Early Identification and Classification of Plant Diseases using CNN Models

Anant Shukla
Concordia University
1455 Boul. de Maisonneuve Ouest, Montreal
me@anantshukla.com

Avni Gupta
Concordia University
1455 Boul. de Maisonneuve Ouest, Montreal
avni.gupta12@gmail.com

Tsega Weldu
Concordia University
1455 Boul. de Maisonneuve Ouest, Montreal
tsegawe55@gmail.com

Mina Alizadeh
Concordia University
1455 Boul. de Maisonneuve Ouest, Montreal
mina.alizadeh94@gmail.com

1. Abstract

The present era is characterized by famines and climate changes, hence plant disease identification and classification have become a necessity. Disease identification by domain experts is time-consuming and costly and cannot be used on a global scale. Due to this, early identification and classification of plant disease is an important research area.

This study uses three datasets that are distinct in size, number of classes and captured in different environmental conditions. Both Training from Scratch and Transfer Learning are explored. Three CNN models, VGG16, ResNet50, and AlexNet are trained on the three datasets. Due to the variety of challenges in different datasets, image preprocessing techniques such as online and offline data augmentation and image normalization have been used.

The study aims to compare the performance by using metrics like time, loss and accuracy plots. Evaluation parameters used for comparison are Accuracy, Precision, Sensitivity Recall, and F1 score.

2. Introduction

As the human population is increasing at an alarming rate, the demand for agricultural products is also increasing significantly. Agriculture is constantly affected by pathogens such as viruses, bacteria, fungi and weeds that cause production and economic losses [10]. Early disease identification and classification helps in health monitoring and taking preventive measures in minimizing the risk of disease spreading and better yield [8].

Plant disease classification using artificial intelligence can prove to be a better resort as it is inexpensive and scalable. Research in the area of deep learning has great potential in terms of increased accuracy [16]. This project aims to classify a few diseases, affecting apple, tomato, blueberry, cherry, corn, grape, orange, peach, potato, and strawberry crops.

However, the main challenges faced in classification are limited dataset availability, complex image backgrounds with surrounding stems, their shades on leaves, and the same disease on look-alike plants making it difficult to see them apart [14]. Some of the datasets have low-quality images collected from the internet which are corrupted by image watermarking and background conditions.

Some research has been done in order to deal with these issues. Briefly, one study proposed combining WGAN-GP with LSR to improve prediction accuracy and address overfitting. They improved the accuracy by 24.4% as compared to 20.2% using classic data augmentation and 22% using synthetic samples without LSR. However, they required significant computational resources and suggested that the problem can be addressed using pre-trained models [18] Another study proposed monitoring plants under field conditions [22]. One used feature extracted from the leaves of diseased plants [24]. Another research used neural networks for the identification of morphological patterns of leaf veins [23]. Most of the research is done on PlantVillage dataset which is captured in a controlled lab environment. In this study, experiments have been conducted on three datasets which were captured in different environmental conditions.

This study tries to address the issue by comparing the results of various deep CNNs on image datasets. Challenges of low-image quality, complex image backgrounds and availability of a small number of images were addressed using various image preprocessing techniques. Hyperparameter tuning over batch size and the learning rate is also done to come up with the parameters that give better results. Additionally, the study tries to compare the results of transfer learning with training from scratch.

The three datasets used are PlantVillage, PlantDoc, and PlantaeK. The datasets are pre-processed to clean the input images, decrease the resolution and enhance the variety of input images wherever required. Then the datasets are fed to various CNN models. The CNN models are chosen based on their suitability to the problem domain and the size of the input dataset. The CNN models ResNet50, VGG16 and AlexNet are selected for both transfer learning and training from scratch.

The datasets are divided into train, validation and test sets. Using the validation set ensures that the model performance is evaluated during the training process. Accuracy curves and loss curves of training and validation datasets are compared to identify the problems of underfitting/overfitting wherever they surface. For the evaluation, various metrics such as model training time, accuracy, precision, recall, F1 score, and confusion matrix are employed.

For visualization purposes, t-SNE distribution is plotted to analyze the class distribution.

2.1. Literature Review and Related Work

Related work can be broadly classified into two categories: one is techniques for plant disease detection, and second is datasets advancing research in plant disease detection.

Researchers have used CNN architecture and its various versions for the classification and identification of plant diseases. Sunayana et al. [6] compared different CNN architectures for disease detection in potato and mango leaves and achieved 98.33% accuracy with AlexNet and 90.85% accuracy with a shallow CNN model. Guan et al. [27] used a pre-trained VGG16 model to predict disease severity in apple plants and achieved a 90.40% accuracy. Jihen et al. [10] used the LeNet model to accurately distinguish healthy and diseased banana leaves, achieving a 99.72% accuracy. Manpreet et al. [5] classified seven tomato diseases with an accuracy of 98.8% using a pre-trained CNN-based architecture known as Residual Network or generally known as ResNet. Rahman et al. [15] proposed a fully-connected deep learning-based network to distinguish Bacterial Spot, Late Blight, and Septorial Spot disease from tomato leaf images with a 99.25% accuracy. Previous work by Sankaran et al. [22] proposed the use of reliable sensors to monitor health and diseases in plants under field conditions. However, this wasn't a helpful idea due to the hardware cost and lack of expertise to operate such sensors. In Patil et al. [24] study, they extracted shape features for disease detection in sugarcane leaves, which obtained a final mean accuracy of 98.60%. In a similar work, Patil et al. [20] used texture features, i.e., inertia, homogeneity, and correlation obtained by calculating the gray level co-occurrence matrix on the image and color extraction for disease detection in maize leaves. Recent work [23] has investigated neural networks to identify morphological patterns in leaf veins. Similarly, feature extraction and ensemble neural network (NNE) have been used to diagnose tea leaf diseases with a final test accuracy of 91% [12].

3. Methodology

3.1. Datasets

In this project, three different plant disease datasets were selected to evaluate the model's performance. The statistical details of the datasets are present in the table below. The lack of availability of sufficiently large-scale non-lab datasets remains a major challenge. Datasets were selected based on different number of classes and the variety of environments they were collected. Figure.1 elaborates a sample of the three selected datasets which gave a variety of results during the training and testing phase.

Dataset Name	No. of Plants	Total Images	No. of Classes	Avg Images /Class
PlantVillage	14	87000	38	2300
PlantDoc	13	2596	27	85
PlantaeK	8	2136	16	9-313

PlantVillage dataset [2] is one of the most common publicly available datasets which are used by most researchers. The PlantVillage dataset has the highest number of images, and those images are taken in a controlled lab environment. This implies all the images are kept inside the lab and captured under good lighting conditions.

PlantDoc dataset [4] is a real-world crop image dataset, and it is prepared by Indian institutes. The images are captured in natural lighting conditions in the fields. The dataset has fewer images as compared to PlantVillage.

PlantaeK dataset [3] is taken from native plants of the Kashmir region of India. The dataset is imbalanced and also has fewer images. This dataset was captured in broad daylight under auto mode with the Nikon D-SLR digital camera. Those images are captured by an 18-55 mm lens and are in JPG format.

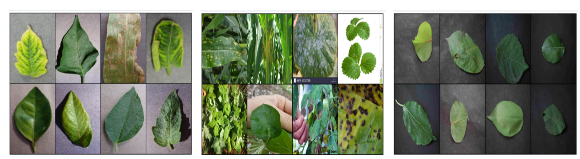

Figure 1. Raw Images for all Datasets

As a preprocessing step, normalization values were calculated for all the datasets. A separate utility script was used to compute the average of the mean and standard deviation of all the images in the dataset. After using these computed values, better results were obtained. Further, online data augmentation techniques were performed to increase the variety of the images in the dataset. Augmentation included resizing, RandomHorizontalFlip, and RandomRotation.

The PlantDoc and Plantaek gave poor training results before preprocessing. For PlantDoc, overfitting and underfitting issues were faced. This was because several images were not of good quality with overlapping text as shown in Figure 2. Data cleaning was performed by deleting such images. To increase the size and variety of images for training the model Random Noise, Horizontal Flip, and Random Rotation were implemented as part of offline data augmentation. This helped in decreasing overfitting and underfitting issues.

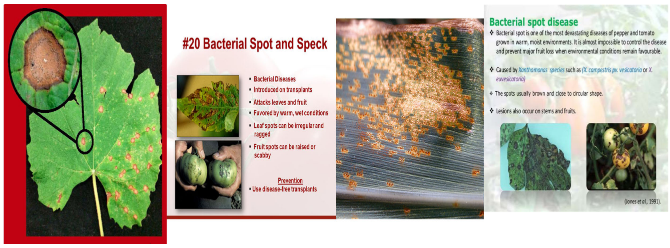

Figure 2. Sample Images of PlantDoc Dataset

Plantaek had class imbalance as shown in Figure 3 and the size of the dataset was 4.3 GB. To tackle this, image compression of 70% and image resize of 1/4th was done using Python Imaging Library while ensuring image quality was not affected. In order to tackle class imbalance Random Noise, Horizontal Flip, and Random Rotation were implemented as part of offline data augmentation. After data augmentation, it gave a good distribution of images per class as shown in Figure 4, and better accuracy.

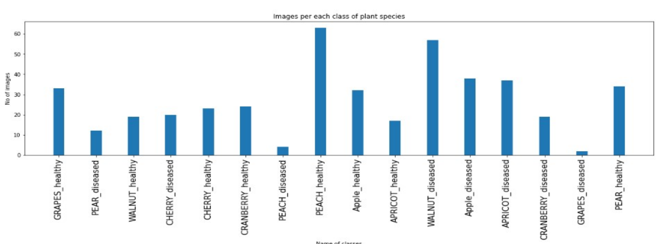

Figure 3. Class Distribution for Plantaek - Before Augmentation

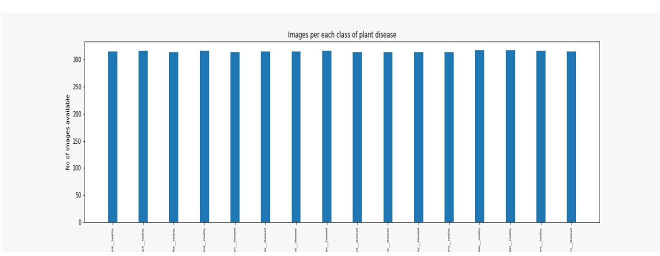

Figure 4. Class Distribution for Plantaek - After Augmentation

The datasets were divided randomly into training, validation, and testing sets in the ratio of 70:20:10. The validation set was used to validate the performance of the model during training, which helped in tuning the hyperparameters of the models accordingly.

3.2. CNN Models

CNN models used in this project were ResNet50, AlexNet, and VGG16. The approaches used for training the models are training from Scratch and Transfer Learning. The selected CNN models are recommended by many researchers [21] for image classification and are also appropriate for smaller dataset sizes. The models also differ in depth and number of FLOPs. CNN models like MobileNet, GoogleNet, and DenseNet are very dense and require larger datasets. Therefore, they are not appropriate for plant disease classification problems.

The Residual network (ResNet) [19] benefits solving complicated tasks and also increases detection accuracy. ResNet tries to solve the difficulties in the training process of deep CNN, the saturation, and the degradation of accuracy. ResNet50 is a residual network with 50 layers (48 convolutional layers, one MaxPool layer, and one average pool layer). It has over 23 million trainable parameters and it is a deeper model and therefore it was able to learn features better.

VGG16 [11] is a 16 layers network with 13 convolutional layers, 5Max Pooling layers, and 3 Dense layers. It has 138M learnable parameters.

AlexNet is a deep neural network created in 2012 by Alex Krizhevsky and it is an 8 layers network with 5 convolutional layers, 3 max-pooling layers, 2 normalization layers, 2 fully connected layers, and 1 softmax layer. AlexNet [9] has about 660K units, 61M parameters, and over 600M connections.

Training a model from scratch requires a huge dataset and a lot of time to train. A pre-trained model is preferred to overcome the plant disease dataset's limitations of smaller dataset size. Transfer learning is a method in which pre-trained models are reused for a new task instead of developing models from scratch. It trains the neural network that has already been trained on different datasets. These models are trained on a large number of images and can improve the accuracy of prediction. In this project, transfer learning using deep tuning is implemented, because it gives higher accuracy and takes less training time. In deep tuning, weights are frozen for the entire network except for the final fully connected layer. The last fully connected layer is replaced with a new one with random weights and only the fully connected layer is trained.

ResNet50 has a lower number of learnable parameters and FLOPs than VGG16 though it is deeper. AlexNet is the least deep network out of the three selected CNN models. However, the training time of ResNet50 was much lesser than that of VGG16. VGG16 took almost double the time than ResNet50 per epoch and three times the training time of AlexNet per epoch. For AlexNet number of

FLOPs is 7.25x10^8, for VGG16 it is 15.5x10^9 FLOPs [17] and ResNet50 3.80x10^9 FLOPs [13]. Moreover, ResNet50 gave higher accuracy in almost all datasets.

3.3. Optimization Algorithm

For this project, Adam optimizer and Cross Entropy Loss function were used. The learning rate was set to 0.001 and a batch size of 32 was used. The training was done for 40 epochs. These values helped the models reach convergence faster, with fewer oscillations.

For model validation, plots for accuracy and loss curves were used. To compare the performance of the models, Accuracy, Precision, Sensitivity Recall, and F1 score parameters were used. Accuracy [26] is the ratio of (TP+TN) / TP+TN+FP+FN. Intuitively, accuracy gives us the percentage of instances that were correctly classified as in a certain class C along with images correctly classified as not in C. Precision is the ratio of TP / (TP + FP). Precision is intuitively the percentage of instances that were correctly classified under a certain class C. Sensitivity Recall is the ratio [25] of TP / (TP + FN). A recall is intuitively the percentage of instances of a certain class C that were classified correctly. F1 Score can be interpreted as a harmonic mean of the precision and recall, where a F1 score reaches its best value at 1 and worst score at 0. F1 Score = 2 * ((Precision * Recall) / (Precision + Recall)).

To visually analyze the performance of all the model KPIs, confusion matrices were generated. This helped in understanding how good the model is at differentiating instances between different classes. For visualization, t-SNE was implemented. t-SNE helps to visualize the distribution of instances as clusters of classes and see how good the model is, in spacially separating the dataset.

4. Results

4.1. Experiment Setup

All the deep learning models were trained on Kaggle P1000 GPU's. Kaggle was selected as it provides 30 hours of GPU per week and the dataset management on Kaggle is much better as compared to other platforms like Google Collab. In order to maintain a uniform comparison, templatized notebooks were created, which help maintain consistency between the results of each model and dataset.

The datasets were preprocessed and then divided into training, validation, and test sets as described in Section 3.1. The selected learning rate, batch size, epoch count, optimizer, and loss function for all models are discussed in Section 3.3.

The training and validation sets were fed into CNN models as mentioned in Section 3.2. Accuracy and loss plots were generated for both training and validation sets to compare the performances. The test results were then compared and visualized using various KPIs as discussed in Section 3.3. t-SNE was also implemented to visualize the distribution of instances as clusters of classes to understand how well the model segregates the dataset into classes.

The hyperparameter tuning selection was taken as part of hyperparameter optimization. These values helped the models reach convergence faster, with fewer oscillations.

4.2. Experiment Results and Analysis

PlantVillage Dataset - For both learning from scratch and transfer learning, the results of ResNet50 were higher than other two architectures (Fig. 12). No underfitting or overfitting was observed. Transfer learning performed good for VGG16 as well. (Fig. 5) The training time for VGG16 was 9.17 hours, AlexNet 3.5 hours, and ResNet50 5.3 hours.

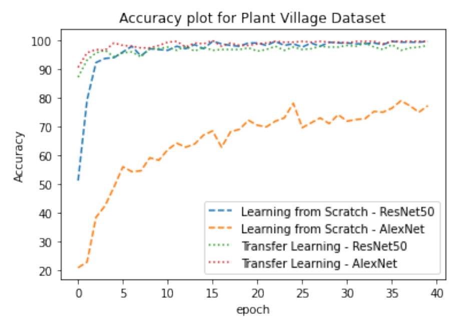

Figure 5. Accuracy comparison between the learning techniques

PlantaeK Dataset – This dataset had a huge class imbalance. With the original dataset, very low accuracy was obtained. After performing data augmentation and balancing the classes both ResNet50 and VGG16 performed well. In training from scratch, overfitting was observed with VGG16 (Fig. 13). AlexNet gave inferior results mostly due to the larger filter kernels. (Fig. 6) The results of ResNet50 were similar in training from scratch and transfer learning while VGG16 performed a bit better in transfer learning.

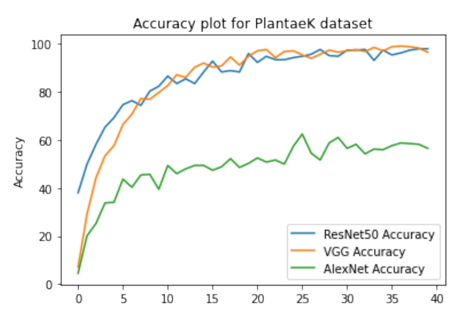

Figure 6. Accuracy comparison for Plantaek Dataset

PlantDoc Dataset – PlantDoc was a very poor-quality dataset. Even after performing preprocessing, the results were not good with training from scratch. ResNet50 performed better than AlexNet but VGG16 was very poor at differentiating the features of the dataset and gave 3% accuracy. Overfitting was observed with ResNet50. With transfer learning, ResNet50 and VGG16 gave better results than training from scratch.

For all the datasets, the best results were obtained with ResNet50, followed by VGG16 and then AlexNet. As corroborated by the literature survey, the performance generally increases from AlexNet to VGG16 to ResNet50 [7]. This is because of the FLOPs calculated and the number of learnable parameters to be computed. With observation, better accuracy was obtained in deeper architectures.

AlexNet gave lower accuracy as it has a bigger kernel size and the depth of this model is lesser than others, hence it struggles to learn features from the training dataset.

The training time of ResNet50 was much lesser than VGG16, even though it has a deeper architecture. This was because VGG16 has more learnable parameters and FLOPs. It was also observed that AlexNet needs more time to achieve higher accuracy compared to its successors.

One point to be noted is, training VGG16 on the PlantDoc dataset was a failed attempt in training from scratch. However, in transfer learning ResNet50 performs better after data argumentation and data cleaning.

Comparing the three datasets, all three models performed better on the PlantVillage dataset as compared to the other two datasets (Fig. 7). This was largely due to the fact that the PlantVillage dataset had considerably large number of input images and more distinguishable features.

Figure 7. Accuracy Comparison for ResNet50 in all datasets

The evaluation metrics for all models of training from scratch and transfer learning are depicted in Table 1 and Table 2 respectively

The confusion matrix (Fig. 8) depicts the actual class vs predicted class by the model. The values in the diagonal are the ones where the predicted class was same as the ground truth and the non diagonal non-zero values are the ones where the model gave the wrong results.

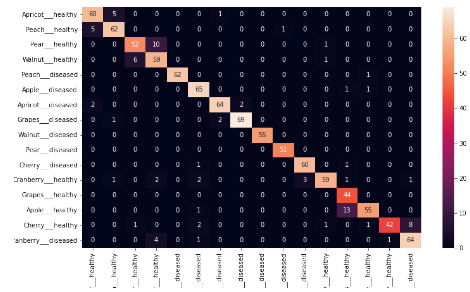

Figure 8. Confusion Matrix for PlantaeK Dataset for ResNet50

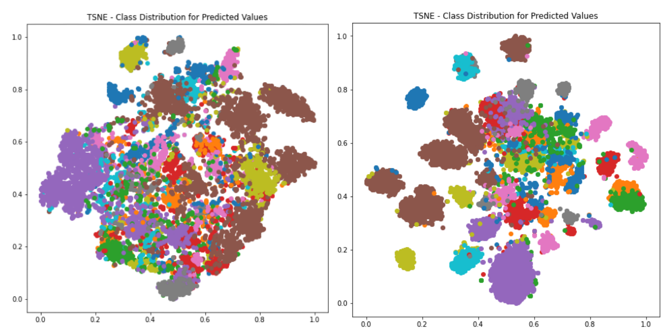

Figure 9. t-SNE plots comparing Learning from scratch (L) and Transfer Learning(R)

The t-SNE plot shows the data separability into clusters of classes.

Comparison among datasets: For instance, in ResNet50, the t-SNE plot has a cohort data pattern for PlantVillage because the model performed well while it showed a scattered pattern for PlantDoc as the model has poor performance due to low-quality images which made it difficult to distinguish the images. (Fig. 14)

Comparison among models: For example PlantaeK t-SNE plots show that ResNet50 performed better than other models as it has a consistent density of distribution. (Fig. 15)

Comparison with and without transfer learning: For example, in the PlantVillage AlexNet t-SNE plot, the class distribution is more distinguishable for transfer learning which shows it was more accurate. (Fig. 9)

4.3. Hyperparameter Optimization

Hyperparameter tuning was done with learning rate: 0.001, 0.0001, 0.00001, and batch size: 8, 16, 32, 64. Tuning was done on the PlantVillage dataset for the ResNet50 to check the variance of accuracy.

Model Metrics	PlantVillage			PlantDoc			PlantaeK		
	ResNet50	**VGG16**	**AlexNet**	**ResNet50**	**VGG16**	**AlexNet**	**ResNet50**	**VGG16**	**AlexNet**
Test Accuracy	97.57	93.28	75.20	61.65	2.59	19.92	91.57	83.04	58.63
Precision	97.67	93.45	74.38	64.03	0.07	23.27	92.12	83.78	60.61
Sensitivity Recall	97.57	93.28	75.20	61.65	2.59	19.92	91.57	83.04	58.63
F1 Score	97.56	93.29	73.69	62.01	0.13	16.98	91.58	83.08	57.81

Table 1. Test Metrics for Training from Scratch

Model Metrics	PlantVillage		PlantDoc		PlantaeK	
	ResNet50	**AlexNet**	**ResNet50**	**VGG16**	**ResNet50**	**VGG16**
Test Accuracy	97.02	95.95	77.99	71.81	91.96	91.17
Precision	97.22	96.02	79.11	73.69	92.37	91.27
Sensitivity Recall	97.02	95.95	77.99	71.81	91.96	91.17
F1 score	97.01	95.91	77.74	72.16	91.96	91.17

Table 2. Test Metrics for Transfer Learning

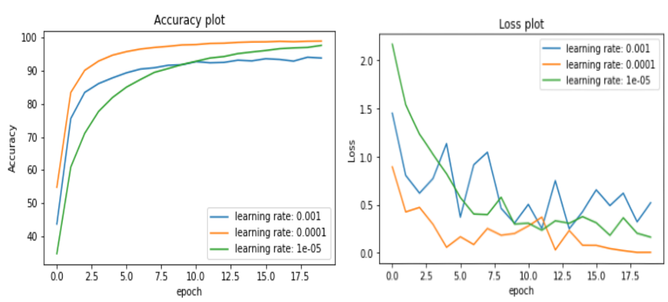

Figure 10. Accuracy and Loss by varying learning rates

	Learning rate (Batch size = 32)		
	0.001	0.0001	0.00001
Test Accuracy	90.05	96.77	94.10
	Batch Size (Learning rate = 0.001)		
	8	16	32
Test Accuracy	75.37	85.26	87.36

Table 3. Test Accuracy by Varying Learning Rate and Batch size

The rationale behind the learning rate chosen is that most of the image classification problems use LR in the range 0.1 to 10^{-7} [1]. Similarly, as per research, the ideal batch size should be 16-64. Analysis of model training performance was done with smaller batch sizes.

Impact of varying learning rate: Learning rate was varied by keeping batch size constant at 32. Best accuracy was observed with LR = 0.0001. LR = 0.01 performed better than LR = 0.00001 for the initial 9 epochs. However, after the 9^{th} epoch, the performance of LR = 0.00001 surpassed (Fig. 10). The oscillating curve of validation loss (Fig. 10) for LR = 0.001 clearly indicates that the model is not able to converge and is overshooting the minima. That is why after 9 epochs model with LR = 0.00001 performs better than the model with LR = 0.001. However, the training time was not affected by changing the learning rate.

The results of test accuracy are summarized in Table 3.

Impact of varying batch size: The batch size was varied by keeping the learning rate constant at 0.001. Higher the batch size better the accuracy (Fig. 11) and lower the training time. The downside of using a smaller batch size is that the model is not guaranteed to converge to the global optima. It will oscillate around the global optima as evident from the zigzag loss curves.

Training time for batch sizes 8, 16, 32, and 64 were respectively 75, 56, 51, and 48 minutes. This implies that batch size of 8 took 1.5 times more time than batch size 64.

The results of test accuracy are summarized in Table 3.

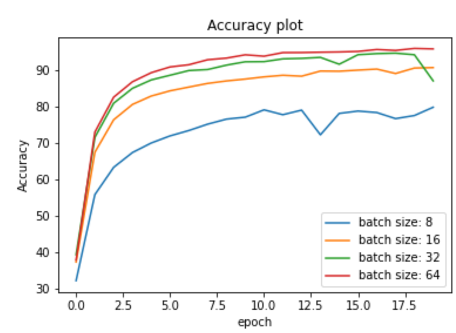

Figure 11. Training Accuracy by varying Batch size

To summarize, models performed the best with the combination of LR = 0.0001 with batch size 32 and the combination of batch size = 64 with LR = 0.001

4.4. Conclusion

The best results were given by ResNet50 on all 3 Datasets. Transfer Learning gave better performance than training from scratch. PlantVillage dataset provided the best results. This was largely due to the fact that the PlantVillage dataset had considerably more input images and more distinguishable features.

References

[1] The frobnicatable foo filter. 6

[2] Plant village dataset. https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset. 2

[3] Plantaek. https://data.mendeley.com/datasets/t6j2h22jpx/2. 2

[4] Plantdoc. https://github.com/pratikkayal/PlantDoc-Dataset. 2

[5] B.; Algergawy A Amara, J.; Bouaziz. A deep learning-based approach for banana leaf diseases classification. In *In Proceedings of the Datenbanksysteme fur Business, Technologie und Web (BTW 2017)—Workshopband*, pages 6–10, 2017. 2

[6] R Arya, S.; Singh. A comparative study of cnn and alexnet for detection of disease in potato and mango leaf. in proceedings of the 2019 international conference on issues and challenges in intelligent computing techniques (icict). pages 1–6, 2019. 2

[7] Viktar Atliha and Dmitrij Šešok. Comparison of vgg and resnet used as encoders for image captioning. In *2020 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)*, pages 1–4, 2020. 5

[8] Asher Levi Victor Alchanatis Bar Cohen, Yael Edan. Early detection of grapevine (vitis vinifera) downy mildew (peronospora) and diurnal variations using thermal imaging. In *Sensors*, pages 1–16, 2022. 1

[9] Hsing-Chung Chen, Agung Mulyo Widodo, Andika Wisnujati, Mosiur Rahaman, Jerry Chun-Wei Lin, Liukui Chen, and Chien-Erh Weng. Alexnet convolutional neural network for disease detection and classification of tomato leaf. *Electronics*, 11(6):951, 2022. 3

[10] Sabella E. Monteduro AG Chiriac'o MS De Bellis L Luvisi A Maruc cio G. Buja I. Advances in plant disease detection and monitoring: From traditional assays to in-field diagnostics. In *Sensors*, pages 1–22, 2021. 1, 2

[11] Carlos A Mamani Diaz, Edgar E Medina Castaneda, and Carlos A Mugruza Vassallo. Deep learning for plant classification in precision agriculture. In *2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, pages 9–13. IEEE, 2019. 3

[12] Guillermo L Grinblat, Lucas C Uzal, Mónica G Larese, and Pablo M Granitto. Deep learning for plant identification using vein morphological patterns. In *Computers and Electronics in Agriculture*, pages 418–424, 2016. 2

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 4

[14] Bo Li Bo Ma Mingyang Wu Kamal KC, Zhendong Yin y. Transfer learning for fine-grained crop disease classification based on leaf images. In *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–5, 2019. 1

[15] R Kaur, M.; Bhatia. Development of an improved tomato leaf disease detection and classification method. In *In Proceedings of the IEEE Conference on Information and Communication Technology*, pages 1–5, 2019. 2

[16] S. Zhang L. Li and B. Wang. Plant disease detection and classification by deep learning. In *Access 2021*, pages 56683–5669, 2022. 1

[17] Wei Li, Zhu, Xiatian, and Shaogang Gong. Harmonious attention network for person re-identification. 02 2018. 4

[18] Guiping Hu Luning Bi. Improving image-based plant disease classification with generative adversarial network under limited training set. *Frontiers in plant science 11*, 2020. 1

[19] Ishrat Zahan Mukti and Dipayan Biswas. Transfer learning based plant diseases detection using resnet50. In *2019 4th International conference on electrical information and communication technology (EICT)*, pages 1–6. IEEE, 2019. 3

[20] Sanjay B Patil and Shrikant K Bodhe. Leaf disease severity measurement using image processing. In *International Journal of Engineering and Technology*, pages 297–301, 2011. 2

[21] Tan Nhat Pham, Ly Van Tran, and Son Vu Truong Dao. Early disease classification of mango leaves using feed-forward neural network and hybrid metaheuristic feature selection. *IEEE Access*, 8:189960–189973, 2020. 3

[22] M.M.; Mahdee G.S.; Kabir M.W.U Rahman, M.A.; Islam. Improved segmentation approach for plant disease detection. In *Engineering and Robotics Technology (ICASERT)*, pages 1–5, 2019. 1, 2

[23] K Shrikant Sanjay B Patil and Bodhe. Betel leaf area measurement using image processing. In *International Journal on Computer Science and Engineering (IJCSE)*, 2011. 1, 2

[24] Reza Ehsani Sindhuja Sankaran, Ashish Mishra and Cristina Davis. A review of advanced techniques for detecting plant diseases. In *Computers and Electronics in Agriculture*, pages 1–13, 2010. 1, 2

[25] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4):427–437, 2009. 4

[26] Alaa Tharwat. Classification assessment methods. *Applied Computing and Informatics*, 2020. 4

[27] Y.; Wang J Wang, G.; Sun. Automatic image-based plant disease severity estimation using deep learning. comput. intell. pages 1–22, 2021. 2

5. Appendix

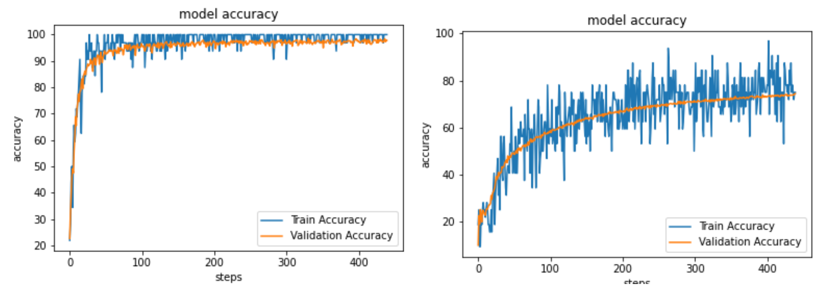

Figure 12. Accuracy plots for ResNet50 and AlexNet

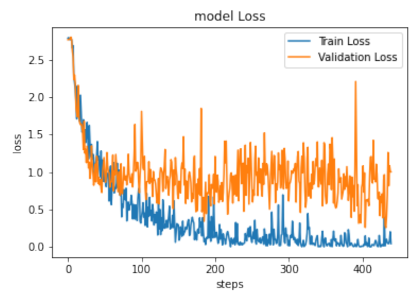

Figure 13. Loss values for VGG in Plantaek Dataset

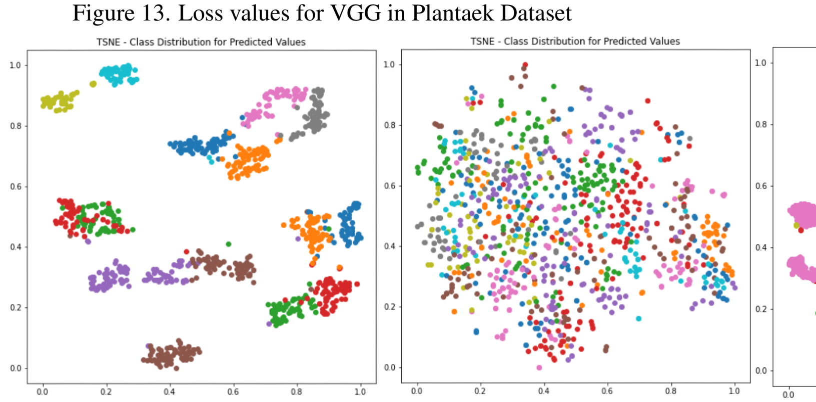

Figure 14. Comparisons of Datasets on ResNet50. Plantaek(L), PlantDoc(M), PlantVillage(R)

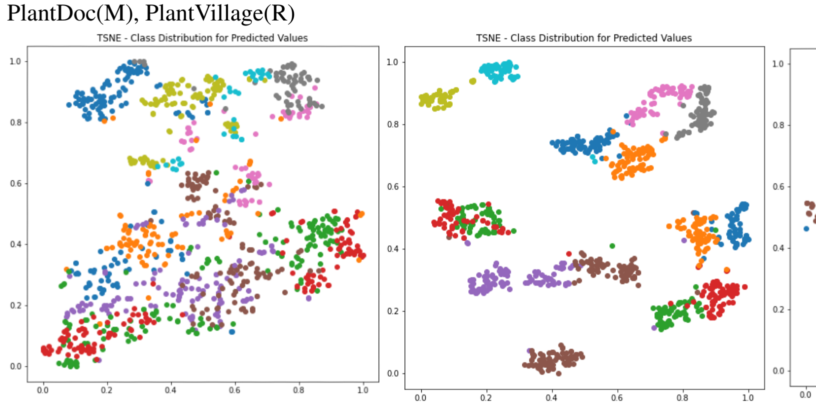

Figure 15. Comparisons of Models on the Plantaek Dataset. AlexNet(L), ResNet50(M), VGG(R)