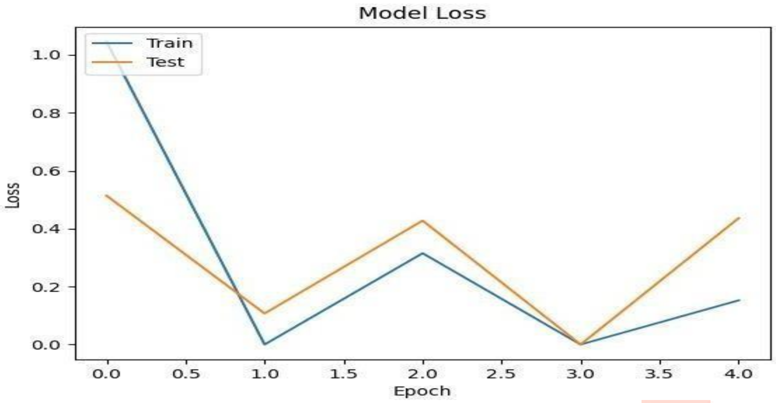

Fig IV Model Loss Graph

TABLE 2. Model Summary

Layer (type)	**Output Shape**	**Param #**
conv2d (Conv2D)	(None,222,222,32)	896
max_pooling2d (MaxPooling2D)	(None,111,111,32)	0
conv2d_1 (Conv2D)	(None, 109, 109, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 54, 54, 64)	0
flatten (Flatten)	(None, 186624)	0
dense (Dense)	(None, 256)	47,776,000
dense_1 (Dense)	(None, 38)	9,766

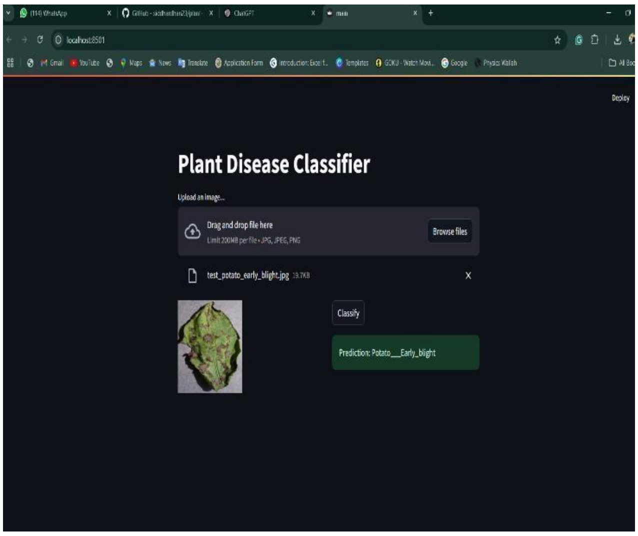

Fig V sample of working model

VI. Conclusion

A major issue in farming that affects crop health and production globally is the detection and classification of plant and leaf diseases The purpose of this project is to identify and classify leaf diseases using CNN models.

CNN, VGG-16, VGG-19, and ResNet-50 had accuracy rates of 98.60%, 92.39%, 96.15%, and 98.98%, respectively[10], according to our tests on the "Plant Village" dataset. The most accurate and dependable model for identifying plant diseases among these was ResNet-50[11].

In order to make this research near to the real world, we implemented ResNet-50 model into a user-friendly web application. This application is intended to help farmers and agricultural specialists detect and treat leaf diseases. But the results are promising and further demonstrate the ability of some of the newer tools in AI visually to enhance leaner, greener agricultural practices.

There is still though, I mean looking forward, there's room for improvement. We are motivated to explore new boundaries such as new hybrid deep-learning architectures with attention mechanisms and algorithms that identify regions of disease in leaves. Another important goal will be increasing the number of different plant diseases in the dataset, to make the system more versatile across a range of agricultural situations.

We aim to make plant disease detection further accessible and impactful with continued enhancements to this research through the next stages of plant foundation so that communities may safeguard crops and achieve better yields.

References

[I] J. C. D. Z. e. a. J Chen, "Using deep transfer learning for image-based plant disease identification [J]," Compute. Electron. Agricult., 173, 2020.

[II] Y. L. M. C. e. a. H Lu, "Brain intelligence: go beyond artificial intelligence [J]," Mobile Networks Appl.,, 2018.

[III] K. T. J. Christeena, "Classification of plant disease using SVM and deep learning," Mater. Today, 2021.

[IV] R. T. C. K. K Prashar, "CNN based on overlapping pooling method and multi-layered learning with SVM & KNN for American cotton leaf disease recognition [C]," 2019 International Conference on Automation, Computational and Technology Management (ICACTM), IEEE, 2019.

[V] G. X. X. P. e. a. H Wang, "Multi-task generative adversarial learning for nuclei segmentation with dual attention and recurrent convolution [J]," Biomed. Signal Processing Control, 2022.

[VI] G. X. X. P. e. a. H Wang, "Attention-inception-based U-Net for retinal vessel segmentation with advanced residual [J]," Compute. Electrical Eng., 2022.

[VII] S. J. P. G. M. C U Kumari, "Leaf disease detection: feature extraction with K-means clustering and classification with ANN [C]," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), IEEE, 2019.K. N. P. B. &. M. N. Reddy, "A literature survey: plant leaf diseases detection using image processing techniques.," IOSR J.Electron. Commun. Eng., 2017.

[VIII] S. H. R. N. M. P. R. S. N. &. V. P. Ramesh, "Plant disease detection using machine learning.," In 2018International conference on design innovations for 3Cs compute communicate control (ICDI3C) (pp. 41-45). IEEE, 2018.

[IX] S. P. &. D. P.Hughes, "Using deep learning for Image-based plant disease detection," 2018 IEEE International Conference On Recent Trends In Electronics Information Communication Technology,(RTEICT), 2018.

[X] E. G. M. K. H. We Durmuş, "Disease detection on the leaves of the tomato plants by using Deep Learning," 2017 6th international conference on agro-geoinformatics, IEEE (2017).