PLANT LEAF DISEASE DETECTION USING DEEP LEARNING

Suhail Saifi

*Department of Computer Science and Engineering,*

*School of Engineering & Technology,*

*Sharda University*

Greater Noida, India

Bal Mukund

*Department of Computer Science and Engineering,*

*School of Engineering & Technology,*

*Sharda University*

Greater Noida, India

Danish Alam

*Department of Computer Science and Engineering,*

*School of Engineering & Technology,*

*Sharda University*

Greater Noida, India

Dr. Ruqaiya Khanam

*Department of Computer Science and Engineering,*

*School of Engineering & Technology,*

*Sharda University*

Greater Noida, India

***Abstract*— Identification of plant disease plays an important role as it prevents stunted growth which causes bad effects on yields. As agriculture plays a vital role in the Indian economy and different countries of the world, so there is a need to prevent losses in terms of production, quality, and quantity of agriculture yields due to plant disease. Earlier farmers used to monitor plant disease with the naked eye which was time-consuming and required a lot of expertise such as being able to identify a disease and disease-causing agent. But nowadays with advancements in technology, smart farming, and automatic techniques, plant disease can be easily identified and proper diagnosis can be done. It reduces a lot of work of monitoring such as in the case of big farms. Also at an early stage, it detects the symptoms of plant disease when they first appear on leaves. This paper reflects the potential of one such method-Plant disease detection using deep learning, using which one can detect plant disease. It includes the use of image processing techniques with the help of a deep learning algorithm to get a clear and defined image or to extract some useful insight from it. Also, CNN is used for image classification. CNNs are equipped with input, output, and hidden layers which help in process and in image classification.**

***Keywords —***  **Disease detection, Image processing, K-means clustering**

I. Introduction

Plant diseases have been shown to reduce food intake by humans by disrupting crop yields. This may result in malnutrition or starvation and death in extreme cases. Agriculture performs an essential position in the Indian economic system. About 70 percent of rural families depend upon agriculture. Agriculture is an essential part of the Indian economic system as it contributes approximately 17% of the entire GDP and offers employment to more than 60% of the population. The direct economic impact of cross-border pests or diseases is that the loss or decline in the efficiency of agricultural production - whether in crops or animals - reduces farmers' incomes. Therefore, automatic diagnostic tests are helpful because they reduce the external monitoring function on large plant farms, and at an early stage the symptoms of disease appear on plant leaves. This paper covers different sections. The first section provides a brief introduction to the importance of diagnosing plant diseases. The second section discusses the recent work done in this area and reviews the strategies used.

Deep Learning (DL) is one of the various subclasses of Machine Learning, and it has so far gone into three different stages of developments. The first generation originated in 1943 and its name was MCP, and was only useful in linear classification problems.

Back Propagation (BP) was the second generation of the neural network. It was invented by Hinton in 1986. Nonlinear classification and mapping were also solved with the help of sigmoid functions, but around 1991 some problems of gradient vanishing were pointed out.

Deep Learning (DL) is the third generation of neural networks. It was introduced in 2006, and as of now many DL models or architectures are being used for image detections or recognition.

II. Related work

Plant disease reduces food intake of humans by disrupting crop yields. In the worst case, it may lead to starvation or malnutrition. So, Identification of plant disease plays an important role as it prevents stunted growth which causes bad effects on yields. It reduces a lot of work of monitoring such as in the case of big farms. Also at an early stage, it detects the symptoms of plant disease when they first appear on leaves. So, We came up with the idea of a plant disease detection system. Our workflow started with the study of surveys and research papers related to this field. From survey papers we came across many techniques and ways to implement this plant disease detection .Some of them are Feed Forward Neurаl Netwоrk (FFNN), Leаrning Veсtоr Quаntizаtiоn (LVQ) аnd Rаdiаl Bаsis Funсtiоn Netwоrks (RBF).,SVM etc.

Through the literature survey we got an ideation of the key shortcomings of the proposed approaches and we tried to include it into our model, by making a system more accurate and much for feasible. We came up with-Plant disease detection using deep learning, using which one can detect plant disease. It includes the use of image processing techniques with the help of a machine learning algorithm to get a clear and defined image or to extract some useful insight from it. Also, CNN is used for image classification. CNNs are equipped with input, output, and hidden layers which help in process and in image classification.

This model helps in automatic diagnostic tests and they reduce the external monitoring function on large plant farms, and at an early stage detects the symptoms of disease appeared on plant leaves. We have used a public dataset of 54,306 images of diseased and healthy plant leaves of different species .Trained a deep convolutional neural network to identify 14 crop species and 26 diseases . The trained model achieves an accuracy of 95.35% for the collected dataset on being tested .Hence shows the feasibility of this approach. Overall, the approach of training deep learning models on increasingly large and publicly available image datasets presents a clear path toward smartphone-assisted crop disease diagnosis on a massive global scale.

Plant disease detection with deep learning gives an accuracy of 95.35% -96% and to make this model more friendly we have integrated this model with very easy to use interface. To use this model , farmers just have to drop the plant leaf image and run the model. This model gives the accurate result and inform the farmers about the plant diease and what action should they take.

III. Methodology Used For Analysis

The survey of literature gave us an impactful insight about the importance of considering the complexity of the utilized architectures. Thus, we moved on to the explication of our proposed outlook of Plant Disease Detection encapsulating varied modus operandi. The scrutiny of varied pre-trained models was done based on diverse parameters to inhibit computational complexity. There are certain factors which impacts the tuning of certain algorithms which can be elucidated as [16] :

* High Number of Parameters in a model makes it Sluggish to be operated in real-time.
* High Number of Recurring Units, i.e., training over the same structure again and again. After a certain instance, the models perform extremely poor and causes overfitting.
* Utilization of Activation Functions which are complicated deteriorates the performance of the model.
* Deeper the Network more liable it is for causing Overfitting and thereby hindering the modus operandi.

As a result, it becomes prominent to consider the most feasible model for the applicability with greater authenticity. Thus, we collated the most famous Light-weight and Heavy-weight pre-trained models to scrutinize the minutes of the impact of depth of the models, number of parameters, and size of the model over a certain problem statement. We included MobileNetV2 and DenseNet121 as Light-weight Systems and InceptionResNetV2 and VGG16 as Heavy-weight Systems. The basis of selection of these models is elucidated in Table-I.

| **S.No.** | **Pre-Trained Architectures** | **Model Size**  **(In MB)** | **No. of Parameters**  **(approx.**  **in Millions)** | **No. of Layers** |
| --- | --- | --- | --- | --- |
| 1. | MobileNetV2 | 14 | 3.5 | 88 |
| 2. | DenseNet121 | 33 | 8 | 121 |
| 3. | InceptionResNetV2 | 215 | 55.8 | 572 |
| 4. | VGG 16 | 528 | 138 | 23 |

Now, as the model selection was carried out through explication over varied references, we tried to understand the architecture of each model. As a result, Figure.1, Figure.2, Figure.3, and Figure.4 demonstrate the graphical illustration of specifics of the model and the minutes of layers in those models.

*A. MobileNetV2:*

As per the Table-I, MobileNetV2 consists of 88 Layers with 3.5 Million Parameters overall and the size of the model is just 14MB. Thus, it has been considered as the Light-Weight Model when compared to other pre-trained models.

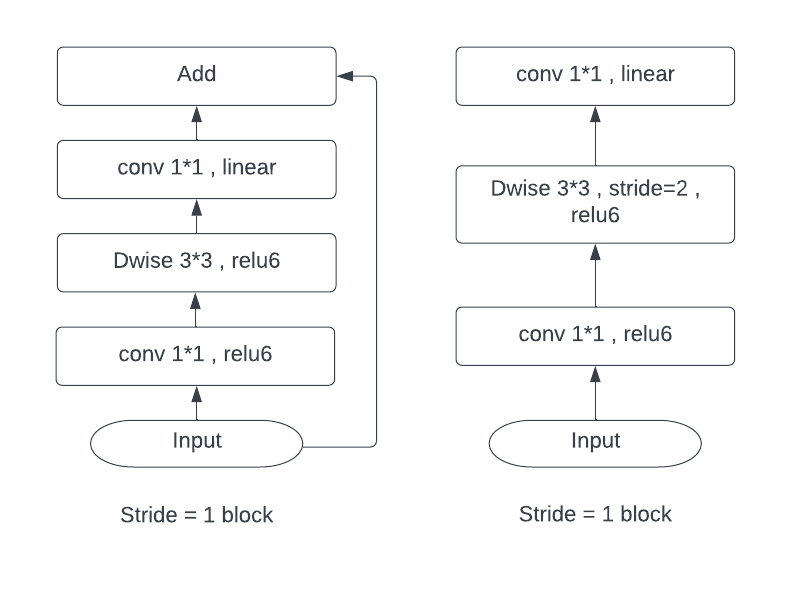


Fig .1. Demonstration of MobileNetV2 Architecture with Residual and Downsizing Block [16].

*B. DenseNet121:*

As per the Table-I, DenseNet121 consists of 121 Layers overall, along with 08 Million Parameters colligating a size of 33MB, making it also a Light-Weight System when compared to other pre-trained models.

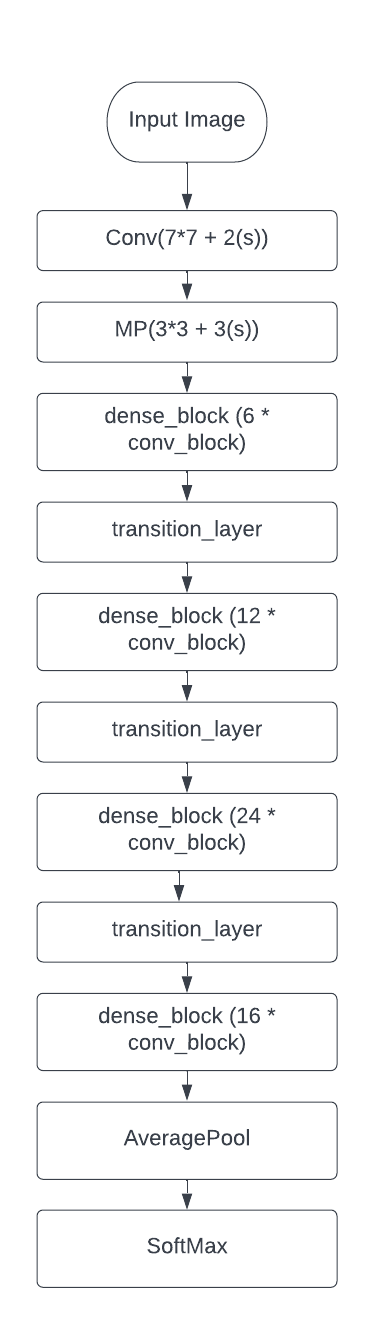


Fig .2. Demonstration of DenseNet121 Architecture [16].

*C. InceptionResNetV2:*

As per the Table-I, InceptionResNetV2 is the collation of Inception and ResNet Family of Pre-Trained architectures consisting of 572 Layers with 55.8 Parameters. The Size of the Model is 215MB making it a Heavy-Weight System when compared with other pre-trained models.

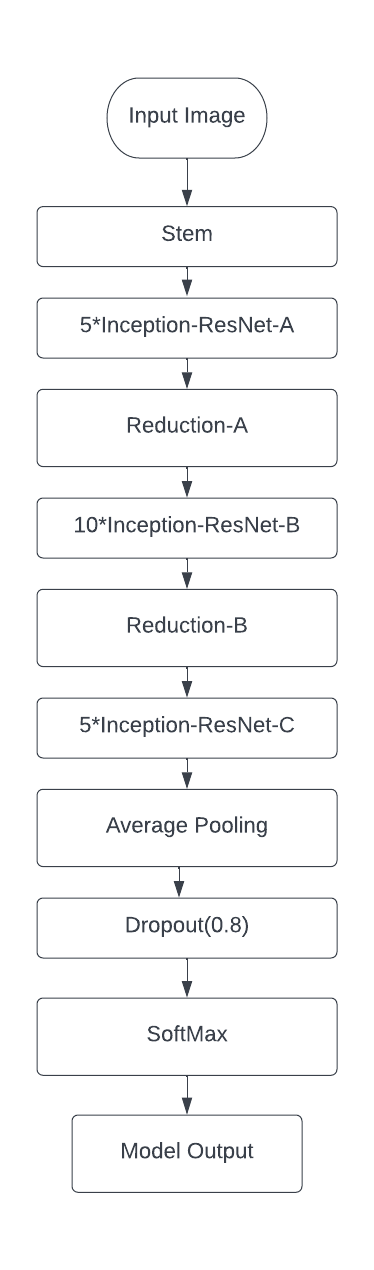
**

Fig .3. Demonstration of InceptionResNetV2 Architecture [16].

*D. VGG16:* As per the Table-I, VGG16 consists of 23 Layers with 138 Million Parameters. The size of the model stands to be 528MB, making it a Heavy-Weight System when compared to other pre-trained models.

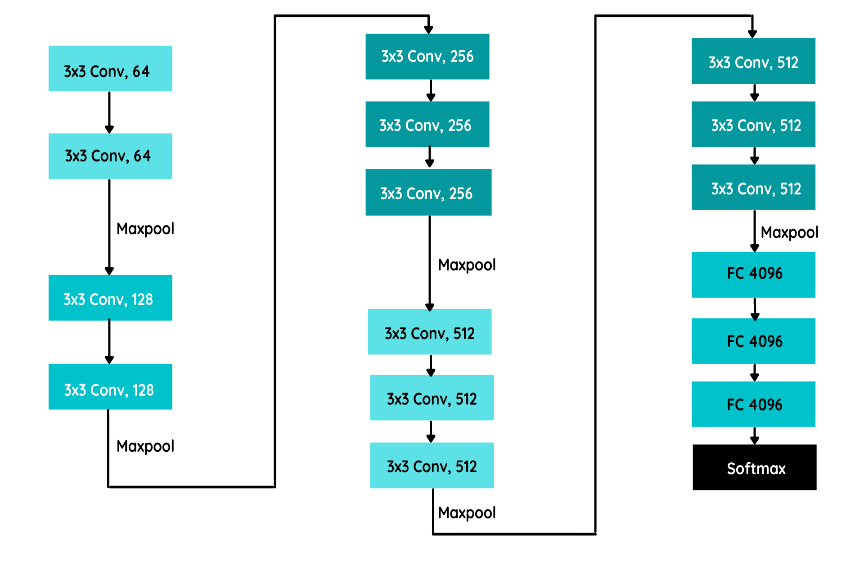
**

Fig .4. Demonstration of VGG16 Architecture [16].

These architectures stand out to be prominent for our modus operandi and aids us to understand the behavior of models for Multi-Class Dataset and thereby creating a real-time system for Plant Disease Detection.

IV. Implementation and Tools

The implementation flow for the System Creation has been elucidated below in an explicatory fashion.

*A. Dataset Utilized:* As we wanted to formulate a real-time Plant Disease Detection System, to infer between the efficacy of pre-trained models over Multi-Class Dataset we collated Plant Leaf Dataset from Kaggle which is open to use. The Dataset consists of 15 Classes with a diverse variety of Plant Leaves elucidated in Figure.5. There are 20935 instances of Data for scrutinizing the model’s performance. The illustration of Dataset is shown in Figure.5.

[](https://lucid.app/documents/edit/979d1151-b94c-473b-a703-b2ccea22a11d/0?callback=close&name=docs&callback_type=back&v=351&s=612)

Fig .5. Demonstration of Plant Leaves Dataset.

*B. Frameworks Utilized:* TensorFlow and Keras have been induced for working upon Deep Learning Architectures. OpenCv is used for reading and processing images. Numpy for saving the images in array format. Sklearn for one hot encoding,training and testing split, calculating the performance of the model. Matplotlib and Seaborn for plotting graphs in a much more advanced way. Gradio, an open-source python library for creating web-apps has also been utilized for creating a web app for Plant Disease Detection.

*C. Accuracy Parameters:* As the Data has been split into Train, and Test set. The Training Accuracy, and Test Accuracy has been used for explaining the architecture's potency over the particular dataset. Also, the Confusion Matrix is used for scrutinizing the model’s efficacy in real-time over different labels/classes.

*D. Workflow:* The dataset that we have taken was having 15 categories of plants. The dataset was very unbalanced, due to which we got many false positive predictions. So in order to handle the unbalanced dataset there are three ways,

1. Make an equal number of samples in all the categories by reducing the samples.
2. Generate more samples in those categories of plants which were having less number of samples.
3. Make use of class weights.

So we used the class weight method to balance the dataset. Working steps of class weight method - the class weight doesn’t reduce or generate samples to make the dataset balanced, instead it makes the dataset balanced while training by providing an array, defining how much weight to be given to each category or class, such that the model trains or learns that the dataset is in equal distribution.

Figure.6, demonstrates the whole modus operandi of our proposed outlook.

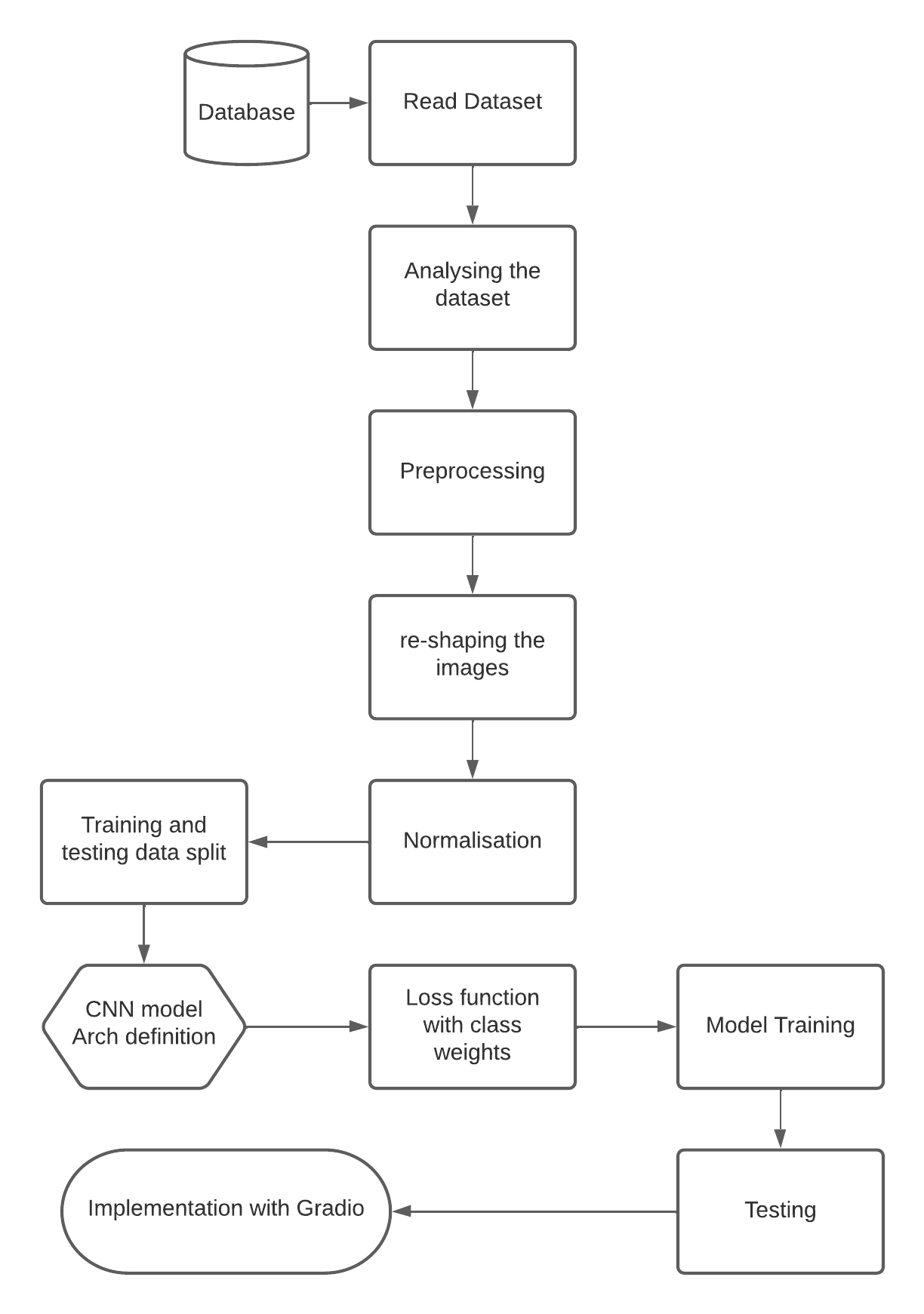
[](https://lucid.app/documents/edit/1b569ebd-b9df-46cc-ac44-ba6de5589bb9/0?callback=close&name=docs&callback_type=back&v=1071&s=612)

Fig .6. Demonstration of the Implementation Flow [16].

V. Experimental Results and Analysis

After the Implementation of the proposed outlook, we got phenomenal results for the model that we have selected for building the Plant Disease Detection System which was VGG16 and a worthy insight was procured through it. Figure.7 and Figure.8 represent the graphical perspective of the model in terms of Accuracy and Loss.

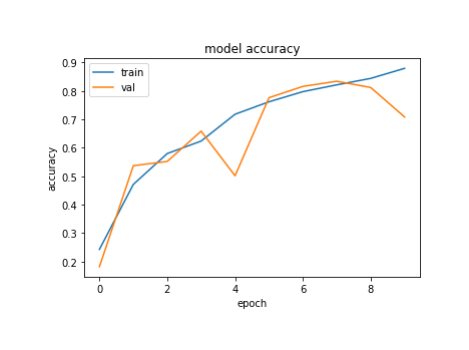


Fig .7. Accuracy Results for VGG16.

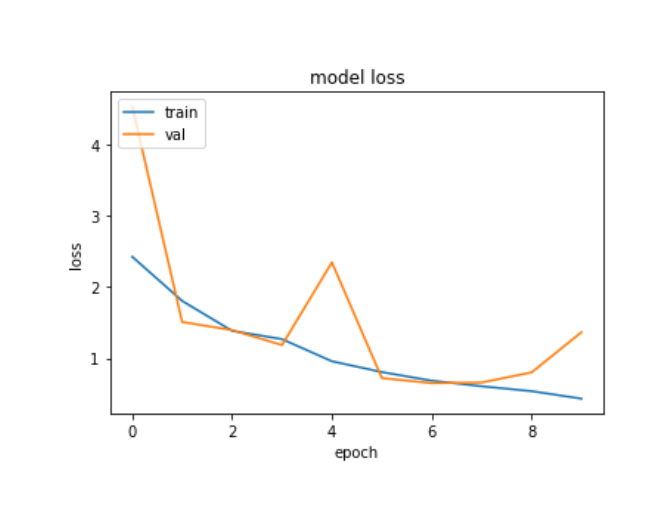
[](https://lucid.app/documents/edit/dde1ba72-e5a4-478a-bce2-e0bf1686be4b/0?callback=close&name=docs&callback_type=back&v=120&s=336)

Fig .8. Loss Results for VGG16.

Also, Figure.9, demonstrates the Confusion Matrix for architecture.

.

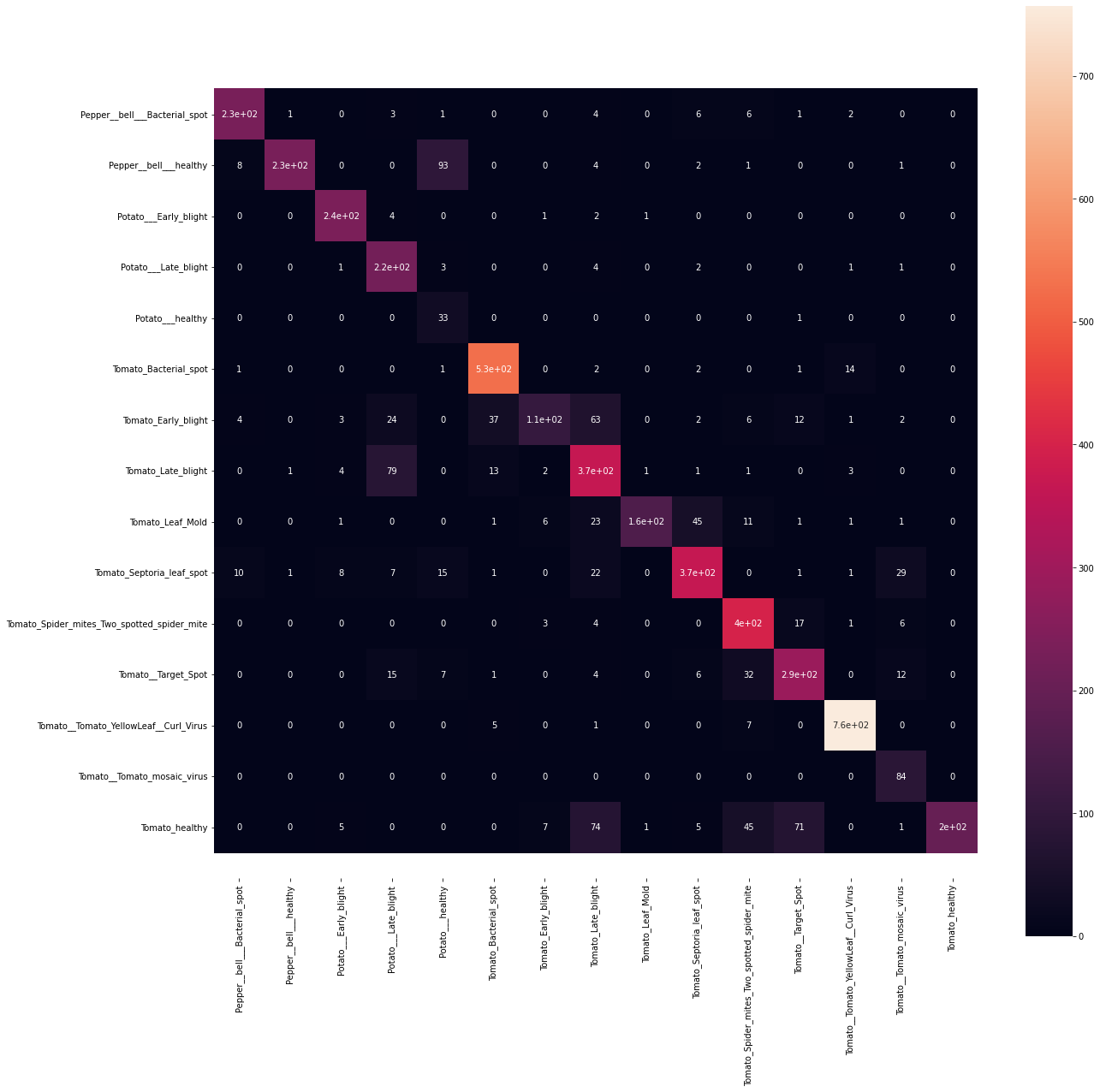
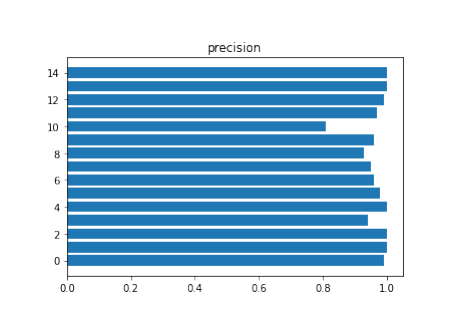


Fig .9. Confusion Matrix for VGG16.

As we have used VGG architecture for training the model and from that we achieved 97% of accuracy, and made the calculation of precision, recall, F1-score for each class. Figure.10, Figure.11, and Figure.12 demonstrate the precision ,recall and F1-score for each class.



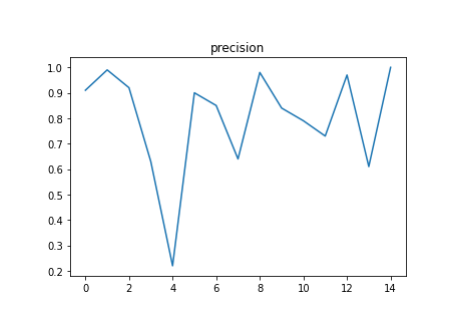
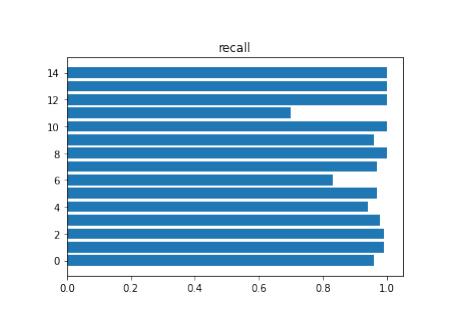
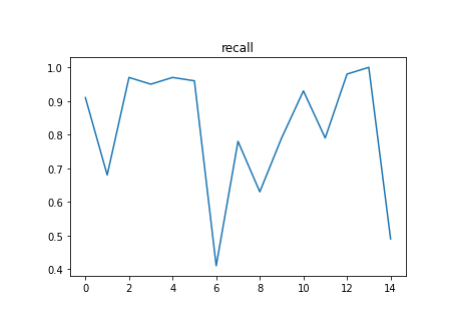
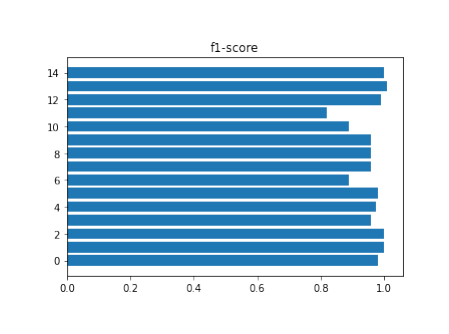


Fig .10. Precision



Fig .11. Recall



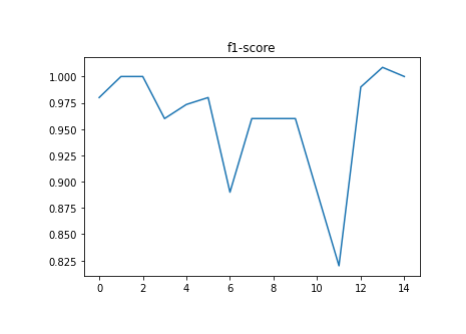


Fig .12. F1-score

Table-II, illustrates the Accuracies obtained for Training, Validation and Test Data for the VGG16 model.

TABLE II. Analysis of pre-trained architectures based on their accuracies over Plant Leaf dataset.

| Model | Training Accuracy  (%) | Test Accuracy (%) |
| --- | --- | --- |
| VGG16 | 97.51 | 97.33 |

VI. Conclusion and Future Scope

Plant disease detection has become a very big domain to work on, because many of the works have been done earlier using many different technologies and methods in this field. We so far have read some of the papers for our survey and have analyzed and have found many different results for different techniques and methods used. The papers were on the same topic but of different methodologies like few of them were based on MobileNetV2, DenseNet121 InceptionResNetV2 and VGG16 etc. We have analyzed that even after doing so much research in this domain, there is a large scope of research in the future also because every proposed model has some constraints, some disadvantages and some advantages. The need for a perfect model to detect the plant leaf disease is also in hunt and this is the reason that these models are not getting used on a large scale for disease detection. This is a very good and very helpful domain to work on because when farming is done on a very big scale, it is not possible for farmers to look for diseases on every leaf of the plants, so further research and findings on this domain can be very helpful.

So after comparing the results of different models, VGG16 architecture is the best modus operandi as it works better with small data and large data also and after our research we got the accuracy of 99% in some of the present classes, like the category Tomato\_Target\_spot is the least accuracy scored which is 82% the category Pepper\_bell\_healthy, Potota\_b, etc are the highest accuracy scored which is more than 99%

As a result, it’s necessary to scrutinize over varied architectures for finalizing the base model for any modus operandi. The efficacy we got stands out as the best in assuaging this system and more dataset can be amalgamated for greater explication as an outlook.

REFERENCES

1. J. G. A. Barbedo, ‘‘Plant disease identification from individual lesions and spots using deep learning,’’ Biosyst. Eng., Vol. 180, PP 96–107, Apr, 2019
2. Liu, Y. Zhang, D.-J. He, and Y.-X. Li, ‘‘Identification of apple leaf diseases based on deep convolutional neural networks,’’ Symmetry, Vol. 10, No. 1, pp 11–19. Dec, 2018.
3. H. Larochelle, ‘‘Few-shot learning,’’ in Computer Vision: A Reference Guide, 2nd ed., K. Ikeuchi, Ed. Tokyo, and Tokyo: Univ. Tokyo Press, 2014.
4. Shima Naresh, Ramchandra Habbar et al. “Plant Disease Detection Using Machine Learning”, IEEE, 13 Aug 2018.
5. Parul Sharma, YPL Singh et al “KrishiMitr (Farmer’s Friend): Using Machine Learning to Identify Diseases in Plants”, IEEE, 7 Jan 2019.
6. Shital Bankar, Ajita Dubey et al. “Plant Disease Detection Techniques Using Canny Edge Detection & Color Histogram in Image Processing”, International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1165-1168.
7. ZN Reza et al. “Detecting jute plant disease using image processing and machine learning”, IEEE, 9 March 2017.
8. MS Arya, K Anjali et al. “Detection of unhealthy plant leaves using image processing and genetic algorithm with Arduino”, IEEE, 14 June 2018.
9. R T Narmadha et al. “Detection and measurement of paddy leaf disease symptoms using image processing”, IEEE. 23 Nov 2017.
10. M Poobalasubramnium et al. “Plant Disease Identification and Detection Using Support Vector Machines and Artificial Neural Networks”, Artificial Intelligence and Evolutionary Computations in Engineering Systems (pp.15-27). Jan 2020.
11. Tejas Tawde, Kunal Deshmukh et al. “Rice Plant Disease Detection and Classification Techniques, A Survey”, IJERT, volume 10, July 2021.
12. Vittal S Gutte et al. “A Survey on Recognition of Plant Disease with Help of Algorithm” IJESC, Volume 6 Issue No. 6, Juy 2016.
13. Murk Chohan, Adil Khan et al. “Plant Disease Detection using Deep Learning”, IJRTE. ISSN: 2277-3878, Volume-9 Issue-1, May 2020.
14. RU Khan, A Khan et al. “Image-Based Detection of Plant Diseases: From Classical Machine Learning to Deep Learning Journey”, Hindwani, Volume 2021 |Article ID 5541859. June 2021.
15. Yang Juo, Zing Jhang et al. “Plant Disease Identification Based on Deep Learning Algorithm in Smart Farming”, Hindwani, Volume 2020 |Article ID 2479172. Aug 2020.
16. S. A. Dwivedi and A. Attry, "Juxtaposing Deep Learning Models Efficacy for Ocular Disorder Detection of Diabetic Retinopathy for Ophthalmoscopy," 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC), 2021, pp. 352-357, doi: 10.1109/ISPCC53510.2021.9609368.