THIRD YEAR B.TECH. MINI PROECT REPORT

A report submitted in partial fulfilment of the requirements for the Award of Degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

By

BT21CSE194 : Gyanbardhan

PLANT DISEASE DETECTION GENERALIZATION

Course Name: Machine Learning

Course Code:CSL422

भारतीय सूचना प्रौद्योगिकी संस्थान, नागपुर

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY,

NAGPUR

(An Institution of National Importance by Act of Parliament)

Survey No.140,141/1 behind Br.Sheshrao Wankhede Shetkari Sahkari

SootGirni,Nagpur, 441108

TABLE OF CONTENTS

1. Introduction

1.1. Introduction

Agriculture plays a significant role in every nation's economy by producing crops. Plant disease identification is one of the most important aspects of maintaining an agriculturally developed nation. The timely and efficient detection of plant diseases is essential for a healthy and productive agricultural sector and to prevent wasting money ,other resources.and preventing economic losses. In this study, we explore the efficiency of various deep learning models, including CNN, AlexNet, VGG-16,VGG-19, ResNet, DenseNet, EfficientNet, and ConvNextLarge, in discerning leaf infections. Leveraging a dataset of over 20,000 images, we assess the accuracy rates of these models, revealing promising results ranging from 96.4% to 99.8%. Furthermore, employing bagging techniques, we amalgamate the predictive capabilities of AlexNet ,VGG-16 and VGG-19, yielding enhanced accuracy and robustness. Motivated by these findings, we propose an application designed to empower farmers in diagnosing plant diseases through leaf image analysis. Central to this application is the utilization of transfer learning models, enabling the discrimination between healthy and infected leaves. By facilitating early disease detection ,our envisioned application aims to foster agricultural sustainability, resource optimization, and economic resilience for farmers worldwide.

1.2. Problem Statement

Developing a rapid, accurate, and accessible method for identifying plant diseases is crucial for agricultural productivity and economic stability. However, the lack of an efficient crop disease prediction system hampers efforts. This project aims to create a versatile deep learning-based model for timely crop disease detection. By analyzing crop images, the model will provide early disease identification, aiding farmers in effective management and contributing to sustainable agriculture and food security.

2. Related Studies

2.1. Existing works

<u>Paper 1</u>

Agronomy | Free Full-Text | Deep Learning-Based Leaf Disease Detection in Crops Using Images for Agricultural Applications (mdpi.com)

In this paper, the team of Agronomy successfully analysed the different transfer learning models suitable for the accurate classification of 38 different classes of plant disease. Standardization and evaluation of state-of-the-art convolutional neural networks using transfer learning techniques were undertaken based on the classification accuracy, sensitivity, specificity, and F1 score. From the performance analysis of the various pre-trained architectures, it was found that DenseNet-121 outperformed ResNet-50, VGG-16, and Inception V4. Training the DenseNet-121 model seemed to be easy, as it had a smaller number of trainable parameters with reduced computational complexity. Hence, DenseNet-121 is more suitable for plant disease identification when there is a new plant disease that needs to be included in the model, demonstrating reduced training complexity. The proposed model achieved a classification accuracy of 99.81% and F1 score of 99.8%.

The result of various models is as follows :

References	Dataset Used	Pre-Trained Model	Multi-Classes	Recognition Accuracy (%)
[53]	PlantVillage	VGG-16	10	91.2
[54]	PlantVillage	ResNet-50	6	97.1
[55]	PlantVillage	AlexNet	7	98.8
Our Work	PlantVillage	Inception V4	38	97.59
		VGG-16	38	82.75
		ResNet-50	38	98.73
		DenseNet-121	**38**	**99.81**

Paper 2

[DeepCrop: Deep learning-based crop disease prediction with web application - ScienceDirect]

The proposed method for detecting plant diseases encompasses a comprehensive process, starting with the collection of input images representing various types of leaves, followed by preprocessing to mitigate noise and ensure data integrity. Training and model construction involve training transfer learning models using a dataset and validating the architecture's performance using test images. The model's evaluation involves splitting data for training and testing, followed by validation to confirm detection accuracy. Finally, performance evaluation considers metrics such as accuracy, precision, recall, and F1-score, leading to the selection of the most efficient model, ResNet-50, for plant disease detection. The study's extensive experiments with convolutional neural network architectures, including CNN, VGG-16, VGG-19, and ResNet-50, using the "plant-village" dataset.

In the experiment, we got the accuracy rate of 98.60%, 92.39%, 96.15%, and 98.98% for CNN, VGG-16, VGG-19 and ResNet-50 models respectively showcase promising accuracy rates.

Among all models, RestNet50 provides a better accuracy rate to detect leaf disease efficiently. So, we employed the proposed higher accuracy model for our web app development to correctly detect plant leaf disease. The web application provides a smart agriculture system for detecting the disease. Hence, the proposed method produced better outcomes for estimating symptom severity than earlier investigations of plant leaf disease

Paper 3

[Frontiers | An effective approach for plant leaf diseases classification based on a novel DeepPlantNet deep learning model (frontiersin.org)]

The research paper delves into the realm of plant disease detection by leveraging advanced deep learning techniques, particularly focusing on digital image processing. The introduced model, DeepPlantNet, presents a novel approach for effectively identifying diseases in plant leaves. Utilizing the Plant Village dataset from Kaggle, which contains images from eight different plant types, the study meticulously outlines its methodology. With a dataset split into training and testing sets, DeepPlantNet's architecture, consisting of 25 convolutional layers, is meticulously optimized through hyperparameter tuning and rigorous training to mitigate overfitting. The evaluation of DeepPlantNet's performance showcases remarkable accuracy, recall, precision, and F1-score, ranging from 93.87% to 98.49% across different disease categories, and achieving near-perfect scores of over 99% in six disease categories.

By employing robust techniques such as k-fold cross-validation and hyperparameter optimization, the research ensures the reliability and generalizability of DeepPlantNet's results. The model's exceptional performance in classifying plant leaf diseases underscores its potential as a valuable tool in agricultural settings, aiding in early disease detection and management. Moreover, the study's meticulous approach in dataset preprocessing and model training sets a strong precedent for future research in the domain of deep learning-based plant disease detection systems, offering promising avenues for further exploration and refinement in agricultural technology and precision farming practices.

Paper 3

[IRJET-V8I2110.pdf]

The paper underscores the critical role of agriculture in India's economy, emphasizing its significant contribution to the GDP, exports, and employment. With over 60% of the population dependent on agriculture, ensuring the health and productivity of crops becomes paramount. However, various diseases pose a substantial threat to crop quality and yield, potentially leading to food scarcity. To address this challenge, the paper outlines a methodology leveraging advanced technologies such as image processing and machine learning for the automatic detection and classification of plant diseases. By employing techniques like image preprocessing, CNN, decision trees, and random forests, the proposed system aims to accurately identify disease symptoms on plant leaves and recommend appropriate solutions, such as pesticides or chemicals, to mitigate crop damage.

The proposed methodology offers a systematic approach to tackle the complexities of crop disease detection, emphasizing the importance of early identification to prevent widespread crop losses. By harnessing the power of machine learning algorithms and image processing techniques, the system aims to provide farmers with timely insights into crop health, enabling them to make informed decisions and take proactive measures to safeguard their harvests. Ultimately, the development of such a system holds promise for enhancing agricultural productivity, sustainability, and resilience in the face of evolving environmental and economic challenges, thereby contributing to the broader goal of ensuring food security for India's growing population.

2.2. Research Gap

The research gap identified in this study pertains to the absence of an approach that utilizes multiple models for disease prediction in plant and leaf detection systems. While the study leverages convolutional neural network (CNN) architectures for disease identification, it primarily focuses on evaluating individual models such as AlexNet, VGG-16, VGG-19, ResNet, DenseNet, EfficientNet, and ConvNextLarge. However, the research overlooks the potential benefits of combining the predictive capabilities of these models to enhance overall accuracy and reliability in disease detection.

By employing a single-model approach, the study may miss out on leveraging the diverse strengths and characteristics offered by different CNN architectures. Consequently, there remains an unexplored opportunity to develop a methodology that integrates the predictions from multiple models, potentially leading to improved disease detection performance

3. Problem Definition

3.1. Problem Statement

Developing a rapid, accurate, and accessible method for identifying plant diseases is crucial for agricultural productivity and economic stability. However, the lack of an efficient crop disease prediction system hampers efforts. This project aims to create a versatile deep learning-based model for timely crop disease detection. By analyzing crop images, the model will provide early disease identification, aiding farmers in effective management and contributing to sustainable agriculture and food security

3.2. Flowchart/Block diagrams

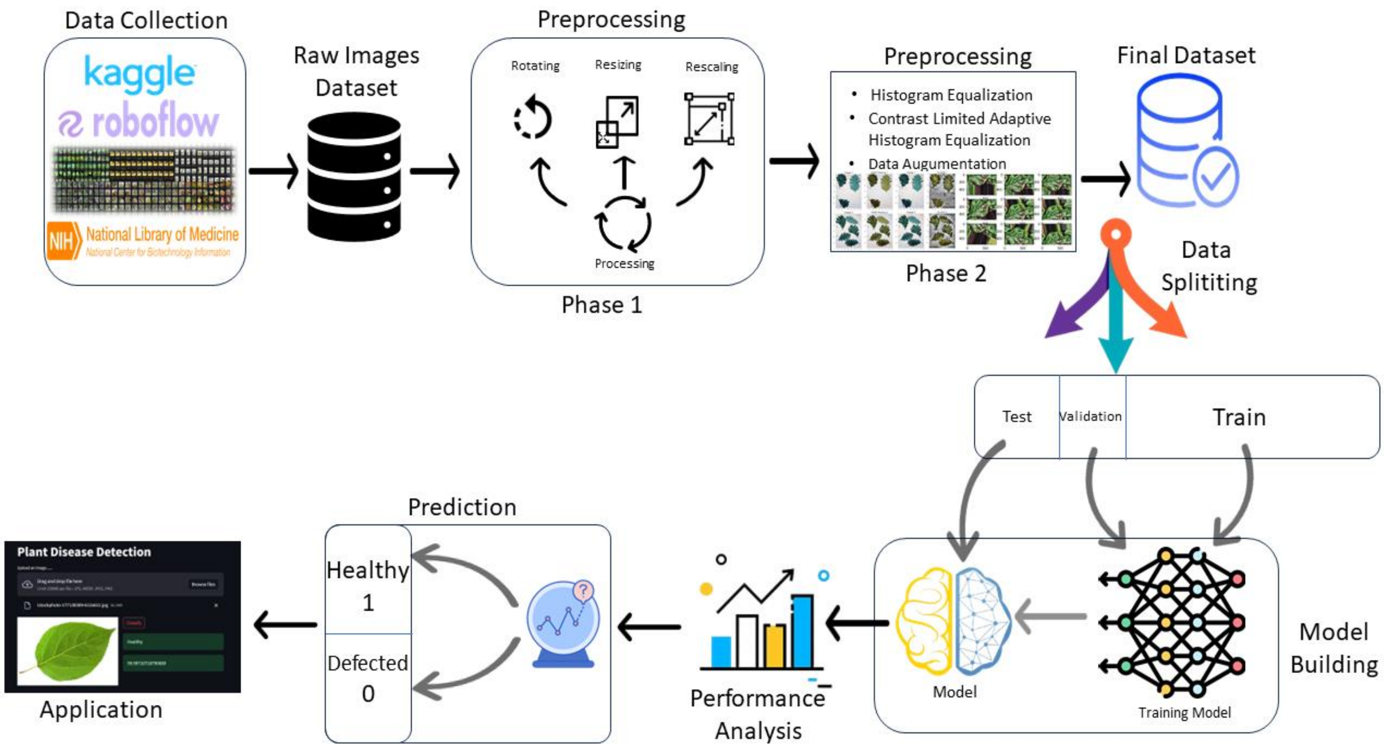


Fig. 1 Flow Chart of Plant Disease Detection

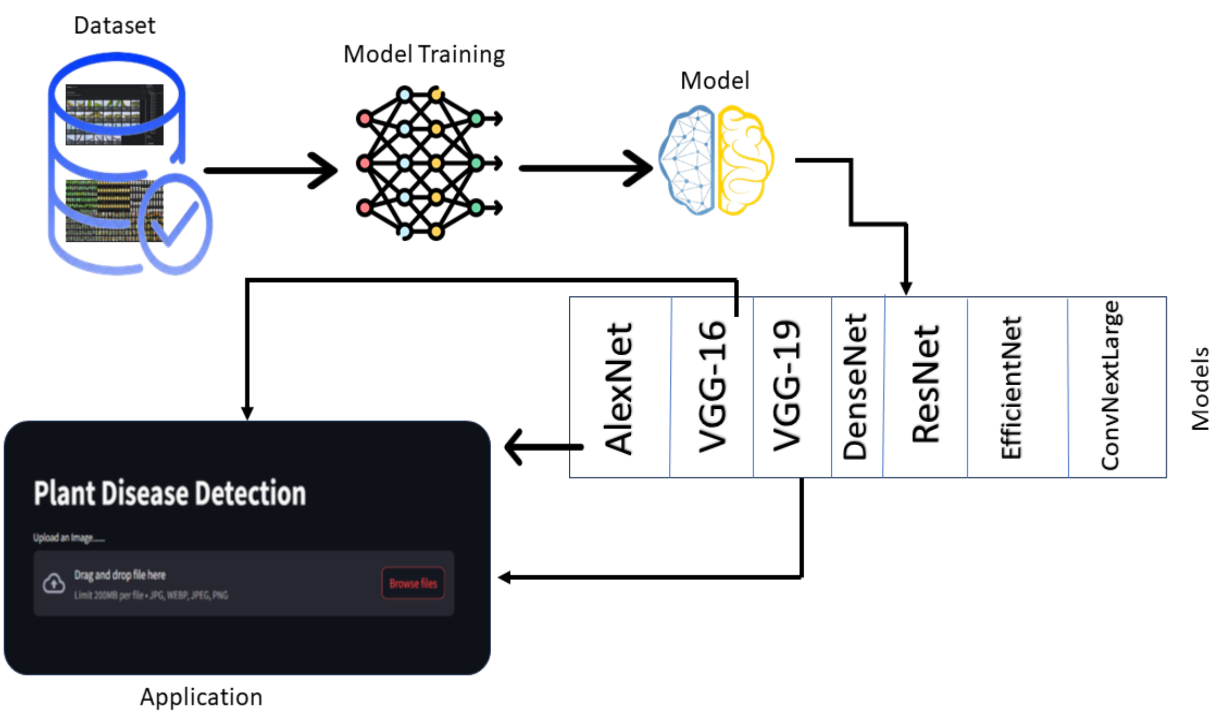


Fig. 2 Flow Chart Model Building

3.3. Dataset Description

The dataset is divided into three main splits: training, validation, and testing, each serving specific purposes in training and evaluating the deep learning model for plant disease prediction.

- **Training Split**:
  - The training split comprises 18,741 images, with 13,128 images depicting defected leaves and 5,613 images representing healthy leaves.
  - These images have undergone preprocessing steps including resizing, rotation, rescaling, and Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance their features and aid in model training.
- **Validation Split**:
  - The validation split consists of 1,410 images, with 900 images showing defected leaves and 510 images showcasing healthy leaves.
  - Similar to the training set, the validation images have been preprocessed using resizing, rotation, rescaling, and CLAHE techniques to ensure consistency and effectiveness in model evaluation.
- **Testing Split**:
  - The testing split contains over 4,000 images, comprising both defected and healthy leaves.
  - These images serve as unseen data to assess the model's generalization ability.
  - Like the training and validation sets, the testing images have been preprocessed to align with the preprocessing steps applied during training and validation phases.

The preprocessing steps, including resizing, rotation, rescaling, and CLAHE, have been employed standardize the images, enhance their features, and mitigate the effects of variations in lighting, orientation, and size, thus facilitating robust and accurate predictions by the deep learning model.

➢ Dataset source and attribute information

The dataset used in this study comprises images sourced from various repositories and platforms, including Kaggle, Roboflow,National Library of Medicine,etc.. These diverse sources ensure the inclusion of a wide range of plant diseases, leaf types, and environmental conditions, enhancing the dataset's representativeness and utility for training and evaluating the deep learning model for plant disease prediction.

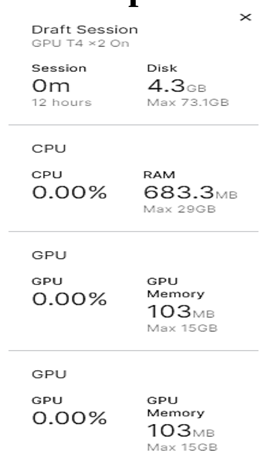
- *Kaggle Plant Village Dataset*
  - Augumented Dataset
  - All .jpg format
  - Size (5kB-50kB)
  - https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset
- *Kaggle New Plant Disease Dataset*
  - Augumented Dataset
  - All .jpg format
  - Size (15kB-100kB)
  - https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset
- *Roboflow Plant Disease Dataset*
  - .jpg,.png,.jpeg format images
  - Size (5kB-100kB)
  - https://universe.roboflow.com/learning-eri4b/plant-disease-tmyq8/dataset/4

3.4. Experimental Setup

The experiments were performed on a Kaggle platform with GPU T4x2 15GBx2,CPU 15GB, 29 GB RAM, 73.1 GB Disc Storage,a Core i5 processor, 64-bit Windows 11, and the Python programming language. Different deep learning models are explored and evaluated with respect to their performance based on the metrics using the confusion matrix

Parameters	Values
Learning Rate	0.00001
Drop Out	0.3
Optimizer	RMSProp
Batch Size	32
Epochs	15
Callback	Early Stopping
Regularization	L2

Description



4. Model Building

Image Preprocessing

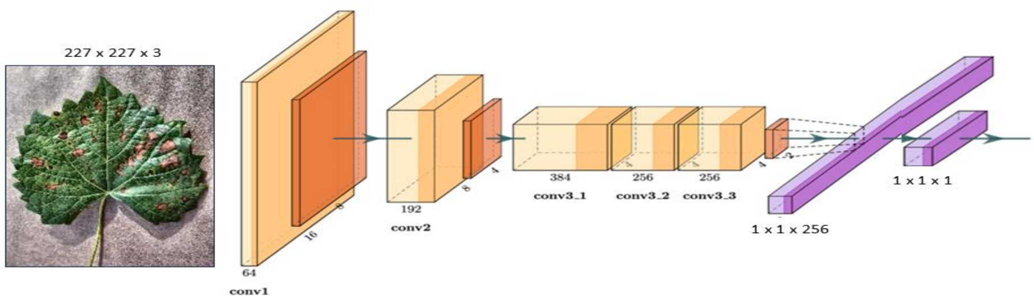
Image preprocessing serves two primary purposes in the context of plant and leaf disease detection. Firstly, it standardizes the input dimensions by resizing images to ensure uniformity across the dataset. This standardization is essential for compatibility with various convolutional neural network (CNN) architectures, such as AlexNet, VGG-16, VGG-19, ResNet, DenseNet, EfficientNet, and ConvNextLarge, each with specific input size requirements. By maintaining consistent image sizes, preprocessing facilitates seamless integration with different CNN models, optimizing the training process and enhancing model performance. Secondly, image preprocessing enhances the quality of images by improving feature visibility and discriminative power. Techniques like histogram equalization and contrast adjustment mitigate the effects of uneven lighting conditions, thereby aiding in effective disease detection and classification.

- **Resizing**: Images resized to dimensions suitable for different CNN architectures. Dimensions for AlexNet: 227x227 pixels. Dimensions for other architectures (VGG-16, VGG-19, ResNet, DenseNet, EfficientNet, ConvNextLarge): 150x150 pixels.
- **Rescaling**: Normalization of pixel intensity values within the dataset. Promotes consistent feature representation across images.
- **Rotation**: Augmentation technique to increase dataset variability. Helps in enhancing model robustness and generalization.
- **Augmentation** : for dataset variability; Increases diversity within the dataset. Enhances model's ability to generalize to unseen data.

- **Histogram Equalization**: Technique to improve image contrast. Enhances visibility of features within images.
- **Contrast Limited Adaptive Histogram Equalization (CLAHE)**: Adaptive version of histogram equalization. Effective in mitigating effects of uneven lighting conditions.

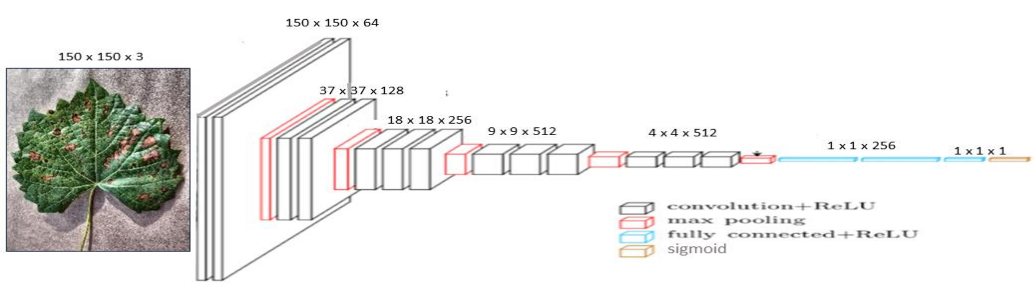
Model Implementation

- **AlexNet**



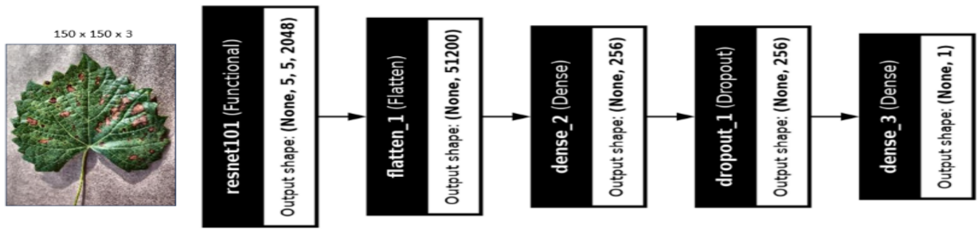
AlexNet Fine Tuned

- **VGG-16,VGG-19**



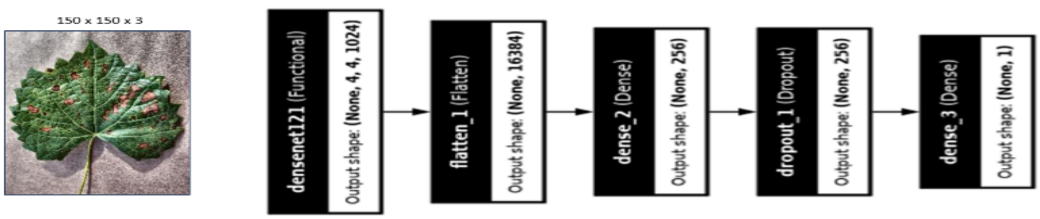
VGG-16 Fine Tuned

- **ResNet**



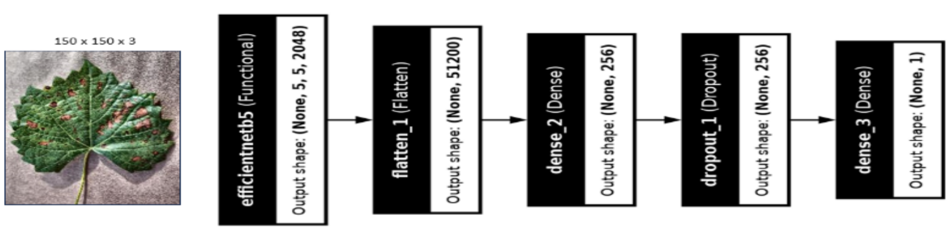
ResNet Fine Tuned

- **DenseNet**



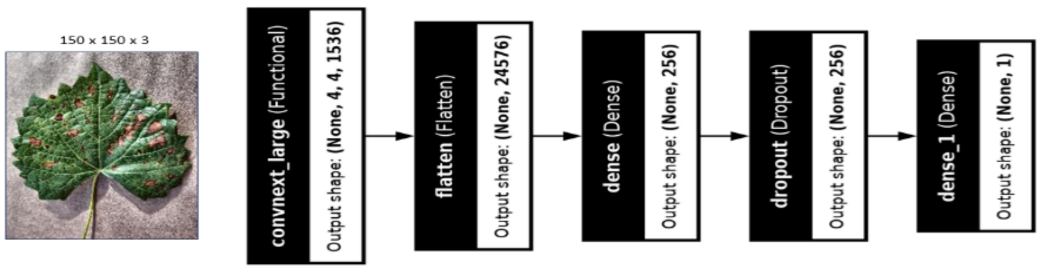
DenseNet Fine Tuned

- **EfficientNet**



EfficientNet Fine Tuned

- **ConvNextLArge**



ConvNextLarge Fine Tuned

Properties	***AlexNet***	***VGG-16***	***VGG-19***	***ResNet***	***DenseNet 121***	***EfficientNet-B1***	***ConvNextLarge***
Image	227x227x3	150x150x3	150x150x3	150x150x3	150x150x3	150x150x3	150x150x3
Weight	Trained	Imagenet	Imagenet	Imagenet	Imagenet	Imagenet	Imagenet
Model Size (Parameters)	162.31 MB	64.13 MB	84.39 MB	162.73 MB	26.85 MB	108.77 MB	748.56 MB
Activation Function	sigmoid	sigmoid	sigmoid	sigmoid	sigmoid	sigmoid	sigmoid
Total Parameters	42549121	16812353	22122048	42658176	7037504	28513527	196230336
Total Layers	13	22	25	51	101	460	-
Max/Average Pool	3	5	5	4	4	1	-

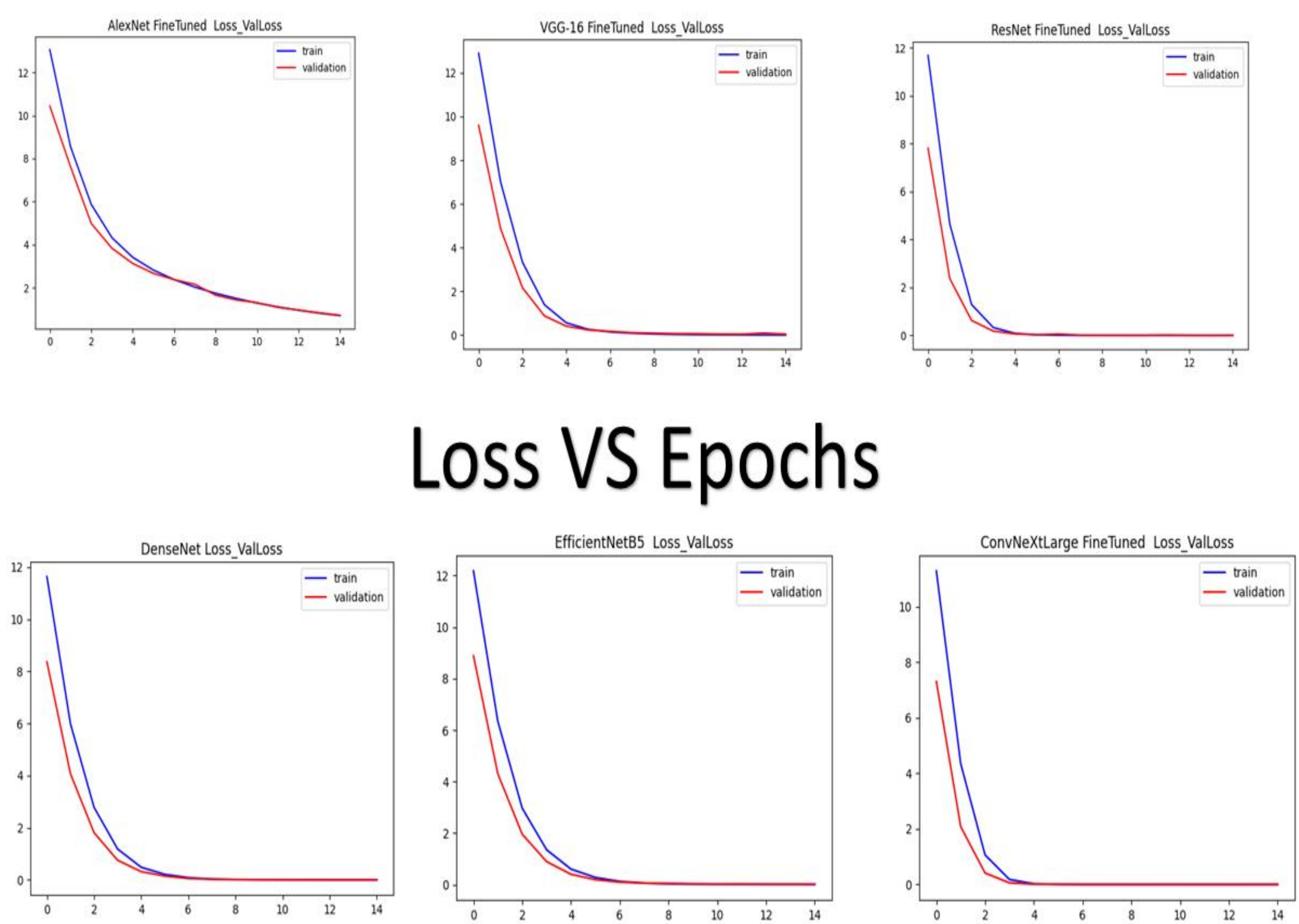
Deployment

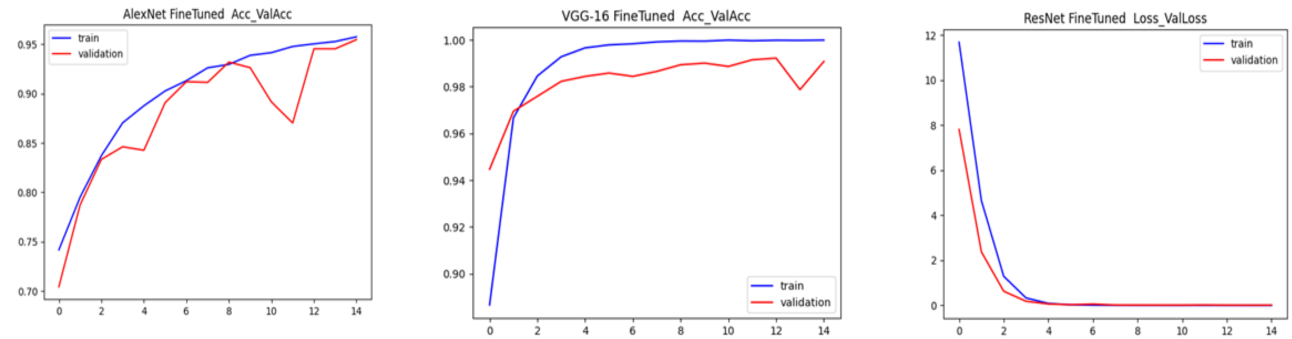
The deployment platform for the plant and leaf disease detection system utilizes Streamlit, a user-friendly web application framework, to provide an intuitive interface for end-users. Upon accessing the deployed application, users can interactively upload images of plant leaves for disease classification. Behind the scenes, the application leverages three pre-trained convolutional neural network (CNN) models: AlexNet, VGG-16, and VGG-19. These models have been fine-tuned and optimized for disease detection tasks. Upon uploading an image, the application processes it through all three models, extracting relevant features and generating individual predictions for each model. Subsequently, a weighted average of the predictions from AlexNet, VGG-16, and VGG-19 is computed, providing a consolidated output that represents the collective classification confidence of the models. Finally, users can tap the "classify" button to trigger the classification process, which yields the predicted disease status of the uploaded plant leaf. This seamless integration of multiple models and the weighted averaging technique ensures robust and accurate disease detection results, empowering users with actionable insights for plant health management

5. Results and discussion

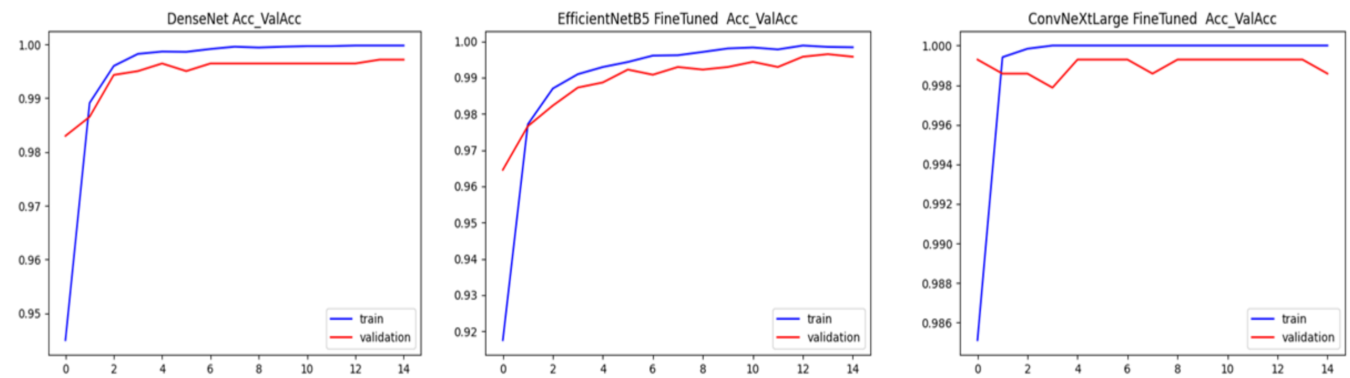
Performance metrics

Metrics	*AlexNet*	*VGG-16*	*VGG-19*	*ResNet*	*DenseNet*	*EfficientNet*	*ConvNextLarge*
ValidationAcc	0.964	0.9807	0.985	0.9864	0.9971	0.993	0.9985
Testing Acc	0.435	0.8472	0.8933	0.9086	0.8192	0.8302	0.86
precision	0.88	0.87	0.88	0.93	0.91	0.86	0.89
Recall	0.42	0.85	0.89	0.91	0.87	0.83	0.86
F1-score	0.49	0.86	0.89	0.91	0.88	0.84	0.87





Accuracy VS Epochs



Results and explanation:

- **Validation Accuracy**:

  All models perform well in terms of validation accuracy, with ConvNextLarge achieving the highest accuracy of 99.85%, followed by ResNet (98.64%) and EfficientNet (99.3%). AlexNet has the lowest validation accuracy (96.4%).

- **Testing Accuracy**:

  Among the models, ResNet has the highest testing accuracy (90.86%), followed closely by VGG-19 (89.33%). AlexNet has the lowest testing accuracy (43.5%).

- **Precision, Recall, and F1-score**:

  VGG-19 achieves the highest precision, recall, and F1-score, followed closely by ResNet. This indicates that VGG-19 and ResNet perform well in terms of both positive prediction and capturing all positive instances. AlexNet has the lowest precision, recall, and F1-score among the models.

  We have chosen VGG16,VGG19 and AlexNet ,taking average of the prediction gives the best prediction. AlexNet is chosen because it predicts accurately for healthy instances

Model Combination Approach(Bagging Approach):

The application plans to take the weighted average prediction of VGG-16, VGG-19, and AlexNet model in ratio 2:3:4. VGG-19 performs the best among the models, so more weight to its prediction in the averaging process

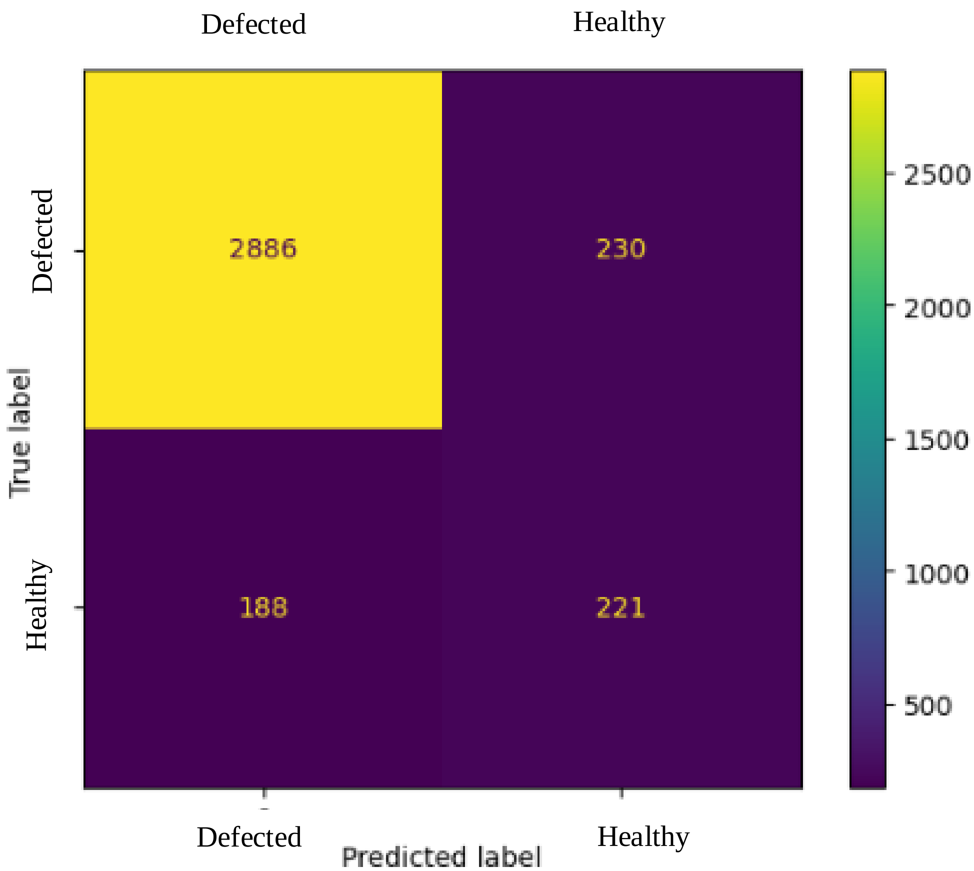
Y_pred=(2*Y_an + 3*Y_v16 + 4*Y_v19)/9

Y_an, Prediction of AlexNet
Y_v16, Prediction of VGG-16
Y_v19, Prediction of VGG-19

Overall Model Performance:

The model mentioned at the end (precision: 0.89, recall: 0.88, F1-score: 0.88, testing accuracy: 0.881418) shows decent performance across the board

```
test accuracy: 88.1418 %
```



	precision	recall	f1-score	support
Defected	0.94	0.93	0.93	3116
Healthy	0.49	0.54	0.51	409
accuracy			0.88	3525
macro avg	0.71	0.73	0.72	3525
weighted avg	0.89	0.88	0.88	3525

6. Conclusion

Plant and Leaf disease detection problems are crucial and challenging problems in agriculture worldwide. This study utilized Convolutional neural networks-based architecture to identify leaf disease and its source. Several models, such as AlexNet, VGG-16,VGG-19, ResNet, DenseNet, EfficientNet, and ConvNextLarge architectures, are adopted to detect leaf conditions. Extensive experiments were conducted with our plant disease dataset. In the experiment, we got the accuracy rate of 96.4%,98.07%,98.5%,98.64%,99.71%,99.3% and 99.85%

for AlexNet, VGG-16,VGG-19, ResNet, DenseNet, EfficientNet, and ConvNextLarge models respectively. Among all models, ConvNextLarge provides a better accuracy rate to detect leaf disease efficiently.

We employed the The application plans to take the weighted average prediction of VGG-16, VGG-19, and AlexNet model in ratio 3:4:2. The application provides a smart agriculture system for detecting the disease. Hence, the proposed method produced better outcomes for estimating symptom severity than earlier investigations of plant leaf disease.

In the future, we want to improve the accuracy rate by developing a new hybrid deep-learning architecture using computer vision developing a localization method to identify the area of disease in the leaf. Besides, we will use better multiple-leaf disease datasets so that it includes all other plant-leaf disease dataset.

7. References

[1]. Islam, Md. Manowarul, Adil, Md. Abdul Ahad, Talukder, Md. Alamin, Ahamed, Md. Khabir Uddin, Uddin, Md n Ashraf, Hasan, Md. Kamran, Sharmin, Selina, Rahman, Md. Mahbubur, Debnath, Sumon Kumar. "DeepCrop: Deep learning-based crop disease prediction with web application." Journal of Agriculture and Food Research, vol. 14, 2023, p. 100764, ISSN 2666-1543, https://www.sciencedirect.com/science/article/pii/S2666154323002715

[2]. Andrew, J., Jennifer Eunice, Daniela Elena Popescu, M. Kalpana Chowdary, and Jude Hemanth. "Deep Learning-Based Leaf Disease Detection in Crops Using Images for Agricultural Applications." Agronomy, vol. 12, no. 10, 2022, p. 2395, https://www.mdpi.com/2073-4395/12/10/2395

[3]. Yadav, Nikita, Kasar, Shivani, Abuj, Dnyaneshwari, Vadvale, Ankita, and Dr. S.C. Dharmadhikari. "Crop Disease Prediction and Solution." International Research Journal of Engineering and Technology (IRJET), vol. 08, no. 02, Feb 2021, e-ISSN: 2395-0056, p-ISSN: 2395-0072, https://www.irjet.net/archives/V8/i2/IRJET-V8I2110.pdf

[4]. Ullah, Naeem, Khan, Javed Ali, Almakdi, Sultan, Alshehri, Mohammed S., Al Qathrady, Mimonah, El-Rashidy, Nora, El-Sappagh, Shaker, and Ali, Farman. "An effective approach for plant leaf diseases classification based on a novel DeepPlantNet deep learning model." Frontiers in Plant Science, Technical Advances in Plant Science section, vol. 14, 2023, Frontiers | An effective approach for plant leaf diseases classification based on a novel DeepPlantNet deep learning model (frontiersin.org)

8. Important Links

- Web Application - *https://plant-disease-detection-h2o8skyxaf2g2f5n6myucs.streamlit.app/*
- Notebooks-
  - AlexNet- *https://www.kaggle.com/code/gyanbardhan/alexnet*
  - VGG16- *https://www.kaggle.com/code/gyanbardhan/project-vgg16-2*
  - VGG19- *https://www.kaggle.com/code/clay108/vgg19*
  - DenseNet- *https://www.kaggle.com/code/clay108/project1-densenet121*
  - ResNet- *https://www.kaggle.com/code/gyanbardhan/project1-resnet*
  - EfficientNet- *https://www.kaggle.com/code/clay108/project1-efficientnetb5*
  - ConvNextLarge- *https://www.kaggle.com/code/gyanbardhan/project1-convnextlarge*
  - Final Model- *https://www.kaggle.com/code/clay108/evaluation-of-vgg16-19-an*
- Final Dataset-
  - Train/Valid- *https://www.kaggle.com/datasets/gyanbardhan/clahe-plant-disease*
  - Test- *https://www.kaggle.com/datasets/gyanbardhan/groundnuttest*
- Models-
  - AlexNet- *https://www.kaggle.com/datasets/gyanbardhan/alexnet123*
  - VGG16- *https://www.kaggle.com/datasets/gyanbardhan/vgg16*
  - VGG19- *https://www.kaggle.com/datasets/clay108/vgg19-123*