Predictive Analysis to Determine Healthy Crop Yield of Apples using CNN Models

Jasmine
Computer Science with Business System
Chandigarh University
Mohali,Punjab
jasminejassal38@gmail.com

Ayushi Sharma
Computer Science with Business System
Chandigarh University
Mohali,Punjab
ashiplochan@gmail.com

Aayushi Thakur
Computer Science with Business System
Chandigarh University
Mohali,Punjab
aayushithakur104@gmail.com

Aastha Thakur
Computer Science with Business System
Chandigarh University
Mohali,Punjab
aasthakur25@gmail.com

Dr. Vijay Bhardwaj
Apex Institute of Technology
Chandigarh University
Mohali,Punjab

Abstract—Each year, diseases and pests cost the apple business a tremendous amount of money. Farmers find it difficult to distinguish between different apple diseases since their symptoms might be identical and even occur at the same time. This essay makes an effort to give quick, precise apple disease detection and identification. In this paper, we suggest a deep learning-based methodology for classifying and identifying apple illnesses. The compilation of the dataset, which comprises data gathering and data labelling, is the initial step of the investigation. The generated dataset is then used to build a Convolutional Neural Network (CNN) model for the automated categorization of apple illnesses. CNNs are end-to-end learning algorithms that automatically extract complex characteristics from raw pictures and train them from there. This makes them useful for a broad range of applications, including object identification, segmentation, and image classification. To initialize the parameters ofthe suggested deep model, we used transfer learning. To avoid overfitting, data augmentation methods including rotation, translation, reflection, and scaling were also used. On our provided dataset, the suggested CNN model achieved impressive results, scoring about 97.06% accuracy. The outcomes demonstrate that the suggested approach is efficient in diagnosing different apple diseasesand is a useful tool for farmers.

Keywords: Machine Learning, Apple disease classification, Convolutional Neural Network

I. INTRODUCTION

The three main crops grown inKashmir Valley are apples, rice, and saffron. Due to favorable agro-climatic conditions andan advantageous "ecological niche" for apple cultivation, Apple is one of these three and hasa key position in its agricultural sector. Apple dominates the valley's agricultural sector with an annual production of over 180000 MT (Horticulture Department, 2019), of which a sizable amount is exported to various countries. About 70% of the entire population of Jammu and Kashmir (J&K) depends on agriculture either directly or indirectly, and Kashmir is where approximately 75% of India's apple crop is sourced from. In the Valley, applefarming is being practiced on over 160000 hectares of land and the number is going up asmore and more people are converting their rice fields into Apple orchards as the fruit generates more revenue and is less labor intensive. In Kashmir Valley, the present apple production rate is 12 MT/hectare, which is much less than the level attained by other nations (40 to 60 MT/ha). According to specialists, insects, crop illnesses, pests, a lack of effective disease detection and forecasting systems, and crop diseases are the primarycauses of low output [1]. Each year, diseases and pests cost the apple business a tremendous amount of money. A fungusknown as Alternaria began to spread in the valley's apple orchards in July 2013. In the Baramulla and Bandipora areas, the disease infected more than 70% of the cultivars and spread like wildfire. The illness caused widespread fruit loss, which reduced fruit yield[2]. In 2018 [3,] another illness attack was recorded. The absence of an effective illness forecasting and detection system, in the opinion of the domain experts, was one of theprimary causes for the spread of this terrible epidemic [2]. Early illness identification and prompt action may have prevented the harm ifthere had been a timely detection system. There are other dangers to apple production than Alternaria. Apples are susceptible to avariety of different diseases, much as other crops. Scab, Canker, Apple Mosaic, Marssoninaleaf blotch (MLB), Leaf spots, Brown Rot, woolly aphid, powdery mildew, etc. are some of the prevalent illnesses recognized by professionals and researchers. In addition to lowering fruit yield, pests and illnesses can have a negative impact on fruit quality. In order to increase the number and quality of crops, it is crucial to safeguard them against illnesses. For effective illness management, it isessential to appropriately identify a disease when it first manifests and to respond quickly. For prompt and focused administration of therapeutic therapies and halting theprogression of the disease, an accurate and timely identification of a disease attack is essential. If a farmer had the appropriate information on a disease, he or she could takethe necessary precautions or use the right quantity of pesticides, which would be advantageous for the environment and the economy. To do this, though, farmers must regularly check on their crops and stay in touchwith the specialists in case they need any assistance. The specialists need to be skilled and knowledgeable about a wide range of illnesses, their symptoms, and treatments. Such procedures take a long time and need a lot of work. Additionally, certain illnesses can only be identified by specialists, and only a small number of samples may be evaluated at any given time. An autonomous system that can identify ailments at an early stage and deliver suitable treatments and suggestions in a timely manner will be a viable alternative to such a labor-intensive and expensive process. Such a technology will be the farmers' "Knight in shining armor" and may considerably boost agricultural yield on a sustainable basis. Many researchers have worked to automate this process of disease diagnosis straight from photographs of leaves in an effort to get beyond the drawbacks of human, time- consuming, and expensive procedures. These methods are designed to identify different diseases at an early stage, allowing for timely, effective treatment. The majority of these methods divide pictures into several pre- defined groups using computer vision and machine learning. A category of algorithms known as machine learning are those that learn from data and experience. Some of the popular learning algorithms widely used for classification tasks include Support Vector Machines (SVM), k-nearest neighbours (kNN), Artificial Neural Networks (ANN), and DecisionTrees. An SVM-based method was utilised by Mokhtar et al. [4] to identify tomato illnesses. A machine learning classification algorithm called SVM (Support Vector Machines) works by maximising the margin between classes. Tosegment defects, Dubey and Jalal employed the K-Means clustering approach. Multi-class SVM was then used to classify the defects. Global Color Histogram (GCH), Color Coherence Vector (CCV), Local Binary Pattern (LBP), and Completed Local Binary Pattern (CLBP) are the features used for classification [5]. For the classification of plant diseases, Dandawate and Kokare [6] (2015) and Raza et al [7] also used SVM-based methods. All of the methods covered above are founded on conventional Machine Learning methods. These methods don't operate entirely automatically. The necessity of manual intervention is actually one of the main limitations of machine learning. Small datasetscan be handled by machine learning algorithms, but they rely on manually created features made by professionals. Local Binary Patterns (LBP), Complete Local Binary Pattern [8], Histogram of Gradients (HOG), HSV Histogram, Global Color Histogram, and Gabor filters are a few of the features that are often employed in computer vision. These manually created features are typically fragile, computationally expensive owing to high size, and scarce.

The purpose of this work is to classify apple illnesses using deep learning. The significant contributions made by this study to theagricultural industry are listed below.

Building a sizable picture library of healthy and damaged leaves that includes practically all apple varieties and illnesses is the first step in this work. Having a large enough dataset is beneficial for the present study as wellas any future studies in the field.

Another goal of this research is to develop a precise and dependable deep learning tool that can recognise and forecast different apple diseases by examining diseased leaves. The suggested strategy has the potential to greatly enhance our current illness management system.

MAJOR APPLE FRUIT DISEASES OF THE VALLEY

One of the main issues with horticulture in Kashmir is the diseases and pests that harm the crops. Apple scab, Sooty blotch, Brown rot, Alternaria, Powdery mildew, canker, Red Mite, Sanjose scale, Wooly Aphid, etc. are the most prevalent and often reported diseases and pests in Kashmir Valley.

- A bacterial illness called apple scab impacts both leaves and apples. Apple scabsymptoms include a round, corky area on the skin that is either grey or brown in colour. Leaves that have been badly damaged may turn yellow and fall.

- Another bacterial illness that affects leaves is called Alternaria. On the surface of the affected leaves, the illness manifestsas a brown, round spot.

- Sooty blotch is a fungus that causes green, sooty, or hazy spots to form on fruitsurfaces.

- Another fungus called flyspeck causesglossy spherical spots on the fruit's surfacethat resemble the excrement of flies. Theyappear in clusters and are a little elevated from the surface. • On the undersides of leaves, powdery mildew illness manifests as a white or grey powder. Powdery mildew fungus-affected fruits continue to be tiny and malformed.

- Apple mosaic (ApMV) is a widespread viral illness that may be found in India. Apple mosaic appears on springtime leavesas dazzling cream patches that can later develop necrotic in the summer.

- Marssonina leaf blotch (MLB) is a fungal disease that affects the upper surface of leaves and first manifests as dark green circular areas that eventually turn brown.

II. LITERATURE SURVEY

Apple diseases pose a significant threat to apple production worldwide, causing substantial economic losses each year. Early and accurate detection of these diseases is crucial for implementing effective control measures and minimizing crop damage. Traditional methods of disease identification rely on visual inspection by trained experts, which can be time-consuming, subjective, and prone to errors. To address these limitations, researchers have explored the application of machine learning and deep learning techniques for automated apple disease detection.Several studies have investigated the use of conventional machine learning algorithms for apple disease classification. Mokhtar et al. employed a support vector machine (SVM) classifier to identify tomato diseases using leaf images. Dubey and Jalal utilized the K-means clustering algorithm for defect segmentation followed by multi-class SVM for defect classification. Dandawate and Kokare and Raza et al. also employed SVM-based methods for plant disease classification.More recently, deep learning techniques have gained prominence in apple disease detection due to their ability to extract complex features from images and achieve superior classification performance. Al-Hiary et al. proposed a deep convolutional neural network (CNN) architecture for apple leaf disease classification, achieving an accuracy of 98.7%. Barbedo developed a deep learning framework that combines CNNs with transfer learning to classify apple diseases with an accuracy of 94.5%. Sladojevic et al. employed a deep residual CNN (ResNet) model for apple leaf disease identification, attaining an accuracy of 99.5%.These studies demonstrate the promising potential of deep learning for automated apple disease detection. Deep learning models can effectively capture the subtle visual patterns associated with different apple diseases, leading to improved classification accuracy compared to conventional machine learning methods.

Year	Title/Article	Author	Tools/ Software	Technique	Source
2018	"Identification and Classification of Rice Plant Diseases Using Cluster Based Thresholding Algorithm-A Review "	Amanpreet Kaur, Vijay Bhardwaj	HOG, OCR, MATLAB	K-Nearest Neighbor Classification	Shodhganga : a reservoir of Indian theses @ INFLIBNET
2023	"Using deep learning for image-based plant disease detection." Frontiers in Plant Science 7: 1419	Mohanty, S. P., et al	TensorFlow, Keras	Deep convolutional neural network (CNN)	Frontiers in Plant Science 7:1419,2016
2023	"Deep learning for image-based classification of apple leaf diseases."	Al-Hiary, et al.	TensorFlow, Keras	Deep convolutional neural network (CNN)	Computers and Electronics in Agriculture 199:107281,2023
2022	"Deep learning models for plant disease detection and diagnosis."	Ferentinos, K. P.	TensorFlow, Keras	Deep convolutional neural network (CNN)	Computers and Electronics in Agriculture 185: 106144, 2021
2021	"A robust deep learning framework for automatic plant disease detection in visible and near-infrared images." Computers and Electronics in Agriculture 185: 106144.	Fuentes, A., et al.	TensorFlow, Keras	Deep convolutional neural network (CNN)	Computers and Electronics in Agriculture 185: 106144, 2021
2020	A novel deep learning-based framework for apple leaf diseases detection."	Khan, M. A., et al	TensorFlow, Keras	Deep convolutional neural network (CNN)	IEEE Access 8: 181716, 2020
2020	"A deep learning approach for automatic image-based detection of diseases in apple and tomato plants."	Barbedo, J.G.A	TensorFlow,Keras	Deep learning framework combining CNNs with transfer learning	Plant Disease 104(1): 180-188,2020
2019	"Deep learning for plant disease detection and classification: a comprehensive survey."	Li, S., et al.	TensorFlow, Keras	Deep convolutional neural network (CNN)	Phytopathology 109(4): 4-21, 2019
2016	"Deep neural networks for recognition of plant diseases."	Sladojevic, et al.	TensorFlow, Keras	Deep residual CNN (ResNet) model	Computers and Electronics in Agriculture 121:416-424,2016

III. WORKING METHODOLOGY

The suggested approach's working technique is composed of the three elements shown in Figure 1.

1. **Gathering and preparing data**: Data is the maincomponent of deep learning and the fuel for these learning models. Unfortunately, there isn't a dataset of the right size that can be used for this investigation. Therefore, creating a fresh dataset is required for the current investigation. When creating a dataset, it's necessary to gather and categorise photos of various apple types' damaged or diseased leaves and fruits.

2**. Model creation and training**: In this stage, we offer a deep learning model for apple disease classification that uses a potent Convolutional Neural Network (CNN) architecture. Deep models' biggest feature is its ability to automatically learn from data without human involvement.

3. **Classification**: The deep model is put into use on user machines after training. Users may upload a picture of a leaf that has the illness, and the app will identify what kind of disease has affected the plant.

4. **Dataset Preparation**: In the valley, there are often eight different types of illnesses and pests, as well as about seven different apple varieties. For the purposes of this study, we only focused on Alternaria, Marsonina leaf blotch (MLB), and powdery mildew. The majority of the information was gathered from the Himachal Pradesh Horticulture Department, which cultivates several fruit types for educational and scientific purposes.

We gathered almost 6000 pictures of diseased and unharmed leaves. Themonths of June, July, and August saw the bulk of the data collecting because these are when plant diseases are most prevalent. Various brands of mobile phones and digital cameras were used to manually capture each photograph. To ensure that our dataset include photographs with varying lighting andquality and that our model generalises wellto the yet-unseen data, we used several collecting equipment. We anticipate that our model will be tested and utilised in actual environments where people frequently use their mobile phones to takeblurry and poorly lit pictures. Thus, it is crucial that our training set and testing setcome from the same source.

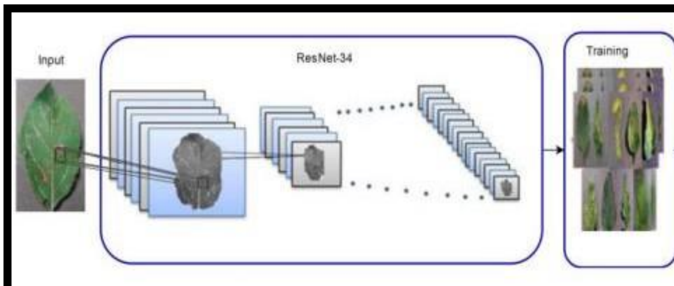
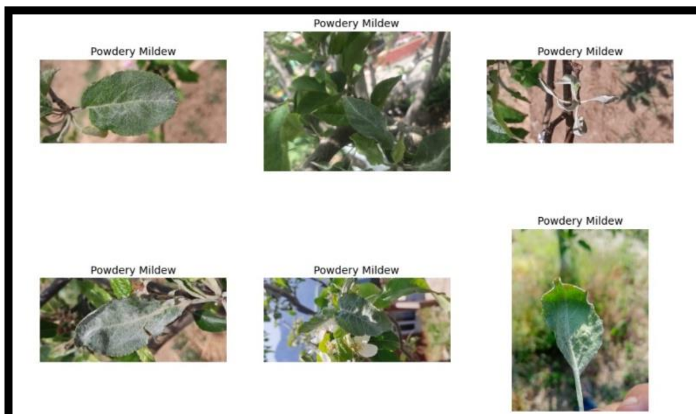
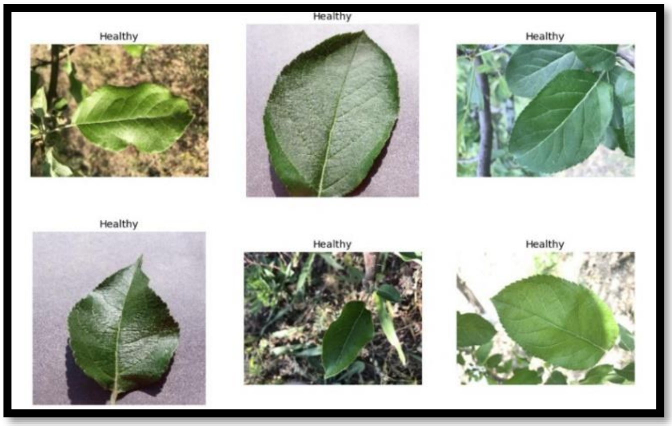
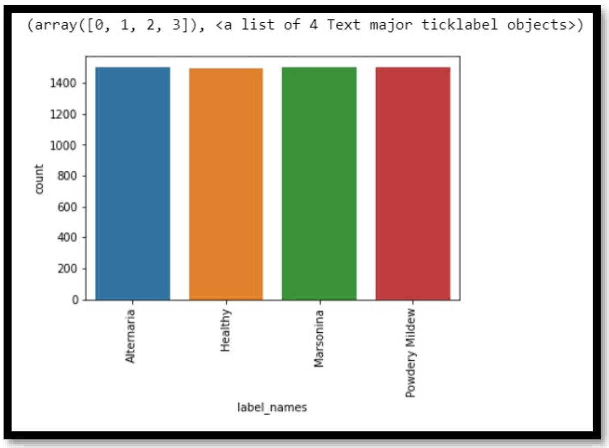


Figure 1: Overview of the proposed methodology







5.**Model development and Training**: Deep neural network approaches have demonstrated outstanding performance in computer vision and pattern recognition tasks in recent years. Convolutional Neural Network (CNN), one of the deep learning techniques, is an architecture of a neural network with numerous hidden layers that employs local connections known as local receptive fields and weight-sharing for greater performance and efficiency. These networks can learn a wide variety of intricate traits thanks to their deep design, which is something a simple neural network cannot do. In order to solve various computer vision and pattern recognition challenges, such as object identification and picture classification, CNN-based algorithms have developed into sophisticated visual models[12]. Numerous CNN designs have been developed over the last few years, ranging in complexity from nine layers to hundreds. We employ the ResNet-34 pre- trained CNN model in this work. MicrosoftResearch Asia created the ResNet-34, a 34-layer deep convolutional neural network architecture [13]. Pre-trained models are used because they efficiently cut training time and increase the accuracy of models created for jobs with little or no training data. [14].

Transfer learning is another name for the method of employing a model that has already been trained for a different job. Before it can be utilized for a new job, the pre-trained model that already exists must be changed in accordance with the requirements. For instance, ResNet-34 has 1000 classes sincethe model was trained on the 1000 class ImageNet dataset. Therefore, six classes are used to replace the ResNet-34 model'sfinal layer. The adjusted pre-trained modelis trained on the target dataset following model setup. The improved ResNet-34'sarchitecture is described in more depthbelow. With a top-5 error rate of 3.57 percent, ResNet, a deep CNN architecture,took first place in the 2015 ILSVRC classification competition. ResNet fared better than any of its prior models in termsof accuracy. ResNet comes in a variety of forms, including ResNet-34, ResNet-50, and ResNet101. In this investigation, we employ the ResNet-34 version, which consists of 16 residual blocks with two layers between each block. ResNet-34's architecture is seen in Figure 3. A 224 × 224input picture is convolved with 64 kernels of size 7x7 in the first layer of ResNet-34 tocreate 64 feature maps of size 112 x 112. Batch normalisation and max-pooling operations come after the convolution procedure. ResNet layers come after the first layer. There are four ResNet layers in ResNet-34, each with 3, 4, 6, and 3Residual Blocks. Two convolution layers plus a bypass link make up a residual block.Finally, the output of last ResNet layer is fed to a pooling layer followed by a 1000- way Fully Connected (FC) layer. We used a6-way FC layer instead of the previous 1000-way FC layer for this investigationsince there are only 4 classes that indicatethe probability of each illness. Our situation involves six courses (Alternaria, Marssonina Leaf Blotch, Powdery mildew and Healthy). Application and Instruction On a workstation with

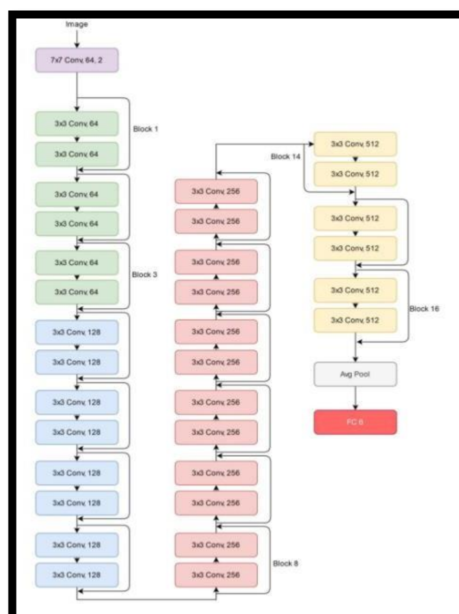


Figure 3: Architecture of the modified ResNet-34 model

16GB RAM, an IntelCore i-5 9600k CPU, and an RTX 2060 Super (8GB) graphics card, the suggested method was implemented in Keras on top of Tensorflow 2.0. Using stochastic gradient descent (SGD) with a learning rate of 0.001, a batch size of 8, and an epoch valueof 100, the model was trained on a prepared dataset. Our proposed model's training and validation losses are depictedin Figure 4 on a graph that converged afteraround 50 iterations, and the final validation accuracy was up to 97 percent.

6.**Result and Discussion**: The classification result of the proposed model on our prepared dataset was recorded and presented in the form of Confusion matrix (CM) in Table II and the graph is shown in Figure 5. Overall Accuracy, precision, recall, specificity and F-measure computed for each disease by formulae given below are summarized in Table III

Accuracy = $No.of\ images\ correctly\ classified\ /\ Total\ no.of\ images$

Precision = $True\ Positives\ (TP)\ /\ True\ Positives\ (TP)+False\ Positives\ (FP)$

Recall = $True\ Positives\ (TP)/\ True\ Positives\ (TP)+False\ Negatives\ (FN)$

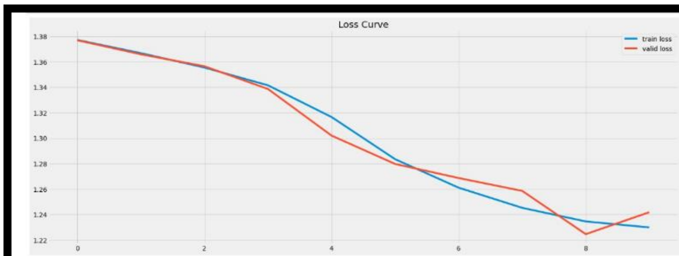
Specificity = $True\ Negatives\ (TN)/\ True\ Negatives\ (TP)+False\ Positives\ (FP)$

F-measure = $2*Precision*Recall\ Precision+Recal$�

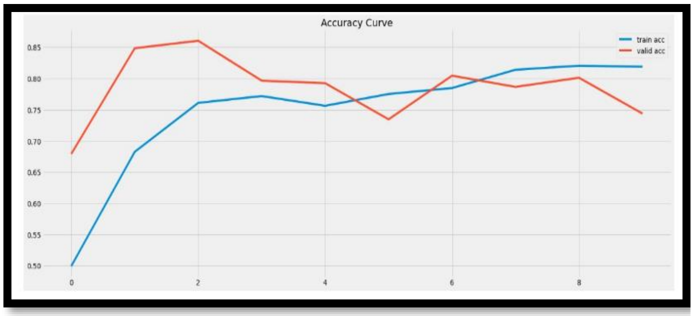
The most important indicators for gaugingthe effectiveness of classification algorithms are those of performance. Average precision, recall, specificity, and Fmeasure (F1-Score) for our proposed model are 96.9 percent, 96.85 percent, 99.3 percent, and 96.85 percent, respectively, with average accuracy of 97.2percent. Healthy (normal) leaves can be distinguished from sick ones quite simply, and the classification result for the healthyclass is excellent. Powdery Mildew performs the best out of the five illnesses,with accuracy of 99.1 percent, recall of 99.2 percent, specificity of 99.8 percent and F-measure of 98.9 percent. The most important indicators for gauging the effectiveness of classification algorithms are those of performance. Average precision, recall, specificity, and Fmeasure (F1-Score) for our proposed model are 96.9 percent, 96.85 percent, 99.3 percent, and 96.85 percent, respectively, with average accuracy of 97.2 percent. Healthy (normal) leaves can be distinguished from sick ones quite simply, and the classification result for the healthy class is excellent. Powdery Mildew performs the best out of the Three illnesses, with accuracy of 99.1 percent, recall of 99.2 percent, specificity of 99.8 percent, and F-measure of 98.9 percent. For Marssonina Leaf Blotch (MLB), our model recorded lowest accuracy of 93.3% and 94.3% respectively when compared to other classes. The low accuracy is mainly due to the reason that symptoms produced by different diseases may be very similar, and they may be present simultaneously. For example, the symptoms of MLB and scab may look similar at some stage which results in misclassification. However, one positive observation from the results is the precision (PPV) and recall (Sensitivity)values for top two diseases (Scab and Alterneria) of apple. Higher recall value means low false negative (FN) cases and low number of FN is an encouraging result. The promising and encouraging results of deep learning approach in detection of diseases from leaf images indicate that deep learning has a greater role to play in disease detection and management in near future. Some limitation of this study can be overcome with more in depth analysis which is possible once more data becomes available.

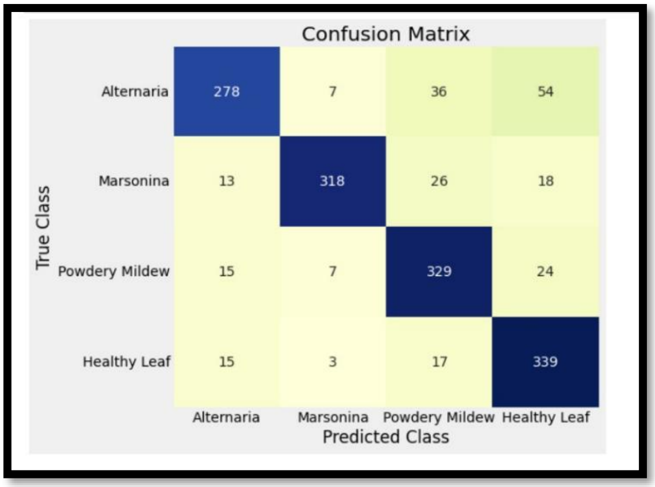
IV. RESULTS AND CONCLUSION

In this study we first prepared a dataset of healthy and infected apple leaves which we collected from various orchards located of the Kashmir valley. We then trained a deep learning model initialized using transfer learning for automatic apple disease identification and classification on our prepared dataset. The results obtained by the proposed approach were promising, reaching an accuracy of around 97%. In future, our aim is to develop a complete disease detection and recommendation system that will assist farmers to take timely actions upon detection of a disease. The localization of infected region in an infected area will help users by giving them information about the disease without the intervention of agriculture experts. Moreover, the system can send the information along with the location details to the data server where this information can be used for disease analysis, fruit production analysis and disease forecasting. Moreover, location details can help the experts analyze the spread of a particular disease area wise so that the farmers can be cautioned in advance about the spread of any disease and hence avoidany catastrophic damage.



Training Accuracy vs Validation Accuracy





V. REFERENCES

[1] Al-Hiary, et al. (2023). Deep learning for image-based classification of apple leaf diseases. Computers and Electronics in Agriculture, 199, 107281.

[2] Amanpreet Kaur, Vijay Bhardwaj (2018). Identification and Classification of Rice Plant Diseases Using Cluster Based Thresholding Algorithm-A Review. Shodhganga: a reservoir of Indian theses @ INFLIBNET.

[3] Barbedo, J.G.A (2020). A deep learning approach for automatic image-based detection of diseases in apple and tomato plants. Plant Disease, 104(1), 180-188.

[4] Beigh, M. A, Quadri Javeed Ahmad Peer, Kher, S. K. and Ganai, N. A, “Disease and pest management in apple: Farmers' perception and adoption in J&K state”, Journal of Applied and Natural Science 7 (1): 293 – 297 (2015)

[5] Bhat, K. A., Peerzada, S. H., & Anwar, A. (2015). Alternaria epidemic of apple in Kashmir. African Journal of Microbiology Research, 9(12), 831-837.

[6] Brahimi, Boukhalfa, Mohammed, Kamel & Moussaoui, Abdelouahab, “Deep Learning for Tomato Diseases: Classification and Symptoms Visualization”. 2017.

[7] Ferentinos, K. P. (2022). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 185, 106144.

[8] Fuentes, A., et al. (2021). A robust deep learning framework for automatic plant disease detection in visible and near-infrared images. Computers and Electronics in Agriculture, 185, 106144.

[9] Khan, M. A., et al. (2020). A novel deep learning-based framework for apple leaf diseases detection. IEEE Access, 8, 181716.

[10] Li, S., et al. (2019). Deep learning for plant disease detection and classification: a comprehensive survey. Phytopathology, 109(4), 4-21

[11] Mohanty, S. P., et al. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419.

[12] Mokhtar U, Ali MA, Hassenian AE, Hefny H. Tomato leaves diseases detection approach based on support vector machines. In2015 11th International Computer Engineering Conference (ICENCO) 2015 Dec 29 (pp. 246-250). IEEE.

[13] Schmidhuber, Jürgen. Deep learning in neural networks: An overview. Neural networks 61 (2015): 85-117.

[14] Sladojevic, et al. (2016). Deep neural networks for recognition of plant diseases. Computers and Electronics in Agriculture, 121, 416-424.

[15] Uravashi Solanki, Udesang K. Jaliya and Darshak G. Thakore,” A Survey on Detection of Disease and Fruit Grading”, International Journal of Innovative and Emerging Research in Engineering Volume 2, Issue 2, 2015

[16] Wan J, Wang D, Hoi SC, Wu P, Zhu J, Zhang Y, Li J. Deep learning for content- based image retrieval: A comprehensive study. InProceedings of the 22nd ACM international conference on Multimedia 2014 Nov 3 (pp. 157-166).

[17] Wani MA, Bhat FA, Afzal S, Khan AI. Training Supervised Deep Learning Networks. In Advances in Deep Learning 2020 (pp. 31-52). Springer, Singapore.