Foilar Disease Classification in Apple Trees

Abhijeet Manoj Pal (200100107)
Electrical Engineering
IIT BOMBAY
Mumbai, India
200100107@iitb.ac.in

Gnanendar Reddy Palagiri (200070053)
Electrical Engineering
IIT BOMBAY
Mumbai, India
200070053@iitb.ac.in

Kanak Yadav (20d070044)
Electrical Engineering
IIT BOMBAY
Mumbai, India
20d070044@iitb.ac.in

***Abstract*—Apple is the third most widely produced fruits in the world. Various states in India such as Jammu and Kashmir, Himachal Pradesh produce apples. However, there are a lot of diseases such as apple scab, apple rust, and brown rot which affect apple cultivation. It is estimated that apple scab causes nearly 70% yield loss. Various types of pesticides are currently used to prevent such diseases in the apple trees. Disease identification is done using manual labour, completely based on visual appearance of leaves. This method is non-reliable, labour intensive and can lead to misdiagnosis. Misdiagnosis of diseases can lead to misuse of chemicals which can result in formation of resistant pathogen strains. Various computer vision models have been used for detection of foilar diseases in apple trees. In this paper, we implement and compare Deep learning models such as VGG16, Dense Net121, ResNet50. We also compare these models to our standard classification models such as Logistic Regression, SVM and Simple Neural Network. We evaluate these models by comparing prediction accuracy, f1 score and confusion matrix. We find that Dense Net121 gives 89.78% accuracy and 0.889 f1 score which is the highest we obtain compared to other methods implemented.**

***Index Terms*—Apple rust, apple scab, Dense Net121, foilar diseases, logistic regression, neural networks, ResNet50, SVM, VGG16**

I. Introduction

APPLE production takes place in the state of Jammu and Kashmir and Himachal Pradesh in India. These areas are mountainous and accessing these areas on a regular basis is costly and risky to human life. Also a large number of apple trees get affected by foilar diseases such as apple rust, apple scab, and multiple diseases. Therefore, the trees should be inspected properly and periodically for checking their health. The type of disease currently is judged based on human inspection. It is important to use some computerized methods so that we can check the type of disease without having to inspect the trees daily. Cameras can be setup but we still need to find the type of disease. Computer Vision based classification techniques can be used. However, the leaves in both rust and scab disease look similar. Also there is a possibility that the tree is affected by both the diseases. This makes it a challenging task. The best method is to use machine learning techniques which can classify the trees as healthy, affected by rust, affected by scab, and affected by multiple diseases. We can use Deep Learning Models such as VGG16, Deep Net121, ResNet50. We can also use basic classification models such as Logistic Regression, SVM, Neural Network. In this paper we have implemented these models and state the accuracy, f1 score and confusion matrix obtained by each model. We find that DenseNet121 gives the best accuracy for classification of the diseases and can be used practically to find the diseases in the apple trees.

Contribution:
Project concept: Abhijeet Pal
Data Collection: Abhijeet Pal
Coding : Kanak Yadav & Gnanendar Reddy
Report Writing : Gnanendar Reddy & Abhijeet Pal
Video Making : Kanak Yadav

II. Background and Prior Work

The dataset'Plant Pathology 2020 - FGVC7' was based on a competition aimed on misdiagnosis of diseases impacting agricultural crops such as apple trees. We learnt that current diseases diagnosis is based on human scouting which is extremely time consuming and expensive. Also the location of apple trees in our country is on hilly areas in the state of Jammu and Kashmir and Himachal Prades, this makes it much more difficult to access such areas. Hence there is a need for classification of apple diseases based on images which can be taken by setting up cameras at appropriate locations in these areas. In the research paper [1] various Deep Learning models such as DenseNet121 and VGG16. The research paper achieved nearly 98% accuracy using DenseNet121 for tree diseases classification. We tried to implement these algorithms ourselves on the Kaggle Dataset and find the accuracy of these models along with other models such as ResNet50, Logistic Regression, SVM and Neural Networks. We have compared these models and stated the accuracy, f1 score and confusion matrix obtained by each method. We obtained nearly 91% accuracy using DenseNet121 which was the highest compared to the other methods.

III. Approach

The dataset consists of 1821 labelled images. We have taken 15% of these images i.e. 273 iages for testing and 25.5 % for validaiton i.e. 464 images and remaining 59.5 % images for training i.e. 1083 images. We have four categories of images, healthy, rust, scab and multiple diseases. We split the data into training, testing and validation according to the above splits. We then used DenseNet from tensorflow library and classified the data set. We have plotted accuracy of both training and validation as a function of number of epochs. We then used VGG16 model and ResNet50 models. We observed that accuracy obtained by DenseNet16 was better than VGG16 and ResNet50 models. We then applied Logistic Regression, SVM and Neural Network to the features obtained using ResNet50 model and have classified the data accordingly. We then concluded that DenseNet16 gives the highest accuracy compared to other models. The results are shown in the Result section and conclusions are made. The dataset class distribution is as follows:-

TABLE I
DATASET CLASS DISTRIBUTION

Class	Distribution
Healthy	516
Multiple Diseases	91
Rust	622
Scab	592

IV. EXPERIMENTS AND RESULTS

Various Deep Learning models such as VGG16, DenseNet121 and ResNet50 were used for classification. Different number of epochs were used to see trend with validation accuracy. Average Pooling layer was used. We observed that as the number of epoch was increased, the validation accuracy increased from 30% to 90%. For VGG16, we observed that as the number of epoch was increased the validation accuracy was increased from 30% to 50%. For ResNet50, we observed that it performs badly compared to the other methods. The validation accuracy is nearly 30% even when the number of epoch is increased.

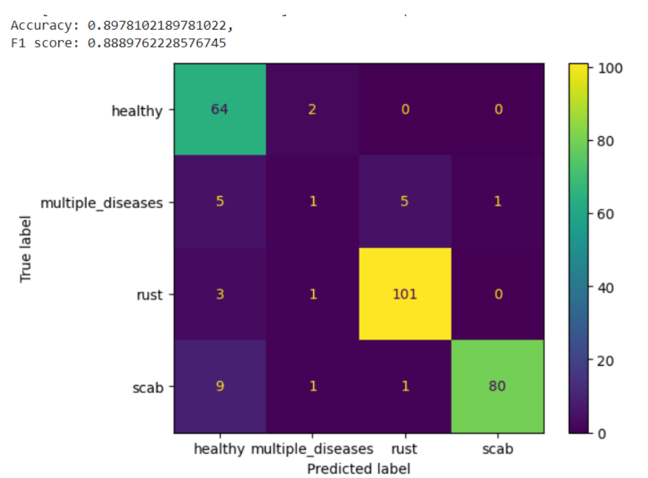

Fig. 1. DenseNet121 Confusion Matrix

The accuracy obtained by Logistic Regression is 52.65% and f1 score is 0.53. Accuracy obtained by SVM is 56.4% and f1 score is 0.539 whereas that obtained by MLP Classifier is 40.21% and f1 score is 0.276. The final results are summarized in the table II

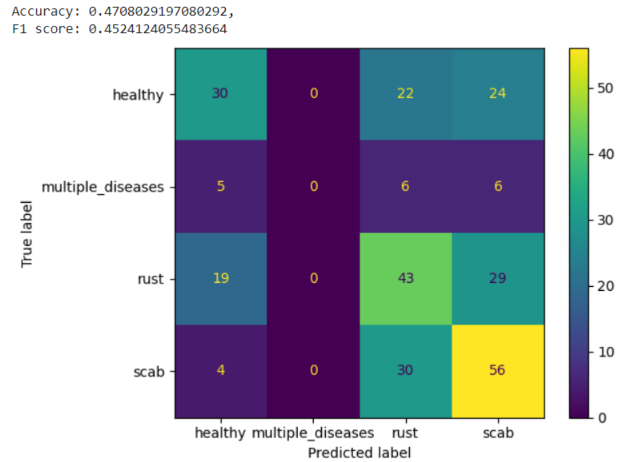

Fig. 2. VGG16 Confusion Matrix

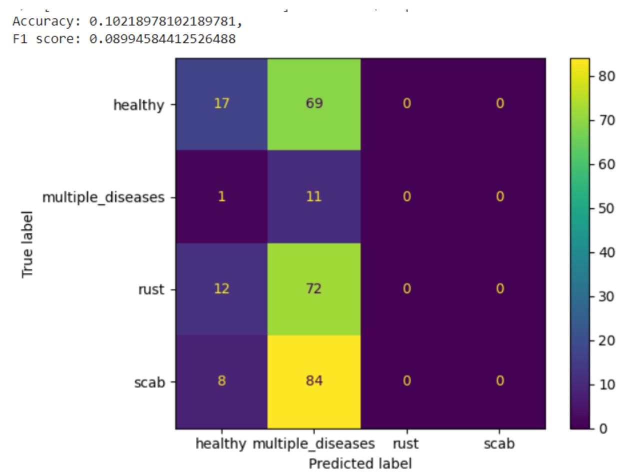

Fig. 3. ResNet50 Confusion Matrix

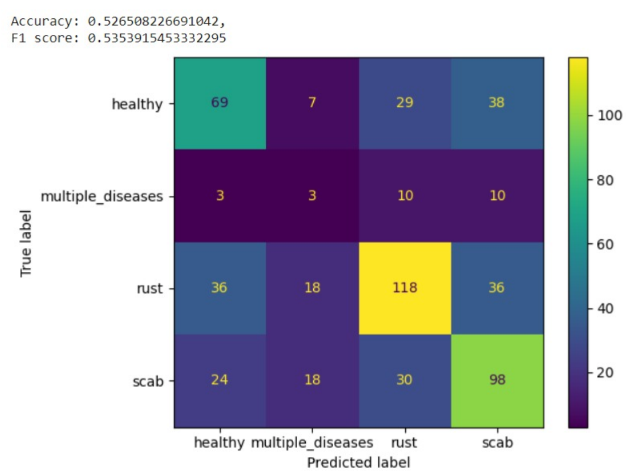

Fig. 4. Logistic Regression Confusion Matrix

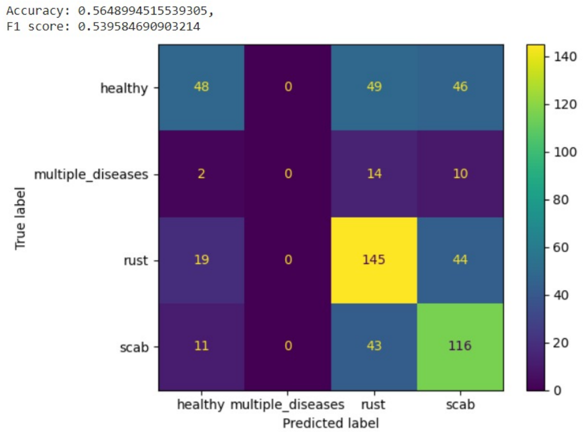

Fig. 5. SVM Confusion Matrix

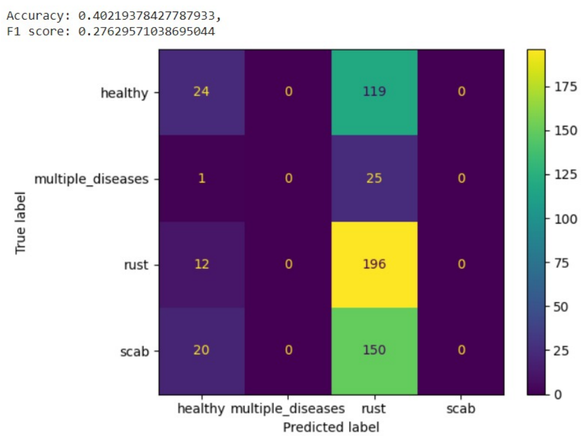

Fig. 6. Neural Network Confusion Matrix

V. Discussion and Conclusions

We find that for classification of foilar diseases in Apple trees, DenseNet121 gives the best accuracy among the models chosen. We can use classification by DenseNet121 on the images obtained using the on-site cameras and can classify the diseases on apple trees with 90% accuracy. This will prevent misdiagnosis of the diseases and will also prevent use of manual labor.

VI. Key Links

Github Repo :- Link
Demo Video :- Link

TABLE II
Performance Metric for Various Models Used

Model	Accuracy	f1_score
DenseNet121	0.89781	0.88897
VGG16	0.47080	0.45241
ResNet50	0.10218	0.08994
Logistic Regression	0.52650	0.53539
SVM	0.56489	0.539584
MLP Classifier	0.402193	0.276295

References

[1] R. Dhivya, R Vennila, G Rohini, S Mithila, K Kavitha "Foilar Disease Classification in Apple Trees," *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation*
[2] Kaggle Dataset used for the project
[3] ResNet50, VGG16, DenseNet121 libraries