Undergraduate Conference on Intelligent Computing and Systems (UCICS 2025)
26-27 February, 2025; Varendra University, Rajshahi, Bangladesh

An Effective Deep Learning Approach to Plant Disease Detection

Md. Ishtiaq Ahmed
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
ishtiaqahmed21291@gmail.com

Md. Mizanur Rahman
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
mizanbd.eee@gmail.com

Md. Nur Hussain
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
nurhussainsany@gmail.com

Md. Musfiqur Rahman Mridha
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
mdmridha100730@gmail.com

Jamil Chaudhary
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
jamil@vu.edu.bd

Md. Arafat Ibna Mizan
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
arafat.cse.ruet18@gmail.com

***Abstract*—Plant diseases are considered a major problem that has an immense impact on agriculture and the economy. Early detection of plant diseases can result in less crop loss and significantly decrease financial losses. This study identifies plant diseases using deep learning (DL) and machine learning (ML) methods with plant leaf images. DL methods such as DenseNet121, DenseNet169, DenseNet201, and VGG19 are used. Aditionally, ML methods including Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR) have been implemented. The plant leaf images are collected from the PlantVillage dataset. DenseNet201 with SVM appears as the outperformer and provides the best results of 100% accuracy with precision, recall, F1-score, and AUC.**

***Index Terms*—Plant Disease, Deep Learning, Machine Learning, Feature extraction, Classifier, Recognition.**

I. Introduction

Plant diseases are a global problem. Many farmers do not know how to protect their crops from various diseases. As a result, the economy suffers as well as the crops.

According to the Food and Agriculture Organization (FAO), around 20-30% of crops are lost annually due to various diseases, with global losses estimated to be around billion dollars. In 2020, India alone suffered a 10% yield loss due to fungal infections, causing an economic loss of $1 billion[1].

It takes a long time to diagnose the disease manually. The large range and lack of efficient detection methods exacerbate the problem. So the need for automated solutions driven by deep learning is very well understood[2].

In 2018, India lost $5 billion in rice yield due to bacterial blight [3].

In 2020, Late blight caused $6.7 billion in losses globally for potato and tomato farmers [4].

Deep learning and machine learning models are applied for the detection of plant diseases. Deep learning is a branch of machine learning that deals with multi-layered neural networks and large datasets. Machine learning is a part of AI that focuses only on trained algorithms and can work with small to medium datasets.

DenseNet121, DenseNet201, DenseNet169 and VGG19 have been used for feature extraction. Used support vector machine (SVM), random forest and logistic regression classifier for classification.

This study introduces a deep learning(DL) and machine learning(ML) approach for plant disease detection using many feature extractor models and classifiers. This study tests different kinds of image-preprocessing techniques. The findings provide valuable insights for the development of efficient plant disease detection systems. The results will help the farmers to save theirs plants from rapid diseases and minimizing the crop losses.

II. Literature Review

Dey et al. used the AlexNet model with connected layers on the PlantVillage dataset. Their model achieved 96.63% accuracy, 92% precision, and 91% recall [5].

Similarly, Alnamoly et al. used an attention-based CNN model to detect tomato leaf diseases. Their model reached 99.04% accuracy, with 99% precision, recall, and F1-score, and an AUC of 99.90% [6].

Banerjee et al. combined CNN with a random forest classifier to detect tomato leaf diseases using the PlantVillage dataset. Their model had 98% accuracy, with 91.91% precision, recall, and F1-score [7].

Khalid et al. applied CNN and MobileNet to classify multiple leaf diseases using the Apple Leaf Disease Dataset. Their model achieved 96% accuracy, 90% precision, and 89% recall [8].

Prince et al. developed a CNN-SVM hybrid model and tested it on multiple datasets. Their model achieved 99.09% accuracy, with 99% precision, recall, and F1-score, and an AUC of 99.98% [9].

III. Methodology

In this section, the methodlogy of the study is descibed in detail. In Fig. 1 we provide a methodology diagram.

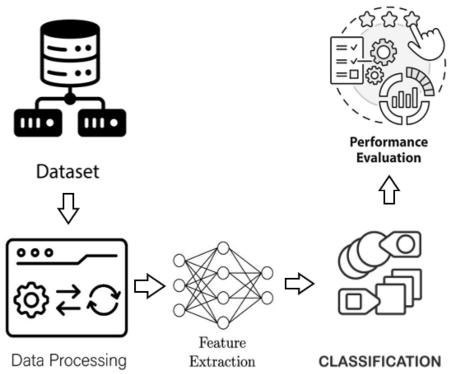

Fig. 1. Full overview of Methodology

A. Data Collection

The PlantVillage dataset was used for this study. It contains 38 classes and a total of 54,303 images[10]. From here, 8 classes have been selected, 4 healthy and 4 unhealthy. A class named 0 is made up of all healthy classes and a class named 1 is made up of all unhealthy classes. The dataset is then renamed to ProjectDataSet. The table below shows the details of the dataset for better understanding-

TABLE I
Summary of datasets used.

Dataset Name	Healthy Images	Unhealthy Images	Total Images
PlantVillage	13,000	41,303	54,303
ProjectDataSet	2,232	2,289	4,521

B. Data Preprocessing Techniques

Collected images from the dataset require many preprocessing techniques to improve the quality and ensure acceptable input for machine learning models by following techniques. CLAHE or Contrast-Limited Adaptive Histogram Equalization is used to improve the contrast of the images, it helps to make the details of the plants in the areas where there too dark or too bright. Laplacian Filter helps to highlight the edges of objects in the images. It helps the model to focus on important features. Normalization scaled the pixel values of the images to a range between 0 and 1.

This project tried to use more techniques such as Gaussian Blur for Noise Reduction, Gamma Correction, Advanced Data Augmentation. After using this, achieved less results compared to CLAHE, Laplacian Filter, and Normalization.

C. Sample Image

Some examples of working images are shown below, including 3 healthy and 3 diseased images. In Fig. 2, the leaves appear without any spots or cracks, indicating they are healthy. In Fig. 3, the leaves show many red spots, indicating they are diseased.

Fig. 2. Healthy Images

Fig. 3. Unhealthy Images

D. Feature Extractor

For feature extraction, here used many CNN-based pre-trained deep learning models. DenseNet121 is a deep learning model that has 121 layers. DenseNet169 has a deeper architecture, which captures more complex details using 169 layers. DenseNet201 is a more deep version that allows the model to learn accurate details with 201 layers and employs dense connections for better feature. VGG19 is a simpler model that used widely for identifying features in imagers using its 19 layers.

E. Classification

After extracting the features, here used machine learning based different kinds of classifiers to predict the plant's diseases-

- SVM: It is used to distinguish different categories depends on features like healthy or unhealthy. It handle high dimensional data that performed best with feature extractors. It is ideal for finding healthy and diseased leaves with high precision.
- Random Forest: It used to build multiple decision trees to make predictions. It is good for handling different types of data.
- Logistic Regression: Provides a baseline for comparison with more complex models.

DenseNet201 with SVM performed best because it can help to learn very detailed features from images and works well with high-quality images. They are helped to separate healthy and diseased leaves accurately. But random forest, logistic regression did not perform well because they couldn't handle complex features.

F. Performance Evaluation

Evaluating the model helps to understand how accurately the model can predict plant diseases. The Performance Evaluation methods are explained below-

Accuracy measures the model got correct how many predictions. If accuracy is easy to interpret, it can not be reliable all time especially if one class has more examples than others.

Precision measures the ratio of correctly predicted positive instances to the total predicted positive instances. This high precision is important to avoid false alarms.

Recall shows how many correctly identified diseased plants from the dataset. The recall is crucial for plant disease detection since missing diseased plants could lead to further crop damage.

F1-Score is the combination of precision and recall. If the number of healthy and diseased plants is imbalanced it can be used. This metric gives a balanced output of the model's work performance.

AUC used in binary classification works. Under the ROC curve AUC measures the area.

G. Computational Requirements

All the models trained on using Gigabyte Intel Arc A380 6GB GPU with 24GB RAM. Each model takes approximately 2 hours for training the dataset.

IV. Result & Discussion

In this section, the results of the study are explained focusing on feature extractors and classifiers which are used for plant disease detection. The evaluation metrics include accuracy, precision, recall, F1-score, and AUC. The full tested results are shown in Table II.

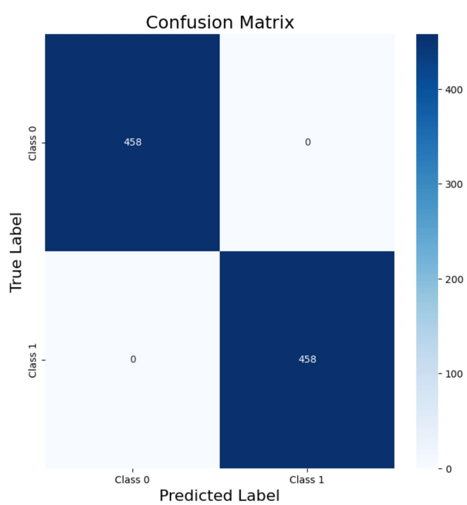

Fig. 4. Confusion Matrix

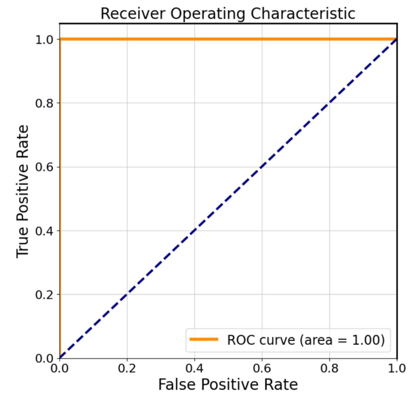

Fig. 5. ROC Curve

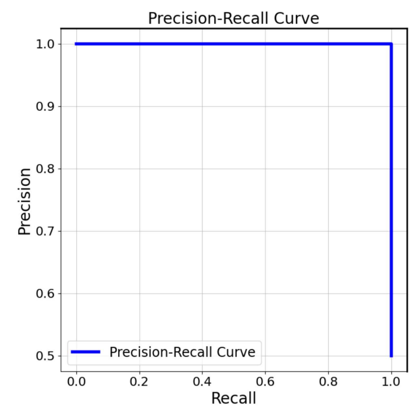

Fig. 6. Precision-Recall Curve

After observing all the data, found the best results combining features and classifiers. DenseNet201 with SVM provided the best results and achieving 100% accuracy . Precision, recall, F1-score and AUC were all perfect at 1.000. The 100% accuracy came beacuse the dataset has very clear with high quality images, making it easier for the model to classify the result and diseases. DenseNet201 with SVM achieving the 100% accuracy because the dataset has high quality images with clear background, well-segmented that may concern the potential overfitting images that means the model cannot perform better on real world datasets. But VGG19 + SVM performed slightly worse, likely due to their shallower depth and lower feature extraction capacity

The confustion matrix of the best result is provided in Fig. 4. ROC Curve measures the model can separate healthy and diseased plants accurately in Fig. 5. The orange line is in the top-left corner which is the better for the model. The Precision-Recall curve shows both precision and recall reach the highest value of 1.0 showing the model is highly accurate as demonstrated in Fig. 6.

The Table-III showing the compared results with my best achieved results.

TABLE II
PERFORMANCE METRICS FOR MODELS.

Models	Accuracy	Precision				Recall				F1 Score				AUC
		Healthy	Unhealthy	Mac. AVG	Wtd AVG	Healthy	Unhealthy	Mac. AVG	Wtd AVG	Healthy	Unhealthy	Mac. AVG	Wtd AVG	
DenseNet121 + SVM	0.9956	0.9913	1.0000	0.9957	0.9957	1.0000	0.9913	0.9956	0.9956	0.9957	0.9956	0.9956	1.0000	1.0000
DenseNet121 + Random Forest	0.9618	0.9434	0.9818	0.9626	0.9626	0.9825	0.9410	0.9618	0.9618	0.9626	0.9610	0.9618	0.9981	1.0000
DenseNet121 + Logistic Regression	0.9934	0.9892	0.9978	0.9935	0.9935	0.9978	0.9891	0.9934	0.9934	0.9935	0.9934	0.9934	1.0000	1.0000
DenseNet169 + SVM	0.9989	0.9978	1.0000	0.9989	0.9989	1.0000	0.9978	0.9989	0.9989	0.9989	0.9989	0.9989	0.9989	1.0000
DenseNet169 + Random Forest	0.9738	0.9597	0.9887	0.9742	0.9742	0.9891	0.9585	0.9738	0.9738	0.9742	0.9734	0.9738	0.9738	0.9981
DenseNet169 + Logistic Regression	0.9989	0.9978	1.0000	0.9989	0.9989	1.0000	0.9978	0.9989	0.9989	0.9989	0.9989	0.9989	0.9989	1.0000
DenseNet201 + SVM	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
DenseNet201 + Random Forest	0.9760	0.9599	0.9932	0.9766	0.9766	0.9934	0.9585	0.9760	0.9760	0.9764	0.9756	0.9760	0.9760	0.9984
DenseNet201 + Logistic Regression	0.9989	0.9978	1.0000	0.9989	0.9989	1.0000	0.9978	0.9989	0.9989	0.9989	0.9989	0.9989	0.9989	1.0000
VGG19 + SVM	0.9880	0.9827	0.9934	0.9880	0.9880	0.9934	0.9825	0.9880	0.9880	0.9881	0.9879	0.9880	0.9880	0.9988
VGG19 + Random Forest	0.9585	0.9449	0.9730	0.9589	0.9589	0.9738	0.9432	0.9585	0.9585	0.9591	0.9579	0.9585	0.9585	0.9915
VGG19 + Logistic Regression	0.9891	0.9828	0.9956	0.9892	0.9892	0.9956	0.9825	0.9891	0.9891	0.9892	0.9890	0.9891	0.9891	0.9990

TABLE III
COMPARISON WITH THE OUTCOME OF RELATED WORKS

Reference	Models	Accuracy	Precision	Recall	F1-Score	AUC
[5]	AlexNet + Connected Layers	96.63%	92%	91%	91%	No Report
[6]	Attention CNN + Soft Attention Mechanism	99.04%	99%	99%	99%	99.90%
[7]	CNN + Rnadom Forenst	98%	91.91%	91.91%	91.91%	No report
[8]	MobileNet + Connected Layers	96%	90%	89%	89%	No report
[]	DenseNet201 + SVM	100%	100%	100%	100%	100%

V. LIMITATIONS & FUTURE WORK

This study produced good results but faced some challenges. Below are the main limitations and future directions:

1) Here PlantVillage dataset used which has high-quality images with edited backgrounds but not real-world conditions. In future, real-world images surely collected and used for practical purposes.
2) The models can perform very well using PlantVillage datasets, but their performance or real image is uncertain. In the future, testing the models with the collected images from farms with complex background, real environmental images.
3) This model have the limited diseases detection system. In futer, using the multi-level classification for detect plant diseases.
4) The study focused on research purposes only, Future efforts will aim to optimize the model including faster predictions, user-friendly applications.
5) This work has no cross validation results. So the main future goal is to find cross validation results with real world images collecting plant leaf images directly from farms under natural conditions which may introduce many variations of leaf background.

VI. CONCLUSION

This study has been conducted to detect plant diseases. Its main objective is that the farmers should be able to detect the disease very easily. In this study, CNN-based deep learning methods and machine learning methods have been used to identify plant diseases accurately.

The focus of the project is to make it useful in real-life situations. Future works aim to improves different conditional use, make it more faster and increase its ability to detect various types of diseases. This project is a step toward helping farmers to grow healthier crops and reduce losses from diseases.

REFERENCES

[1] Food and Agriculture Organization (FAO). (2022). Impact of plant diseases on global agriculture. Retrieved from FAO.

[2] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” Front. Plant Sci., vol. 7, p. 1419, 2016.

[3] A. Gupta, “Rice bacterial blight impact,” Journal of Crop Diseases, 2018.

[4] J. Smith, “Economic analysis of late blight,” Agricultural Economics, 2020.

[5] “ Puja Dey, Tanjim Mahmud, Sultana Rokeya Nahar,” in Plant Disease Detection in Precision Agriculture: Deep Learning Approaches, Karl Andersson, 2024.

[6] An Improved Lightweight Attention Convolutional Neural Network Model for Tomato Leaf Diseases Classification for Low-End Devices Published 2024 • Mahmoud H. Alnamoly, Anar A. Hady, Sherine M. Abd El-Kader.

[7] R. Cnn, D. Forest, V. Banerjee, and P. Kukreja, CNN and Random Forest-Based Tomato Disease Urgency Prediction: A Vital Tool for Agriculture Published Jan 24, 2024 · D. 2024.

[8] Deep Learning for Plant Disease Detection Published Nov 18, 2023 • Munaf Mudheher Khalid. Oguz Karan.

[9] R. H. Prince et al., “CSXAI: a lightweight 2D CNN-SVM model for detection and classification of various crop diseases with explainable AI visualization,” Front. Plant Sci., vol. 15, 2024.

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