



---

# Fruits Leaf Disease Detection Using Convolutional Neural Network

---

Deepak Pantha\*

\*Aadikabhi Bhanubhakta Campus, Nirmal Secondary School, Vyas Municipality, Tanahun, Nepal.

Corresponding Email: \*[deepakmphil@gmail.com](mailto:deepakmphil@gmail.com)

**Received:** 28 September 2023    **Accepted:** 17 December 2023    **Published:** 01 February 2024

**Abstract:** *Due to the traditional agricultural system, losses of millions of money have been loses every year. Farmers were always ready in agricultural work without risking their lives. If smart methods can be adopted in the agricultural system, the farmers will not have to suffer much damage. Using machine learning and testing with Convolutional Neural Network algorithm (mobileNet method), in this research to find out the actual accuracy, 3642 photos of apple leaves of Kaggle dataset and CSV files are used. In this paper, using Python language with the help of Jupyter notebook, Eposes has been tested 15 times to create confusion metrics. In this paper, precision, recall, f1\_score and average accuracy have been found and studied. An average accuracy of 95 percent has been obtained from the study. 95% accuracy is considered as a good result of the test using machine learning. By adopting this method, we can also give more motivation to the agricultural sector.*

**Keywords:** *Convolutional Neural Network, Deep Learning, Smart, Augmentation, Fruits Leaf Disease.*

## 1. INTRODUCTION

The traditional agricultural system is a lot of hard work and suffering but after the decline in income, the agricultural system is gradually being transformed into a smart agricultural system. In developed countries, like every other work, agriculture has been using various technologies, but it did not seem to make any changes in Nepal. In Nepal, the meaning of smart is still not understood or understood by stakeholders. A smart system is one where the machine makes the machine work.

Machine learning's robust computational capabilities have significantly influenced large-scale farming, particularly in fruit disease detection system deployed for on-field analysis and yield prediction. These systems play a vital role in identifying and predicting farm-related diseases



and defects. Moreover, they contributed extensively to agricultural research, aiding in comprehending and addressing various issues affecting the industry.

It is usually not easy to distinguish whether the fruits that people buy and eat are infected with any disease or not. In this way, the plants that are being produced and grown by the farmers will be uncomfortable to come in contact with agricultural experts and food experts at any time. In this way, farmers are forced to suffer losses of billions of rupees every year in agriculture. People who pay money to eat fruit are also being forced to eat rotting fruit whether they want to eat fruit without disease or not.

Nowadays, there is a need for an application that can distinguish whether the fruit is healthy or diseased by purchasing it with the help of the mobile phone that every person has. The application made with the help of machine learning will be able to distinguish the condition of the fruit based on a photo. It has been a few years since such knowledge was developed and used in developed countries. Nepali people do not know the generality of such knowledge. With the development of such technology, the farmer will be able to treat the plant in time and not have to suffer major damage.

Identifying fruits also minimizes the time and labor required for fruit sorting in supermarkets and eliminates the necessity for extensive handling of farm produce across the supply chain. As new strains of viruses and bacterial pose global health concerns, automating processes could mitigate unnecessary and improper handling of farm produce by non-experts. This automation has the potential to address issues like food poisoning, among others, stemming from mishandling of produce (Darwish, 2020).

There are many problems in Nepal's agricultural system, farmers are not motivated. Similarly, various research questions are hypothesized as this study progresses. Isn't there a decrease in production in the agriculture sector due to the lack of modernization or smart methods? Isn't there a decrease in the use of such technologies because of the fear that more problems may arise with the use of smart methods?

The main goal of this paper is to use machine learning of some samples to detect Convolutional Neural Network method. Using this method, the accuracy percentage is to be tested and we are giving confidence that this work can be done by the accuracy percentage. Such activities have not been done in our country. It is hoped from this research paper that anyone who studies this paper in our place and country will get knowledge and help to make new plans even if it is little.

The concept of smart agriculture system is also an opportunity in the agricultural system of the country. As per the United States Farm Bureau, by 2050, we'll need to increase food production by 70% to sustain the growing global population. This necessitates not only expanding the number of farms but also enhancing productivity per unit of farmland. Therefore, it's crucial to adopt more efficient farming practices and analyze our farms to devise strategies that improve yields (Darwish, 2020).



The main objective of this study is to measure the accuracy by using machine learning in the agricultural system and to give more confidence and motivation to the farmers and to add the main revolution in the agricultural system. So that agricultural production is not allowed to decrease and agricultural crisis is not allowed to occur in the future.

To simplify this study, specific research boundaries have been established. This study will involve identifying fruits using image and differentiating diseases through photos of fruit seeds and symptoms. Rather than utilizing all plant components, only fruit leaf image will be employed for disease identification. Only CNN model is used for the image segmentation. In the same way, detection of 3 types of diseases will be studied and average accuracy will be studied. I used only MobileNet algorithm to classify disease of fruits.

Currently, image processing methods have streamlined numerous tasks, including applications in agriculture. Despite challenges within the agricultural system, abandoning this professional isn't an option, especially considering the production of fruits. Identifying fruit species and their diseases isn't straightforward. This research aims to address these issues, offering solutions. using photos, it will diagnose various fruit diseases and detect issues that are imperceptible to the naked eye. These images will enable even non-expert to recognize fruit problems and administer necessary treatments.

### **Literature Review**

Reviewing papers present its own challenges. To enhance the credibility and effectiveness of this research, a thorough examination of publications by various researchers was conducted. The following provides a summary of the papers analyzed for this review.

This technology employs essential infrastructure to enhance the quality and quantity of agricultural yields. Its focus lies in developing an effective method for detecting diseases to tomato plants using advanced computer vision technology. The study aimed to identify three tomato plant disease, pharma rot, leaf miner, and target spot using a dataset containing both healthy and diseased leaves. By training a deep convolutional neural network with this data, the study sought to enable disease identification. Implementing a convolutional Neural Network, the system successfully recognized the presence of tomato diseases in monitored plants. The automated image capturing system achieved an impressive accuracy of 91.67% in identifying disease in tomato plant leaves.(De Luna et al., 2019).

Agriculture holds considerable importance in our daily lives. Farmers commonly face challenges in pinpointing leaf diseases, resulting in reduced crop yields. Yet, leveraging video and images of leaves can provide agricultural scientists with enhanced insights into plants, aiding in disease resolution. This paper centers on employing image processing methods for disease detection in plants. Researchers utilized a dataset featuring 5000 images of both healthy and diseased plant leaves, applying semi supervised techniques to discern crop types and identify four disease classes(Suma et al., 2019).



The research paper explores five distinct diseases that impact apple leaves: area leaf spot, Brown spot, Mosaic, Grey spot, and Rust. Employing deep learning techniques, the paper aims to improve the efficacy of convolutional neural networks in identifying these diseases. Researchers trained the INAR-model using 26,377 images of apple leaf disease in the testing dataset and successfully detected all five common apple leaf diseases. Test results illustrate that the INAR-SSD model achieves a detection rate of 78.80%, operating at a high detection speed of 23.15 FPS. This indicates that the innovative INAR-SSD model serves as a remarkably efficiently tool for early apple leaf disease diagnosis, enabling real-time detection with greater precision and speed compared to prior methodologies (Jiang et al., 2019).

The scientists trained the convolutional neural network model with an extensive open dataset encompassing over 39 classes of plant leaf diseases and background images. The dataset incorporated six diverse data augmentation methods, such as gamma correction, image flipping, principal component analysis color augmentation, rotation, noise injection, and scaling. They observed a substantial enhancement in the model's performance due to data augmentation. Employing various ranges of epochs, batch sizes, and dropout rates during training, they conducted comparisons with transfer learning approaches. Their model exhibited superior accuracy results in validation data compared to these transfer learning methods. Through simulations, their proposed model achieved an impressive classification accuracy of 96.46%. Additionally, the CNN model's accuracy surpassed that of the transfer learning approaches (Geetharamani & J., 2019).

This research paper includes agricultural images of vegetables, fruits, crops, and flowers depicting leaf diseases specific to each agricultural product types. These diseases manifest in different product components like roots, seeds, and leaves, aiding remote disease identification. The research divides into main steps. Firstly, a ring project-based segmentation model is established to analyze leaf image features. Once these features are identified, they are applied to a PNN classifier to detect the presence of diseases. The study tests this method on randomly gathered leaf images from various plants sourced from the web(Soni & Chahar, 2017).

Agriculture has progressed from solely feeding an expanding population to addressing concerns regarding crop quality, often compromised by diseases. Detecting leaf disease stands as a pivotal step in curbing agricultural losses. This research aims to develop a software solution for automatic identification and classification of these diseases. The process encompasses multiple stage image acquisition, pre-processing, segmentation, feature extraction, and classification to facilitate disease detection. Utilizing leaf images, the study employs image processing techniques for the identification and classification of agricultural diseases(Teenu Sahasra et al., 2021).

Plant diseases present a substantial risk to crop yield, impacting both food security and farmers' financial stability. This research paper delves into two methodologies aimed at distinguishing and categorizing healthy and diseased tomato leaves. The initial method involves segregating tomato leaves into healthy and unhealthy categories using the k-nearest neighbor technique. The second approach employs a probabilistic neural network in conjunction with the k-nearest



neighbor method to classify unhealthy tomato leaves. Various features such as GLCM, Gabor, and color and harnessed for effective classification. Experimental results highlight that the fusion approach incorporating a PNN classifier surpasses other methods in terms of accuracy and efficiency. (Devaraj et al., 2019).

Accurately delineating affected areas within plant leaf images is pivotal when employing machine learning to detect plant diseases. To establish a reliable reference, two method typical ROI segmentation and reduced ROI techniques were employed on a renowned Plant Village dataset. Textural and color features were extracted and utilized to construct a linear classifier. Comparative analysis of classification outcomes highlighted the benefits of the suggested method in extracting dissimilarity features. This research facilitates the extraction of more nuanced characterization features crucial for classifying and estimating the severity of plant diseases(Sahithya et al., 2019).

This procedure, also recognized as Region of Interest segmentation, involves distinguishing color-varied symptom lesions from the surrounding green tissue, enabling the extraction of distinctive features. This study introduces an automatic extension of region of interest segmentation to encompass information on symptom progression. This extension broadens the border region to include some healthy tissue by applying color homogeneity thresholding. To establish a benchmark, both the typical ROI segmentation and a reduced ROI approach were employed on the well-known Plant Village dataset. Textural and color feature features were separately extracted to construct a linear classifier. A comparison of the classification outcomes highlighted the benefits of the proposed method in extracting dissimilar features. Through this research, more intricate characterization features can be derived, enhancing the classification and severity estimation of plant diseases.(Sahithya et al., 2019).

Detecting plant diseases early is crucial for maintaining a steady food supply. Swift detection significantly impacts the survival or devastation of crops affected by these diseases. Deep neural networks, a notable aspect of artificial intelligence, have sparked extensive research in image processing and computer vision. This research centers on employing a deep learning-based method to identify diseased plants by analyzing leaf images through transfer learning. Using the NAS-net architecture for convolutional neural networks, the study trains and tests the model using the publicly available Plant Village project dataset, which encompasses diverse plant leaf images exhibiting various infection levels and locations. The model achieves an accuracy rate of 93.82%(Adedoja et al., 2019).

Tomatoes, a widely consumed vegetable offering notable health advantages, necessitate vigilant disease detection to avert substantial losses. This study presents a methodology for precise identification and categorization of common diseases affecting tomato plants such as Bacterial leaf spot, Sectorial leaf spot, and Yellow Leaf Crul-leveraging the comprehensive Plant Village dataset. This dataset encompasses 54,306 image depicting 14 crops afflicted with 26 diseases. The proposed methodology involves three key steps: data acquisition, pre-processing, and classification. Utilizing the Plant Village dataset for image acquisition, the image undergo resizing before being inputted into a modified version of the LeNet deep



learning convolutional neural network model for classification. The proposed system achieves an impressive accuracy rate of 95% (Irmak & Saygili, 2020).

## 2. RESEARCH METHODOLOGY

In this research, disease classification in plant leaves relied on two primary methods: a through literature review and experimentation aimed at addressing specific research inquiries. The literature review served to gather insights and understanding from prior approaches, focusing particularly on CNN and Deep Learning with various classifiers. An experimental approach was adopted to discern apple leaf disease symptoms, providing answers to the defined research questions. To set up the experiment, comprehensive knowledge was gathered from the literature, including insights on experiment design and setup. A crucial aspect of this process involved gathering datasets, pivotal for the experimental outcomes, obtained from publicly available sources.

Python served as the programming language for this research project, chosen for its open source nature and extensive support libraries. Its efficiency and ease of implementation for classification models made it the preferred choice. To optimize training time for plant leaf images and align with study's requirements, tools like Google Colab and Jupyter Notebook were utilized. Several Python libraries such as pandas, numpy, os, cv2, sklearn metrics, confusion matrix, classification report and matplotlib pyplot were employed to support various aspects of the project, from data processing to model evaluation.

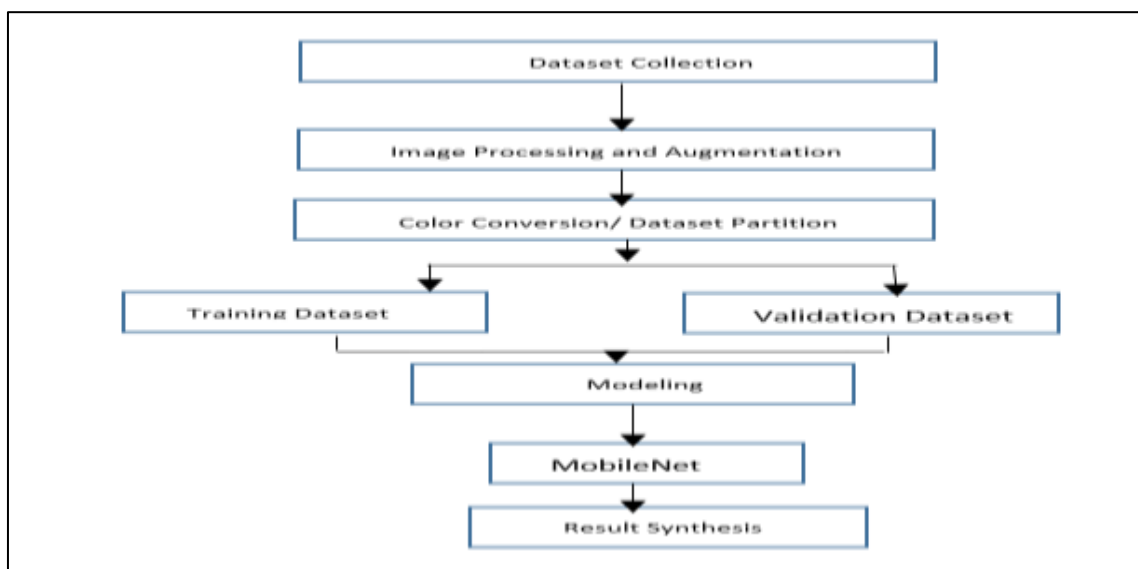


Fig: 1

### Data

Data extracted from Kaggle dataset has been used for this study. The 3642 photographs there have been used for the study. Out of which 28.3 percent are healthy, 5 percent multiple disease, 34 percent rust and 32.5 percent scab disease. Which is presented in figure 4.

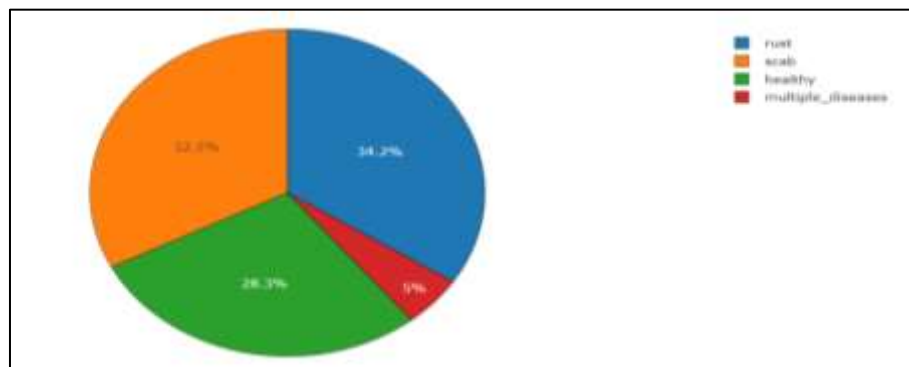


Fig: 2

**Healthy:** Within our dataset, approximately 28.3% of the leaves are classified as healthy. As depicted in figure 3, healthy leaves exhibit no signs of disease and appear completely free of any spots or blemishes.

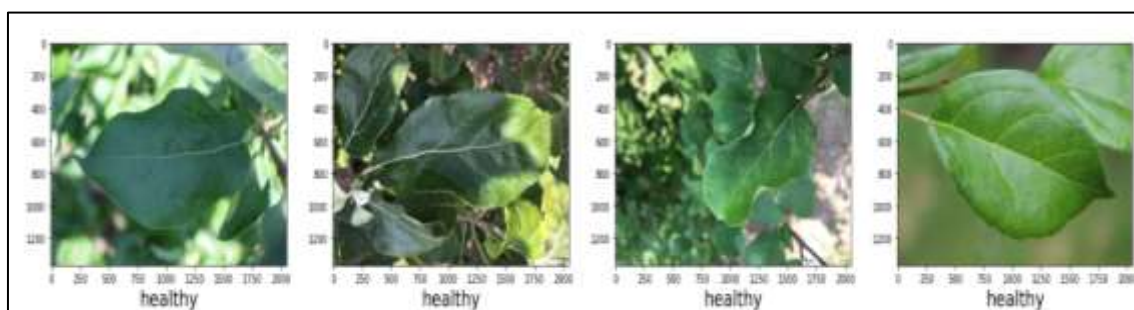


Fig: 3

**Scab:** Scab, commonly triggered by a fungal infection affecting both leaves and fruits, renders the fruit unsuitable for consumption. Within our dataset, approximately 32.5% of images depict apple scab, evident in the presence of brown spots or makes on the leaves.

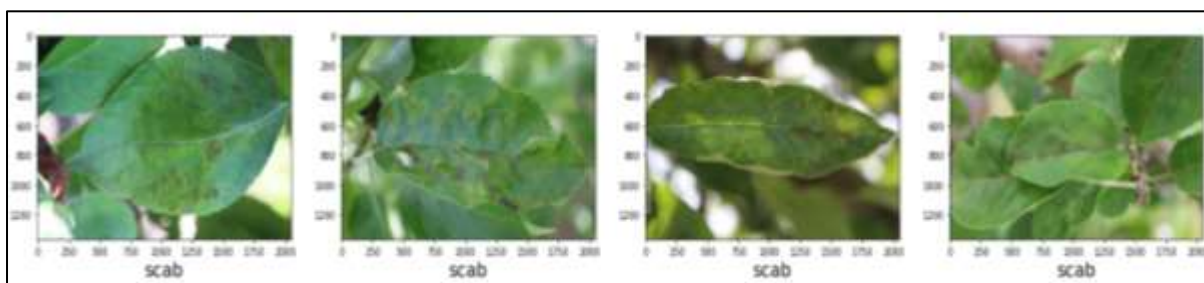


Fig: 4

**Rust:** Observing the apple tree's leaf exhibiting cedar apple rust, it displays dense, yellowish marks. Rust, typically instigated by a distinct fungus called 'rust', is prevalent in plants. In our dataset 34.2% of images showcase this particular type of rust infection, characterized by these visible symptoms.

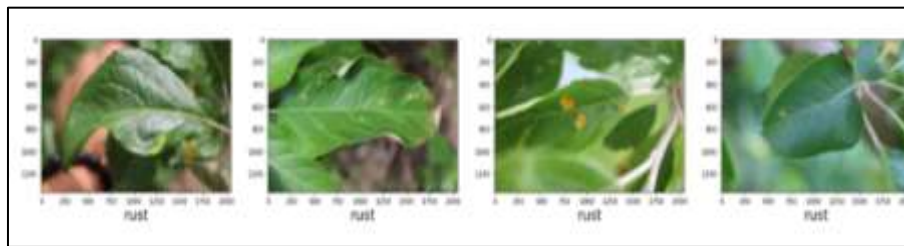


Fig: 5

**Multiple Diseases:** Leaves exhibiting multiple diseases, such as both apple scab with brown spots and cedar apple rust displaying yellow marks simultaneously, are depicted in the provided figure. These leaves are extensively damaged and pose considerable challenges for treatment. Within our dataset, only 5% of the images depict such complex cases where multiple diseases afflict the leaves.

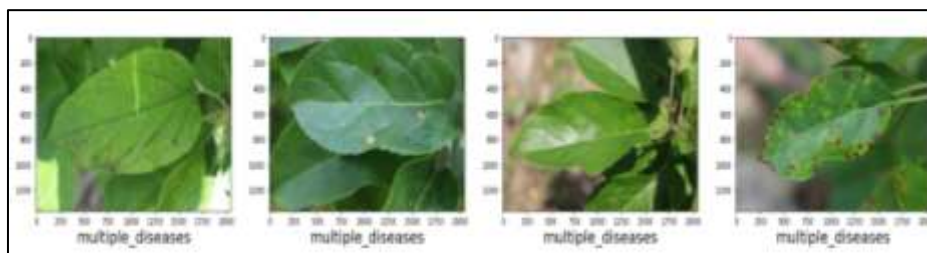


Fig: 6

### Pre-processing / Augmentation

Image augmentation involves enriching our dataset through diverse techniques like rotation and flipping to generate additional images. This process offers dual benefits. Firstly, when dealing with a limited dataset, image augmentation enables expansion without the need for manual collection of new images. Secondly, it enhances model performance by mitigating overfitting in deep learning, thereby contributing to the construction of more robust models.

**Canny Edge Detection:** Edge detection greatly simplifies image analysis and training by reducing observed area, while crucially preserving the structural details within the boundaries. John Canny introduced the influential Canny edge detection method, comprising five steps illustrated in a figure. The outcome is a two-dimensional binary map pinpointing the image edges. In this research, the application of this algorithm to our dataset's images is depicted in the accompanying figure.

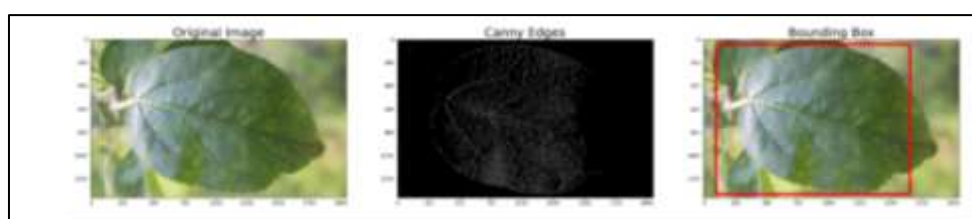


Fig: 7



**Flipping:** Flipping refers to reversing the sequence of rows (Horizontal Flipping) or columns (Vertical Flipping) within the channels of an image.



Fig: 8

**Convolution:** Convolution involves the gradual movement of a 2D matrix or mask across an image, computing the dot product at each window along its path.

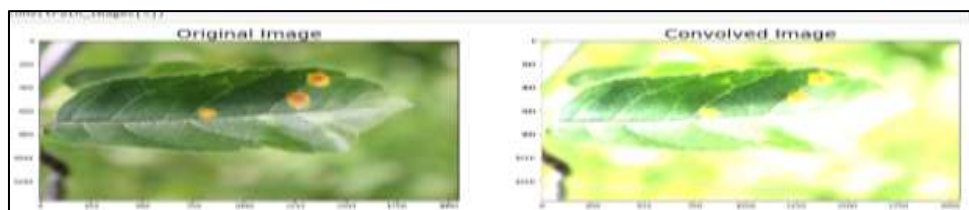


Fig: 9

**Blurring:** Blurring encompasses the act of reducing clarity or sharpness, broadly affecting image details and potentially distorting them in the realm of image analysis.

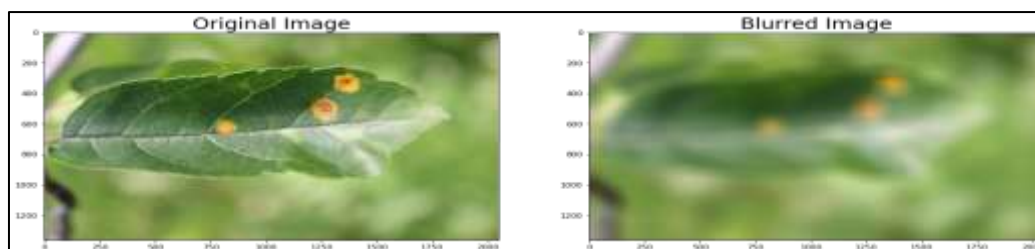


Fig: 10

### Dataset Partition

In our dataset, there are 3642 apple leaf images. These were partitioned into training and validation sets at an 85% to 15% ratio, resulting in 3095 images for training and 547 for validation. To expand the training dataset, we utilized image augmentation. Prior to the split, shuffling was performed to eliminate existing groups or collections within the dataset.

### Modeling

Convolutional Neural Networks, also known as convnets, are widely recognized in computer vision applications, specializing in analyzing visual imagery. This architecture excels in object recognition with in images or videos and finds applications in image and video recognition, neural language processing, among others.

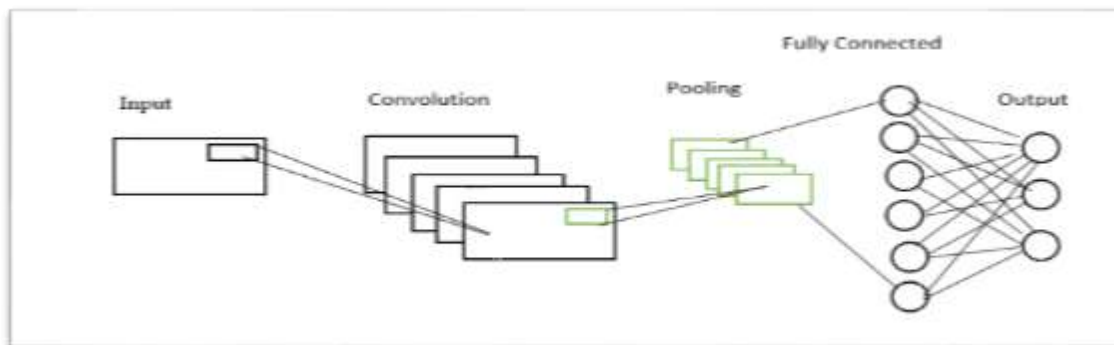


Fig: 11

MobileNet stands out as a popular and user-friendly CNN model. Tailored for mobile applications, it represents TensorFlow's inaugural mobile computer vision model. Employing depth-wise separable convolutions, MobileNet substantially reduces parameters compared to networks using regular convolutional of similar depth. This results in streamlined deep neural networks. The depth-wise separable convolutional involves two operations: depth-wise and point-wise convolutions. As an open-source CNN by Google, MobileNet provides an excellent foundation for training classifiers that exceptionally compact and swift. These models, optimized for various resource constraints, serve multiple purposes, including classification, detection, embedding, and segmentation.

### Data Analysis

In research, data analysis holds pivotal importance. The confusion matrix incorporates crucial terms:

True Positive (TP): Denotes accurate predictions where actual and predicted data align.

True Negative (TN): Represents accurate rejections where both actual and predicted data are false.

False Positive (FP): Signifies instances where false data is predicted as true.

False Negative (FN): Reflects incorrect predictions where true data is mistakenly rejected.

Performance metrics for classification models, such as accuracy, precision, and recall, assess the outcomes of experiments. In plant disease detection, correctly identifying diseases in an image constitutes true positives, accurately rejecting diseases stands as true negatives, incorrectly flagging diseases represents false positive and mistakenly rejecting diseases is classified as false negatives.

Accuracy: The proportion of accurately predicted disease symptoms among all input images.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Precision: It's the ratio of accurately predicted images to the total number of categorized images.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall: It represents the ratio of all positive predictions to the total leaf images, also referred to as sensitivity or the true positive rate.

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1\_Score: The f1\_score, a statistical measure, calculates the harmonic mean of a model's precision and recall, serving as a metric to evaluate model performance.

$$\text{F1\_Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

### 3. EXPERIMENTAL RESULTS

The sample collected in this research was tested with the help of Google Colab and Jupyter Notebook. Photos of apple leaves downloaded from Cagle dataset and two csv files. The files were trained through machine learning. In this processing CNN algorithm and 15 epochs were tested in that work where mobileNet method was used. Training validation loss and training validation accuracy were tested. This diagram is presented as follows.

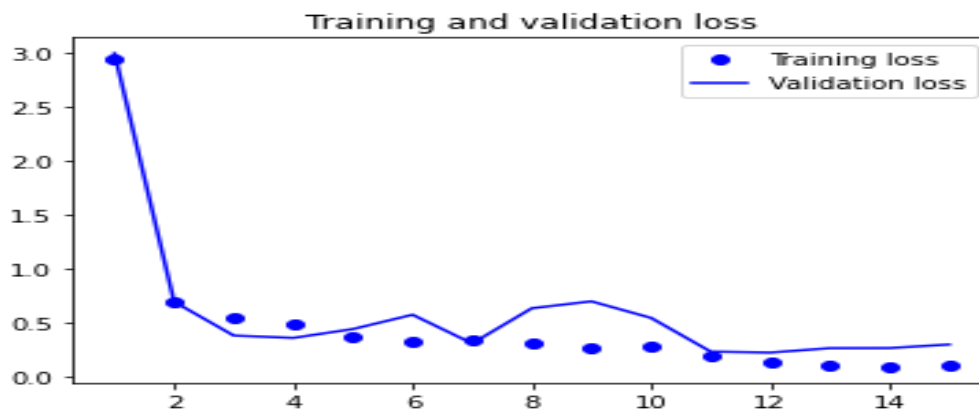


Fig: 12

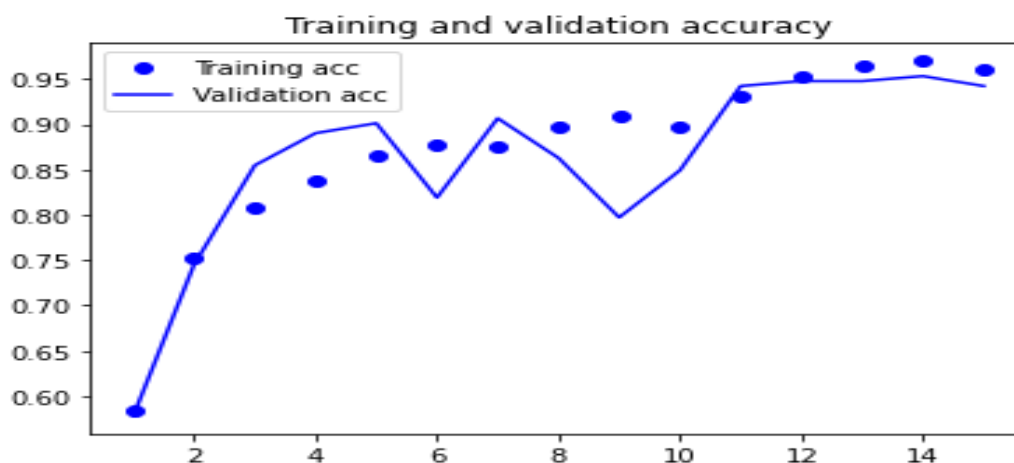


Fig: 13

From this test, the following confusion matrix is formed as follows.

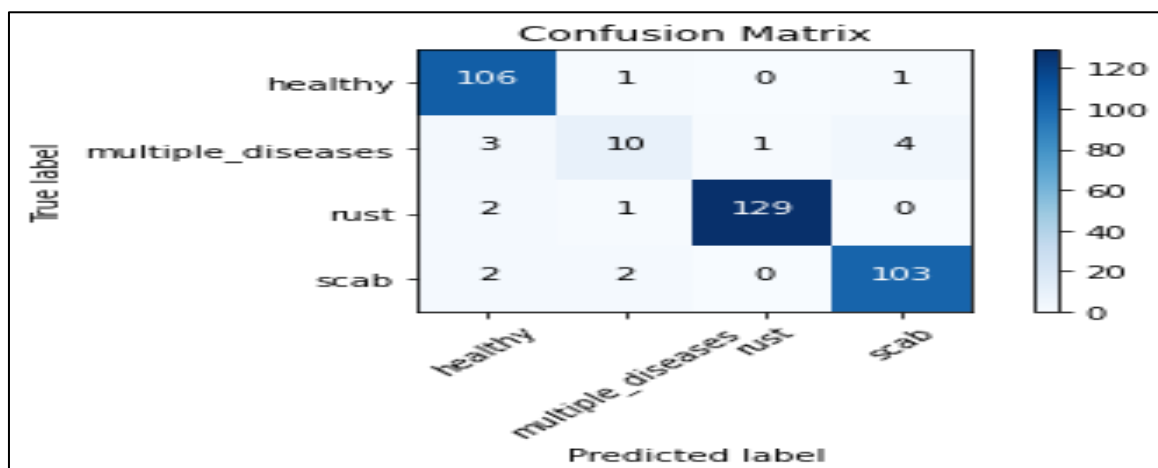


Fig: 14

With the help of confusion metrics, the results are classified and presented in the table.

Table: 1

Types of Diseases	Precision	Recall	F1-Score
Healthy	94%	98%	96%
Multiple Disease	71%	56%	63%
Rust	99%	98%	96%
Scab	95%	96%	96%

The confusion matrix offers a comprehensive overview of how classification models perform and the efforts they encounter. Assessing the validation dataset, the classifiers demonstrated a commendable performance, evident from the notably higher values of true positives and true negatives. This evaluation resulted in calculated accuracy of 95% for the proposed model.

#### 4. CONCLUSION

Machine learning has been used in the preparation of this research paper, which has been prepared with the intention of helping to develop the concept of smart agricultural systems in this area. Is there a decline in the agricultural sector because the agricultural system is based on traditional methods? Likewise, can fruit diseases be researched with photos taken from mobile phones? In this paper, it has been possible to easily experiment such questions of research questions. 95 percent accuracy has been obtained in this test done by mobileNet method of algorithm. 95 percent accuracy is considered a good success. In this way, if the relevant bodies can find, interest and encourage the project that has been successfully tested, then smart designers like us will be able to help with the solutions of other problems easily by successfully testing them.



## 5. REFERENCE

1. Adedoja, A., Owolawi, P. A., & Mapayi, T. (2019). Deep learning based on NASNet for plant disease recognition using leave images. *IcABCD 2019 - 2nd International Conference on Advances in Big Data, Computing and Data Communication Systems*, 1–5. <https://doi.org/10.1109/ICABCD.2019.8851029>
2. Darwish, M. (2020). FRUITS CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK. *Engineering, Construction and Architectural Management*, 25(1), 1–9.
3. De Luna, R. G., Dadios, E. P., & Bandala, A. A. (2019). Automated Image Capturing System for Deep Learning-based Tomato Plant Leaf Disease Detection and Recognition. *IEEE Region 10 Annual International Conference, Proceedings/TENCON, 2018-October(October)*, 1414–1419. <https://doi.org/10.1109/TENCON.2018.8650088>
4. Devaraj, A., Rathan, K., Jaahnavi, S., & Indira, K. (2019). Identification of plant disease using image processing technique. *Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing, ICCSP 2019*, 749–753. <https://doi.org/10.1109/ICCSP.2019.8698056>
5. Geetharamani, G., & J., A. P. (2019). Identification of plant leaf diseases using a nine-layer deep convolutional neural network. *Computers and Electrical Engineering*, 76, 323–338. <https://doi.org/10.1016/j.compeleceng.2019.04.011>
6. Irmak, G., & Saygili, A. (2020). Tomato Leaf Disease Detection and Classification using Convolutional Neural Networks. *Proceedings - 2020 Innovations in Intelligent Systems and Applications Conference, ASYU 2020*, 2–4. <https://doi.org/10.1109/ASYU50717.2020.9259832>
7. Jiang, P., Chen, Y., Liu, B., He, D., & Liang, C. (2019). Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks. *IEEE Access*, 7, 59069–59080. <https://doi.org/10.1109/ACCESS.2019.2914929>
8. Sahithya, V., Saivihari, B., Vamsi, V. K., Reddy, P. S., & Balamurugan, K. (2019). GUI based detection of unhealthy leaves using image processing techniques. *Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing, ICCSP 2019*, 818–822. <https://doi.org/10.1109/ICCSP.2019.8697946>
9. Soni, P., & Chahar, R. (2017). A segmentation improved robust PNN model for disease identification in different leaf images. *1st IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems, ICPEICES 2016*. <https://doi.org/10.1109/ICPEICES.2016.7853301>
10. Suma, V., Shetty, R. A., Tated, R. F., Rohan, S., & Pujar, T. S. (2019). CNN based Leaf Disease Identification and Remedy Recommendation System. *Proceedings of the 3rd International Conference on Electronics and Communication and Aerospace Technology, ICECA 2019*, 395–399. <https://doi.org/10.1109/ICECA.2019.8821872>
11. Teenu Sahasra, M., Sai Kumari, S., Sai Meghana, S., & Devi, P. R. (2021). Leaf Disease Detection Using Deep Learning. *Proceedings of the 2nd International Conference on Electronics and Sustainable Communication Systems, ICESC 2021*, 1797–1804. <https://doi.org/10.1109/ICESC51422.2021.9532909>