**Classification Prediction of Kashmiri Apple Plant by Deep Learning Techniques**

***Abstract*: “Apple” proudly represents the fruit industry of Kashmir, hereby representing about nighty-eight percent of the total fruit production in the valley. It is the backbone of the economy of Kashmir valley particular in Shopian district. Sadly, there has been no significant increase in the production of this industry during the last few years due to various reasons but one of the most prominent reasons is the infection spread across the crops. India is an agricultural nation, and it is an important issue for the prompt detection of diseases among plants. The early detection of apple plants will enable apple producers to take the necessary steps or precautions to save the fruits from infection. The manual or traditional methods for apple plant disease predictions are very time consuming and laborious, which includes the lab diagnoses to be provided. But here is where the new technology comes up, with the advent in machine learning and deep learning it is now possible to quickly determine if a plant is infected or not with a reliable accuracy.**

**Keywords**

Convolutional Neural Network, Pooling, Apple Disease.

I. INTRODUCTION

In a big and crowded nation like India, agriculture is one of the most crucial parts, and one of the major problems is the illnesses that these agricultural plants contract or are affected with. Given that the world's population has dramatically increased and most goods on the market have become more expensive, the identification or forecast of plant diseases is a critical issue that requires attention. Thus, obtaining food continues to be a necessity for all living things. India is renowned as an agricultural country thanks to its fertile soil, which supports the economy and meets the country's basic needs. For profitable agriculture, the agricultural industry has been creating hybrid plants and premium seeds. Early detection of plant disease enables farmers to immediately remove it or take measures to stop it from infecting other plants. **[1]** Around 80% of all apples produced in India are grown in Kashmir Valley, which is well known for its specialty throughout the country. Ambri, Treal, Maharaji, Kesaer, and other apple types are indigenous to Kashmir. In the Budgam, Baramulla, Kupwara, Anantnag, Shopian, Srinagar, and Ganderbal districts, apples are farmed. These plants are vulnerable to a number of ailments, the most prevalent of which being apple rot, apple scabs, and Alternaria leaf blotch.**[2]** In Kashmir, the apple business employs 33 lakh people directly and indirectly. **[3]** Apple producers throughout Kashmir are concerned about this year's early fruit fall, as well as the continued spread of crop diseases Alternaria and leaf blotch in other areas. All of this necessitates the development of a system for identifying illnesses in their early stages. We can construct a model for Plant disease detection, specifically for Apple plants, using machine learning and deep learning. The catch is that when a plant becomes sick, the leaves are the first to exhibit indications, therefore the model can anticipate by analysing photos of the leaves. A paradigm or framework like this will assist us in better diagnosing illnesses.

The Plant Disease Predictor will be made up of two primary components:

1. A feature extractor

2. A Classifier

as shown in the Figure1.

The feature extractor takes visual information from input photos and delivers them to the training model. During training, the model use prediction algorithms to analyse the supplied photos and forecast whether they are contaminated or not. **[4]** Machine learning methods are used for feature extraction in the early stages, but deep learning image processing, such as Convolutional Neural Network (CNN), eventually replaces the obsolete methodologies. Object identification, image alignment and stitching (to form a panorama), 3D stereo reconstruction, and navigation for robots/self-driving automobiles are all examples of jobs that require feature extraction.

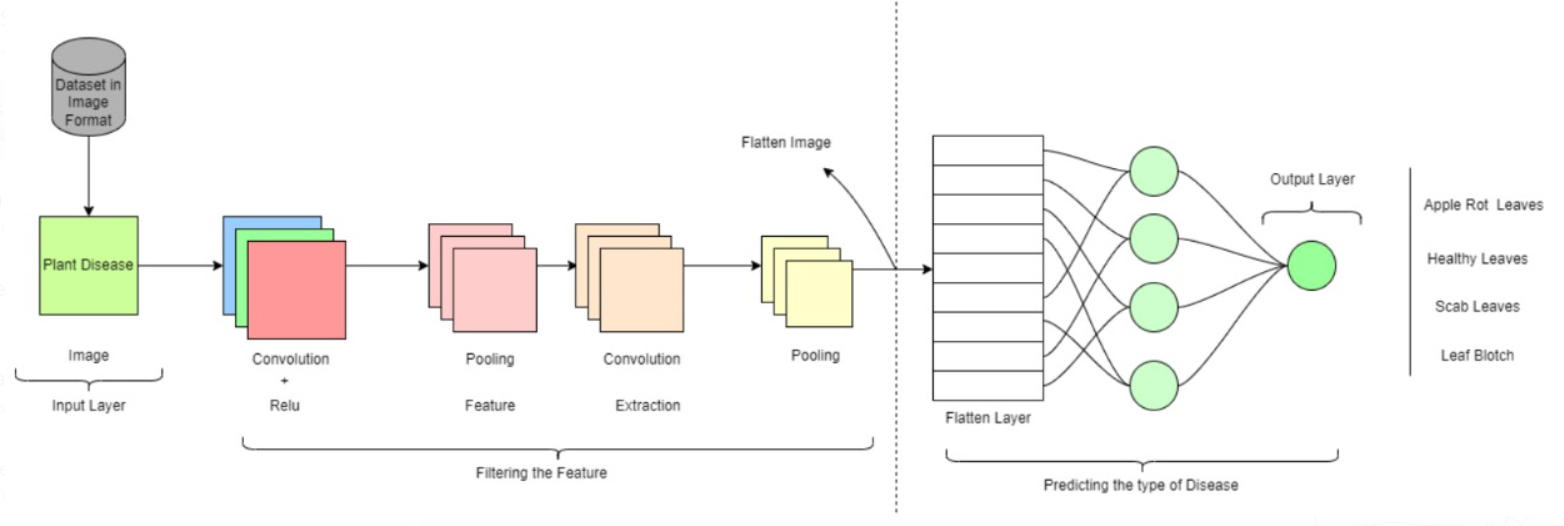


Figure 1. Structural Outline of Model

Traditional feature detection methods include Harris Corner Detection, Shi-Tomasi Corner Detection, Speeded-Up Robust Features (SURF), and others.

Traditional feature extractors can be replaced with convolutional neural networks (CNNs), which have a strong capacity to extract complex features that represent the picture in greater detail, learn task-specific characteristics, and are significantly more efficient. Deep Learning feature extraction approaches include Super Point, D2-Net, LF-Net, and others.

Despite being exposed to all of these current tools, Plant Disease Prediction for Apples remains one of the industry's most unexplored topics. Despite these considerations, several models are produced throughout the year with various adjustments to improve accuracy and dependability. We analysed some work on CV-aided plant disease prediction frameworks and discovered that relatively little effort has been done on the Apple Plant Disease Prediction. As a consequence, we created a deep learning-based architecture for the Apple Plant disease using the most recent dataset available. The collection includes photographs of illnesses such as Apple rot, Apple scab, and Alternaria leaf, as well as images of healthy plant leaves.

II. LITERATURE SURVEY

Apple plant disease as discussed is very common in the Northern parts of the country and there has been studies for detecting apple plant disease. Some of the studies by the great intellect minds are:

Xiaopeng Li and Shuqin Li [12] have used convolutional structures and Transformer structures; the convolutional structure is used to extract the global features of the image, and the Transformer structure is used to obtain the local features of the disease region to help the CNN see better. The parameters and FLOPs (Floating Point Operations) of the model are significantly reduced by using depth wise separable convolution and linear-complexity multi-head attention operations. The model used by them gave a stupendous 96.85 % of recognition after 100 epochs.

[13] Helong Yu, Xianhe Cheng, Chengcheng Chen, Ali Asghar Heidari, Jiawen Liu, Zhennao Cai & Huiling Chen had developed a model based on residual network (ResNet50) and further improving it. Their experimental results show that the average precision, recall, and F1-score of the proposed model for leaf disease identification are 0.957, 0.958, and 0.957, respectively.

[14] S. Aleqethami, B. Almtanni, W. Alzhrani, M. Alghamdi from the agricultural sector of Saudi Arabia proposed three models in their studies that are, SVM, CNN and KNN, with different image processing methods to detect and classify the apple plant leaves as healthy or diseased.

In our model we have used fully connected convolutional neural networks with sufficient dense layers and max pooling for the diseases like apple rot, alternaria and leaf blotch as well as scab leaves. We can further dive deeply into the methodology used in our study.

III.METHODOLOGY

Our proposed model for plant disease prediction is built on deep learning-enhanced feature extraction, categorization, and prediction approaches. The following sub-components make up the entire model:

A. Data Analysis

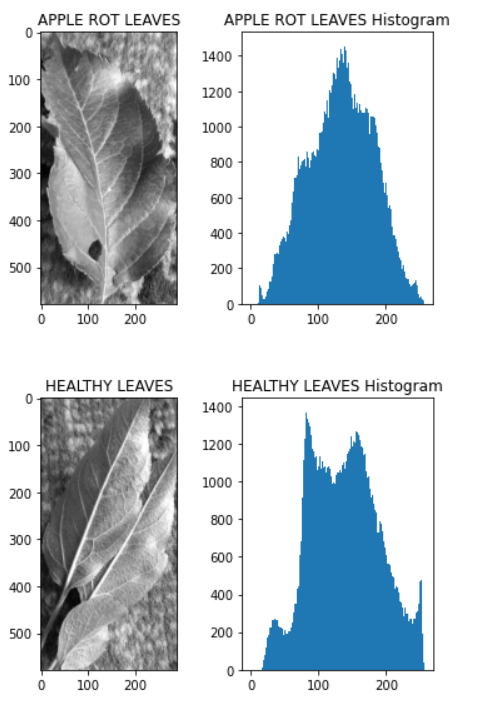
**[5]** Data analysis is the technique of working with data to extract valuable information from it so that educated decisions may be made. We can make better judgments when we can extract meaning from facts. Furthermore, we live in an era where we have an abundance of data at our fingertips, therefore it is a sensible habit to examine it all the time. We examined the dataset for Kashmiri Apple plants. First, we studied the photographs provided to us, some of which are as follows:

Graphical user interface

Description automatically generated

Figure 2. Infected Leaves

The following graphs show all the varieties in our dataset:



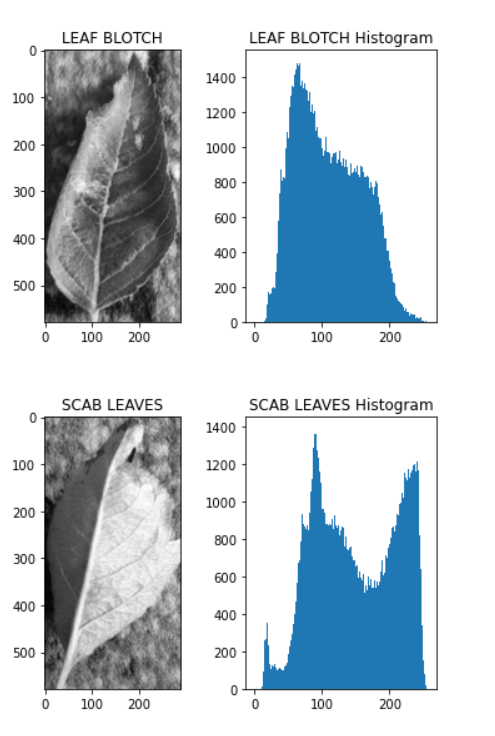


Figure 3. Graphs of the leaves

The data we examined provided us a solid notion of the picture intensity and data distribution.

B. Data Pre-processing

Data pre-processing is a data mining technique used to change the raw data into a format that is both practical and effective. Pre-processing is a multi-step procedure since it is one of the main factors influencing the model's performance.[6] The data may not be sufficient for a suitable model to train or test, as we discovered after analysing the data. This may also be construed to mean that a deep learning model often performs well when given a large amount of data. Having a sizable dataset to feed the deep learning network, however, is not always feasible. To get beyond this obstacle, we employed picture augmentation. The key reason for the picture augmentation was that a deep learning model would not be able to effectively learn patterns from the data if there wasn't a sufficient quantity of data available. By modifying the current data, image augmentation creates new data that may be used for model training. We didn't apply all of the augmentation strategies, but we carefully considered the facts before making our decision. For instance, using the following instructions in our model produced the following results for a single image:

*Chart, surface chart

Description automatically generated*

Figure 4. The Original Image

Chart

Description automatically generated

Figure 5. Augmented Image 1

Chart, surface chart

Description automatically generated

Figure 6. Augmented Image 2

Chart, surface chart

Description automatically generated

Figure 7. Augmented Image 3

**[7]** Then, using picture decolorization, we converted the photograph from colour to black and white. Widely employed in single-channel image processing, black and white printing, etc., it is known as the method to convert a colour image to a grayscale one. We adopted this method because it is easier to correctly identify illnesses like apple scab and apple rot that are caused by fungal or bacterial infection. This offers the model the advantage of being able to train more effectively. Additionally, we labelled the target, or the diseased plant leaves, using Label Encoder. We divided each class of sickness into numbers like 0,1,2,3 because the data we obtained was in string format, which made it difficult to train and develop the model. As a result, Label Encoder was utilised to get around the impediment.

C. Selection of model and Training

**[8]** There are several libraries available that provide easy-to-use machine learning libraries such as scikit-learn and keras, making it simple to fit many different machine learning models on a given predictive modelling dataset. The issue of applied machine learning is deciding which model to utilise for a given task. Aside from performance, other aspects to consider when selecting a model include complexity, maintainability, and available resources. For our model, we employed Convolutional Neural Networks (CNN), a deep learning approach. The major purpose for doing so was to improve performance by making full use of the patterns found in the images.

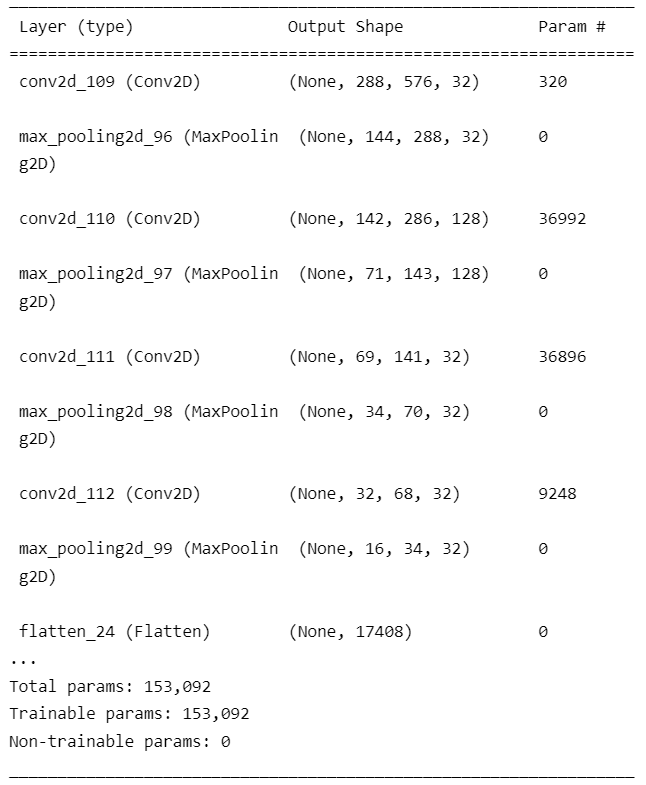


Figure 8. Convolutional and Pooling Layers

**[9]** CNN is a type of deep neural network that is commonly used to evaluate visual information. The main line is that ConvNet's purpose is to compress the pictures into a format that is easier to process while retaining elements that are crucial for making accurate predictions. Following the convolutional layer, the pooling layer is used to reduce the spatial size of the Convolved feature. By lowering the size, the computer power required to process the data is reduced. We applied max pooling. So, what we did was find the greatest value of a pixel in a region of the picture. This removes all noisy activations and conducts de-noising as well as dimensionality reduction. The model's block diagram during training is shown in Figure 8. The prior figure shows us about the convolutional and pooling layers used in the model.

IV. TESTING

[10] Testing is the process of evaluating the performance of a fully trained model on a testing set. The testing set is made up of samples that have been split from the training and validation sets. This is the stage at which our models begin to make accurate predictions. The significance of testing a model is that many models may exhibit great accuracy during the training phase but fail to perform well when presented with unknown data. So, by testing our model, we can establish its boundaries and weaknesses, allowing us to fine-tune it. In our situation, we first deployed a model without any data augmentation, and the results are shown in the figures below: This demonstrates the model's poor performance in the absence of sufficient data. We used epoch to train our model. [11] An epoch is a word used in machine learning to describe the number of passes the machine learning algorithm has made across the full training dataset. Figure7 depicts the accuracy after 10 epochs, which is likewise inconsistent and unsatisfactory. Even after raising the number of epochs to 50, we were unable to get the desired outcomes. Figure8 depicts the inconsistency of the accuracy, which is either stable or declining dramatically after 50 epochs. This is due to the low number of images available in the dataset and a way to represent how less dataset can affect the performance of a deep learning model. Thus, if a model does not have a good precision and predictive ability then the model should be tuned in a better way.

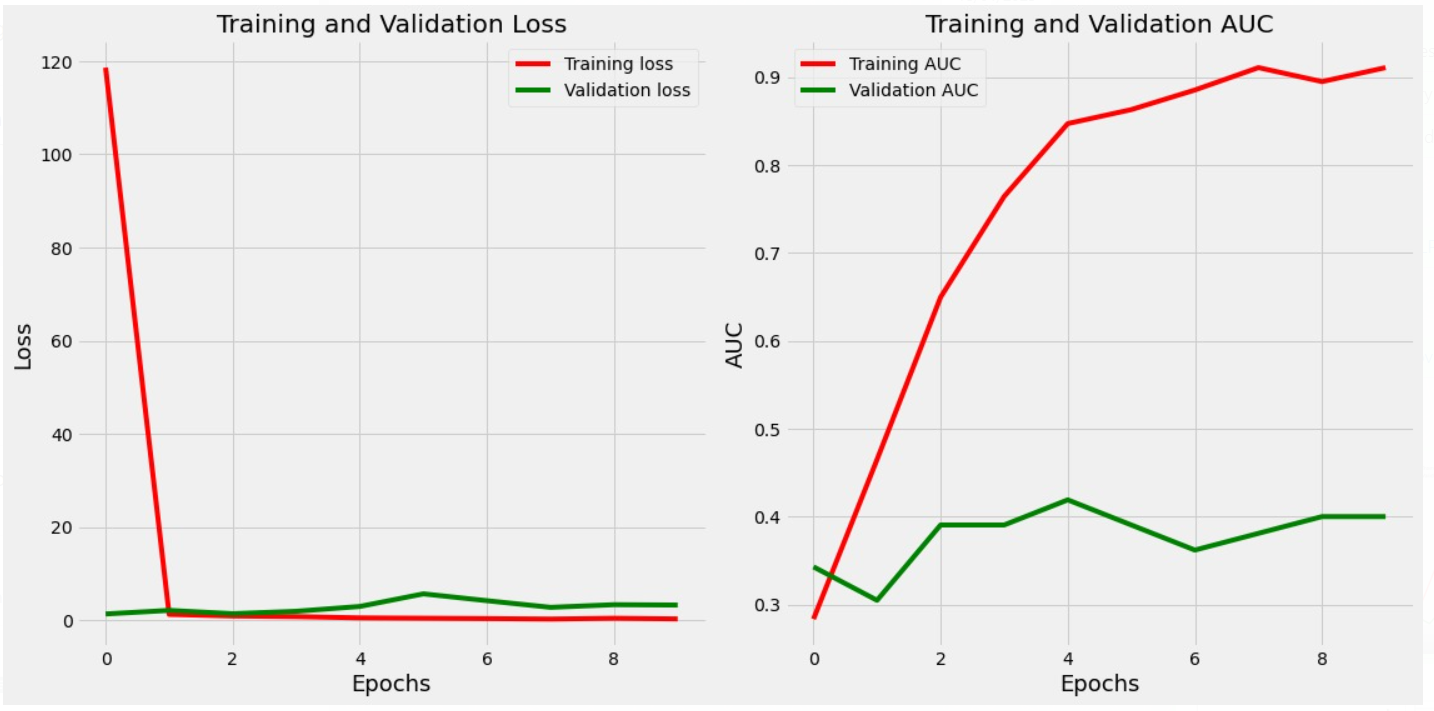


Figure 9. After 10 Epochs

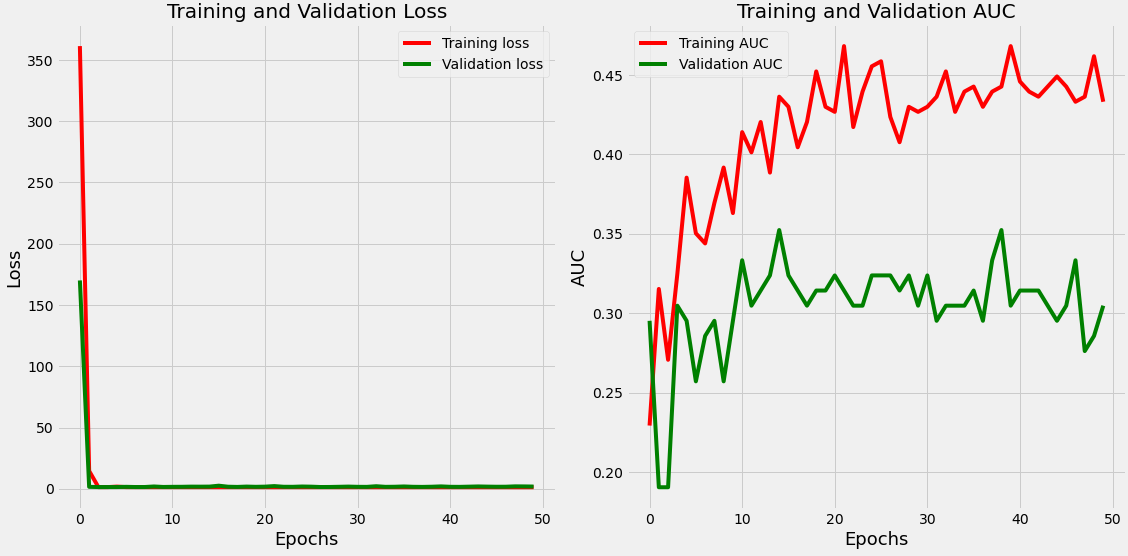


Figure 10. After 50 Epochs

Then, using data augmentation and increasing the amount of data, we tweaked our model. This expanded data was then sent into the system. Now, the below Figure 11 shows us the accuracy and performance of the model after image augmentation.

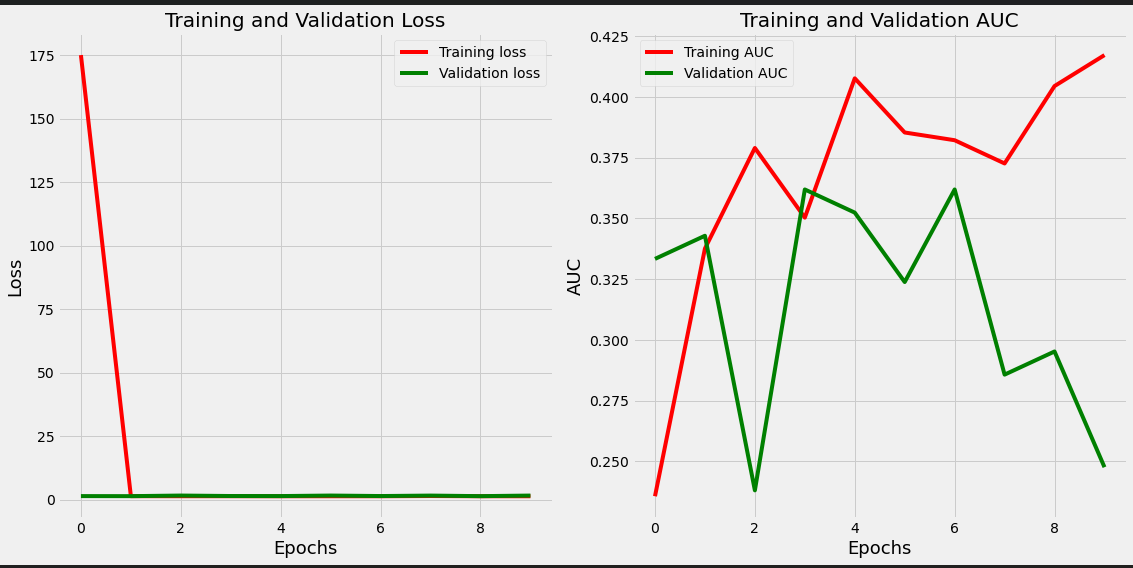


Figure 11. After Data Augmentation

After the model is created, testing data confirms that it can generate correct predictions. If the training and validation data contain labels to monitor the model’s performance metrics, the testing data should remain unlabelled. Test data is a final, real-world verification of an unknown dataset to ensure that the ML algorithm was properly trained. As a result, we now know that if the dataset given for training is insufficient, even a large number of epochs will not boost the model's efficiency. As a result, data or image augmentation played a major role in enhancing and tuning the model. This concludes the testing phase as well as the Methodology used in our model.

V. CONCLUSION

The findings of this study will be extremely beneficial to farmers and other Kashmiri apple growers, allowing them to implement countermeasures against contaminated trees. According to the study's findings, the predictive accuracy of the model for Apple Plant Disease Detection is 79.84% in the

testing phase, which is excellent for prediction and assessment. This will mainly ensure that the model will predict the diseases correctly and will also help the farmers to detect all the diseases to take countermeasures. Based on the findings of this paper, subsequent researchers will have a lot simpler time working on more areas of this discipline and delivering high-quality research. As a result, our publication will serve as a more trustworthy source of information for everyone. We can surely make it better, as we all know there is always room for development and polishing. In addition to healthy plants, our model can identify three common diseases: apple rot, apple scab, and Alternaria leaf blotch which are most prominently found. This research can also help to design more robust models capable of detecting a wide range of disorders. The fusion of features from multiple networks can also be implemented or implanted. Furthermore, PCA and LDA which are the dimensionality reduction techniques can be used to obtain better and efficient results. This research may also be utilised to not only construct models for Kashmir apples or any specific plant, but also to forecast illnesses in a range of plants at the same time. Other Convolutional models, such as R-CNN, faster R-CNN, and ResNet, may be used to assess the results in the future.

VI. REFERENCES

[1] Kashmir Observer, Apple Economy In Jammu & Kashmir: Changing Paradigm, https://kashmirobserver.net/2021/03/22/apple-economy-in-jammu-kashmir-changing-paradigm/

[2] Kashmir Observer, Kashmir Apple Grapples with Politics and Pandemic, https://kashmirobserver.net/2020/09/28/kashmir-apple-grapples-with-politics-and-pandemic/#:~:text=There%20are%20113%20varieties%20of,some%20of%20the%20known%20varieties.

[3] The Citizen, Kashmir Orchardists Warn of Premature Fruit Fall, Leaf Blotches, https://www.thecitizen.in/life/kashmir-orchardists-warn-of-premature-fruit-fall-leaf-blotches-288888

[4] Towards Data Science, Image Feature Extraction: Traditional and Deep Learning Techniques, https://towardsdatascience.com/image-feature-extraction-traditional-and-deep-learning-techniques-ccc059195d04

[5] H. Sharma, “D-KAP: A Deep Learning-based Kashmiri Apple Plant Disease Prediction Framework.” Institute of Electrical and Electronics Engineers (IEEE), Oct. 01, 2022. doi: 10.36227/techrxiv.21210320.v1.

[6] Analytics Vidhya, Image Augmentation Techniques for Deep Learning Models,https://www.analyticsvidhya.com/blog/2021/03/image-augmentation-techniques-for-training-deep-learning-models/

[7] X. Zhang and S. Liu, “Contrast preserving image decolorization combining global features and local semantic features,” The Visual Computer, vol. 34, no. 6–8. Springer Science and Business Media LLC, pp. 1099–1108, May 03, 2018. doi: 10.1007/s00371-018-1524-8.

[8] Machine Learning Mastery, Gentle Introduction to Model Selection for Machine Learning,https://machinelearningmastery.com/a-gentle-introduction-to-model-selection-for-machine-learning/

[9] Analytics Vidhya, Introduction To CNN, https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/

[10] H. Wang and H. Zheng, “Model Testing, Machine Learning,” Encyclopedia of Systems Biology. Springer New York, pp. 1405–1405, 2013. doi: 10.1007/978-1-4419-9863-7\_231.

[11] Radiopaedia, Epoch Machine Learning, https://radiopaedia.org/articles/epoch-machinelearning#:~:text=An%20epoch%20is%20a%20term,of%20data%20is%20very%20large).

[12] https://doi.org/10.3390/s22010173

[13] Apple leaf disease recognition method with improved residual-network, https://link.springer.com/article/10.1007/s11042-022-11915-2

[14] Disease Detection in Apple Leaves Using Image Processing Techniques,

https://www.etasr.com/index.php/ETASR/article/view/4721