Comparative analysis of different types of CNN methodologies for prediction of Cotton Blast Disease

Amaan Naseh , Teena Singh, Priyanshu , Dhruv Taneja, Aditya Choudhary, Ankit Gupta, Dr. Khyati Chopra ,University School of Automation of Robotics, Guru Gobind Singh Indraprastha University East Delhi Campus Corresponding author

Abstract

Cotton is the backbone of the textile industry , without it the sector may not survive and growth may no longer be possible . It is the most widely grown crop in the world economically and in India around 24 million farmers cultivate it every year , making it not just a crop , but a livelihood for millions of families . Cotton makes up about 22% of the global fiber market, with around 80% of it used in clothing, 15% in home textiles, and 5% in industrial applications. In India alone, cotton is grown by over 6 million farmers and supports nearly 60 million people through related industries . Even as cotton holds its place as a global leader in production and export , the crop still faces challenges that impact its yield and quality . With rapid increase in Artificial intelligence and machine learning , new and smarter ways are there to detect plant disease , especially through Convolutional Neural Networks (CNNs) . It has shown good accuracy in recognizing disease patterns .The overall length/quantity of the dataset is 3694 images, in which 2956 images belong to training and 738 images belong to testing, to the three categorized classes to avoid overfitting . After setting the optimal parameters we have trained our model through MobileNetV2 .The validation accuracy obtained in MobileNetV2 is 98.13, whereas the validation loss is 0.0168 respectively. To make the final CNN model accessible and user - friendly a web application was developed using HTML , CSS , React and fast api on the backend . At its core , the overall purpose of our research is to take a meaningful step toward smart agriculture by using technology to help farmers to detect cotton diseases early and improve crop health and productivity.

Keywords : Cotton Curl , Convolutional Neural Networks (CNNs) , MobileNetV2

1. Introduction

Cotton (Kapas) , a cycle from farm to fashion markets , which weaves comfort , style us almost daily . Because of its softness and comfortness cotton industry has ruled the world for decades . India is the number one producer and exporter of raw cotton in the world because of that it plays an important part in our economy . Its comfort , strong natural bond and versatile property makes it an ideal fabric which we can use in clothes , homewares , hospitals , hotels , tents

,school and army uniforms and even astronaut’s clothing choices when inside a space shuttle . For some fashion and stationary brands cotton plays a major role in their production , to make a positive difference in people’s minds and many brands focus on pure cotton cloth even Australia is the largest global fashion group . Despite decades of cotton cultivation, the fundamental challenges in farming practices remain largely unaddressed. Issues such as soil degradation, pest resistance, water scarcity, and climate variability remain unresolved . Since cotton is vulnerable to infections by pathogens like Fungi , bacteria , and viruses , it significantly leads to reduced crop yield and reduces the quality of the crop . The reason for yield loss is the type of disease the plant has , time of infection , environment problem , soil problem . The estimated average yield loss due to cotton diseases in India ranges from one-tenth to one-sixth of the total yield. However, during severe outbreaks, losses may exceed half of the yield. Effective steps will help us to reduce the loss. Since government is taking revolutionary steps , including advanced irrigation system, incorporation of machine, chemical and organic farming methodologies to enhance the productivity and so on , but still india is facing the same issue , still we are in the struggling face so the concept of genetic engineering was introduced which used modified genes to inject it into plants directly to provide protection against some diseases. This method is not applicable for all diseases and requires a lot of money and resources. Therefore, many research scholars implemented the concept of “Artificial intelligence” (AI) for building up smart agricultural systems to reduce dependence on chemical and other less efficient methods. We have used the concept of Convolutional Neural Networks (CNN) with different transfer learning methodologies to develop an AI model which can predict cotton curl disease based on the image input of the user.

2. History of Cotton

India held a global monopoly in manufacturing cotton textiles from the ancient era and had been exporting cotton to other countries since ancient times . The production gained its pace during the 16th century to 18th century . The production was done majorly in the Bengal region and was exported to Europe through the Bay of Bengal . As the cheap cotton was introduced to Europe by the British East India company and various other companies . After the partition of India the cotton growing states were divided into 3 major zones: north , central and south zone in 1958 .To improve cotton production cotton research institutes were set up in 1976 . The govt introduced the cotton to MSP minimum support price so that the farmers can get minimum pay for the yield and in the late 20th century hybrid varieties were introduced which were flexible to the region and climate required .

3. Varieties of Cotton

In India basically there are three types of cotton that are grown :

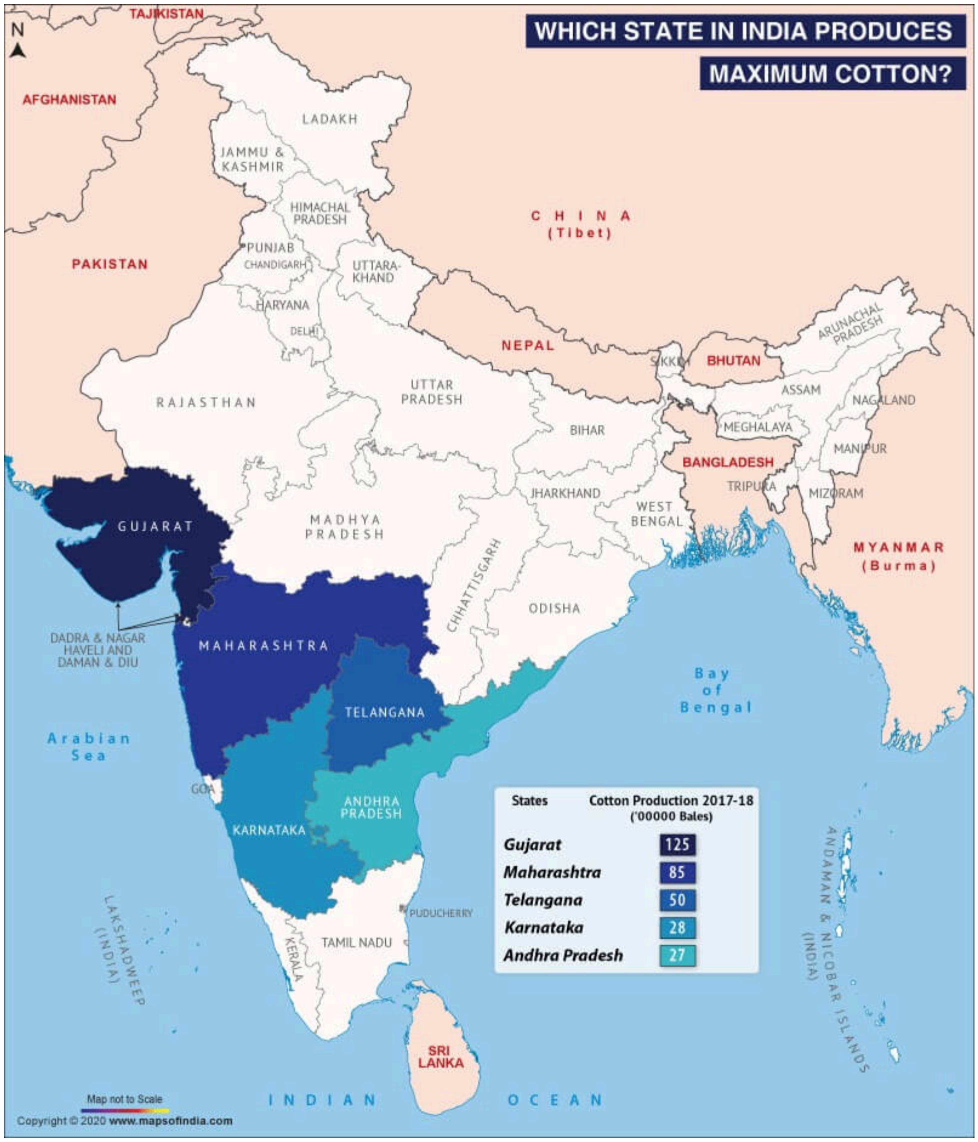
- Gossypium hirsutum : Due to its adaptable capabilities of climatic changes and long stable fibres and high production ,This is the most grown cotton variety in India .
- Gossypium arboreum : This is also known as Asiatic cotton . It's pests resistance quality , diseases and drought conditions, making it well-suited for low-input and sustainable farming systems . However its productivity is low as compared to Gossypium hirsutum which make it inactive in commercial practices .
- Gossypium barbadense : It is known for its extra long staple fibres . It has great genetic traits as it provides protection against pests like the cotton bollworm . They generally grow in warm climates .It has a lower yield but offers higher fiber quality, making it ideal for premium textiles .

4. Major State of Cotton Production

In India there are ten major cotton growing states divided into three zones :

- North Zone : It includes Rajasthan , Punjab and Haryana .
- Central Zone : It includes Madhya Pradesh, Maharashtra and Gujarat.
- South Zone : It includes Andhra Pradesh, Telangana, Karnataka and Tamil Nadu.
- East Zone : It includes states like Orissa .

On a smaller scale Cotton is Grown in many states like Uttar Pradesh , West Bengal , Tripura and many .



5. Area, Production & Yield of Cotton in India

India currently cultivates around 130.61 lakh hectares of area for cotton with a yield of 447 kg/ha . The central zone , which includes Gujarat , Maharashtra and Madhya Pradesh was the largest cotton producing region in India 2023

Cotton Year	Cotton Acreage (in lakh hectares)	Cotton Yield (Lint in Kg/ha)
2017-18	125.86	500
2018-19	126.14	449
2019-20	134.77	460
2020-21	132.85	451
2021-22	123.71	428
2022-23(P)	130.61	447

6. Key Drivers of Cotton Production

6.1. Climate & Soil Type :

Cotton generally grows in high temperature which ranges between 21°C to 30°C . Temperature below fifteen restrict the seed germination while extreme heat above 40°C can cause boll shielding .Cotton grows well in regions with 50 to 100 cm of annual rainfall. However, excessive rainfall or drought can negatively affect growth and yield. Cotton prefers black cotton soil (Regur soil), loamy soil, or alluvial soil with high fertility and moisture-holding capacity . Highly acidic or saline soils reduce productivity and a PH of 6.0 - 7.0 is ideal for the growth of plants . In north india like punjab cotton is grown in alluvial soil . Soil , rich in potassium , Nitrogen , phosphorus and organic matter promote healthy plant growth .

6.2. Crop Season & preparation of Land

Cotton is a kharif crop whose growth varies depending on region and climate . In North India Haryana , Punjab , Uttar Pradesh , Gujarat , it is sown between April and May and its harvesting time is September to October , whereas in south India like karnataka , Andhra Pradesh , kerala , Tamil Nadu seeds are sown in August and September . Ensuring fertile and well-structured soil is essential for strong root development and optimal nutrient uptake. Protecting soil from moisture , weeds , any virus and supplying a good amount of nutrients to them , this will minimize the soil -borne diseases .

6.3. Fungicides & Pesticides Used:

As in India , cotton farming is susceptible to pest , insects , and fungal infection, so use of fungicides and pesticides are crucial for the production of good quality crops.

Pesticides Used:

Pest	**Pesticides**
Red Cotton Bugs, Dusky Cotton Bugs	Malathion, Lambda-Cyhalothrin
Thrips & Mites	Abamectin, spyromesifin
Bollworms	Chlorantraniliprole, Emamectin Benzoate, Spinosad
Aphids, Jassids, Whiteflies	Thiamethoxam, Dimethoate

Fungicides Used:

Fungal Disease	Fungicides
Alternaria Leaf Spot	Mancozeb, Chlorothalonil
Anthracnose & Bacterial Blight	Copper Oxychloride, Zineb
Root Rot & Wilt (Fusarium & Verticillium Wilt)	Carbendazim, Trichoderma-based biofungicides

6.4 Fertilizers Used :

In India, there is availability of various hybrid seeds to grow cotton in different regions because of which fertilizer becomes a crucial part as the soil lacks the required nutrients .

Nutrients	Fertilizers
Nitrogen (N)	Urea, Ammonium Sulphate
Potassium (K)	Muriate of Potash (MOP), Sulfate of Potash (SOP)

Sulfur (S)	Ammonium Sulphate
Phosphorus (P)	Single Super Phosphate (SSP), Diammonium Phosphate (DAP), Rock Phosphate
Zinc (Zn)	Zinc Sulphate ($ZnSO_4$)

7. Government Initiatives to Enhance Cotton Productivity

The cotton development Programme is launched by the Department of Agriculture & Farms Welfare with operations in 15 major cotton growing states across India .
Its motive is to enhance cotton production and productivity of the major state with respect to total cultivated land and the yield of cotton .
Cotton is introduced to the Minimum support price MSP so that the farmers can get minimum pay to fetch the 1.5 times cost of production; this secures the economic interest of cotton farmers while also ensuring the cotton supply to the textile industry .

Cotton corporation of India (CCI) controls the MSP operations , its function becomes operational when the market prices of fair average quality grade seed cotton fall below MSP rates.and ensures fair pricing for cotton .
The Indian Government, through the Ministry of Agriculture & Farmers Welfare, has sanctioned a Large-Scale Demonstrations Project.This project is part of the broader National Food Security Mission (NFSM) checks whether the best practices are going on to enhance the production of cotton . The projects aims to develop and use HIgh density planting system approach

8. Cotton Leaf Curl Disease

Cotton Leaf Curl Disease (CLCuD) is a viral disease which impacts the production of the cotton worldwide . It is caused by the Cotton Leaf Curl Virus (CLCuV), a member of the Geminiviridae family, and is primarily transmitted by the whitefly (*Bemisia tabaci*). Its symptoms generally occur within 2-3 weeks of infection and it varies from environmental conditions . Affected cotton plants give smaller bolls due to which the quality of fibre decreases and its results in yield loss . Cotton Leaf Curl Disease has been responsible for major cotton production declines in India, Pakistan, and other cotton-growing regions.

8.1. Transmission & Spread

- Whiteflies (*Bemisia tabaci*) are the vectors which transmits virus from an infected plant to healthy plant .
- Due to climate change , heavy use of pesticides , chemical fertilizers and poor pest management it lead to increase in chance of disease spread
- The disease generally spread in humid and warm climate .

8.2. Symptoms

Plant infected by this virus become weak and fragile day by day . There are many factors responsible for this disease like soil type , environmental changes , plant age at infection , pesticides and fertilizers used .

- Leaf Curling and Distortion happens in the leaf , the newly leaf exhibits upward or downward curling . Due to Severe infections leaves appear like cup - shaped
- The main vein and lateral vein of the leaf become swollen and raised
- Small, wart-like enation occur downside the leave and it become a web like projection
- Due to this virus plants show poor growth and development as they do not get proper nutrients .
- Due to lack of nutrients , no growth of buds and flowers occur
- The infected plant becomes smaller day by day and becomes under developed.
- Chlorotic areas occur between the veins.

8.3. Economic impact

The disease reduces both yield and fiber quality, leading to financial losses for farmers, industries, and national economies. The reduction in the yield directly impact farmer's income and increases their dependencies on chemical fertilizers , pesticides , insecticides . The additional expenditure on pest control significantly raises production costs, making cotton farming less profitable. The cotton produced by infected crop is generally weak and have poor quality fibre which makes it completely useless for the textile industry . This results in financial losses for ginning factories, mills, and garment manufacturers, leading to higher production costs and reduced competitiveness in the global market. Many textile industry uses high quality of cotton and due to cotton curl disease they need to export this on a very high rate , which lead to sudden increase in the price of the product and it would lead to a great impact on the international trade market . The decline in production lead to unemployment issues as a huge number of labour works in this market .

8.4. Manage and Control Strategies

As the problem is very high so effective management of cotton curl disease should be implemented to escape from the loss . Firstly the disease is transmitted by a vector named whiteflies , so controlling them should be primary goal .

- Developing and using resistant or tolerant cotton variety
- Using Biotechnology and DNA recombinant technology to make an pest resistant seed
- By controlling vectors through biological control , chemical control and botanical pesticides.
- Monitoring the plant at an early stage so that we can detect the virus and can take major steps .
- Growing non-host crops like cereals between cotton seasons breaks the disease cycle.

8.5. Disease Progression

- **Early-stage** : Within two three weeks of virus mild curling and vein thickening on leaf occurs .The affected leaves start showing curls upwards and downwards . This gradually worsen the condition of the plant , supply of nutrients and water got disrupted which lead to progress in the disease
- **Mid-stage** : The symptoms become more severe within three four weeks of infection .The leaves become extremely twisted and get deformed .Due to Vein thickening , there is huge restriction in nutrient flow . At this stage , the disease significantly impacts the growth and production .
- **Late-stage** : This is the last stage of the virus . Plants exhibit its stunted growth, failing to reach their normal height due to severe nutrient disruption. The disease effect boll development , leading to complete crop failure

9. Convolutional Neural Network & its terminologies

Python, along with the TensorFlow and Keras frameworks, has emerged as a leading language for developing deep learning models, especially Convolutional Neural Networks (CNNs). TensorFlow offers a robust and flexible platform for building and deploying complex neural network architectures, providing fine control over computational processes and model scalability. Keras, now integrated within TensorFlow, simplifies the design of CNNs by offering a high-level interface with pre-built components like layers, optimizers, and loss functions. This combination allows for efficient model prototyping and experimentation, making it easier to implement CNNs for tasks such as image recognition and object detection. Python's concise and readable syntax further enhances the overall development experience, ensuring both ease of use and advanced functionality in deep learning applications. Convolutional Neural Networks (CNNs) consist of multiple layers designed for feature extraction and classification. They typically include convolutional layers, pooling layers, and fully connected layers. Convolutional layers extract spatial hierarchies of features from input images, while pooling layers reduce spatial dimensions. The fully connected layers at the end of the network make final predictions, making CNNs highly effective for image-related tasks such as classification and detection. Transfer learning leverages pre-trained models like MobileNetV2, which is optimized for mobile and edge devices, to speed up model training. MobileNetV2 uses depthwise separable convolutions to reduce model size while maintaining performance, making it a suitable choice for tasks with limited computational resources.

Hyper-parameters of CNN:

- **Batch size :** Refers to the number of training samples used in one forward/backward pass of the model. A larger batch size accelerates training but requires more memory.
- **Epochs:** The number of times the entire dataset passes through the network during training. More epochs can improve learning but may also lead to overfitting.
- **Activation function:** Functions like softmax that introduce non-linearity to the model, helping it learn complex patterns.
- **Loss function:** Measures the difference between predicted and actual values. Common loss functions include categorical crossentropy for classification tasks.
- **Optimizer:** Algorithms like Adaptive Moment Estimation (Adam) adjust weights to minimize the loss function during training.
- **Metrics:** Include accuracy (correct predictions), loss (error), confusion matrix (class distribution), classification report (precision, recall, F1-score), and Receiver Operating Characteristic/ROC curve & Area under the Curve/AUC (performance across various thresholds). The respective formulae for evaluating a convolutional neural networks model is given below:

Here TP means True Positive, FP means False Positive, TN means True Negative and FN means False Negative, respectively. The trained model evaluated using the metrics may have the following conditions, which are tackled by us in our methodology:

- **Over-fitting:** Occurs when a model learns the training data too well, capturing noise and details that don't generalize to new data. It results in high accuracy on the training set but poor performance on the test set.
- **Under-fitting:** Happens when a model is too simple to capture the underlying patterns in the data, resulting in poor performance on both the training and test sets.

9. Methodology

9.1. Working Equipment & Versions

The machine used for training the CNN model is a MacBook with an Apple M3 chip, equipped with an 8-core CPU, 8-core GPU, and a 16-core Neural Engine for high-performance computing. It features 16GB of unified memory and 256GB SSD storage, ensuring smooth multitasking and fast data access. The python version 3.7.8 is used during training, having the python libraries tensorflow (v2.11.0),keras (v2.11.0) & Flask (v2.2.5), respectively. The specification of machine and specific software versions is highly crucial in machine learning for stability & accuracy of models

9.2. Dataset Collection

The dataset is collected from Kaggle https://www.kaggle.com. Kaggle is an online platform used by data scientists and machine learning enthusiasts . Authentic high-quality datasets compatible with ML algorithms are available on each domain , anyone can access that data , can communicate with anyone , can learn new algorithms . It becomes more attractive as you can publish your dataset here , and can participate in many international competitions . For our research we have collected dataset from https://www.kaggle.com/datasets/fatimashakeel1/cottonleafcurl-dataset .
The dataset is based on cotton_Curl_leaf composed of test and train classes . Further we have categorized the data into three major domains “Healthy Leaf” , “Invalid Data” , “Cotton Curl leaf” . “Healthy Leaf” consists of all leaves that are not suffering from this disease . “Invalid Data” consists of all the inappropriate images (like blurred images , low quality images , duplicates , pen , notebook etc..) that can cause disruption in the dataset . We have used a common splitting technique to split our dataset into training and testing subsets in the ratio of 80:20 percentages, where 80% images data belongs to training and 20% images data belongs to testing, respectively. The overall length/quantity of the dataset is 4563 images, in which 3650

images belong to training and 913 images belong to testing, to the three categorized classes. In training dataset, we have used 1200 images of Healthy Cotton leaf, 1250 invalid images and 1200 images of diseased Cotton leaf (having cotton curl) whereas in testing dataset, we have used 300 images of healthy leaf, 313 invalid images and 300 images of diseased cotton leaf (having cotton blast) as shown in the **Table 3.0.**

Class	Train	Test
Healthy	960	241
Invalid	800	200
Leaf Blast	1196	297
Total = 3694	Total = 2956	Total = 738

Table 3.0.

8.3. Dataset Pre-Processing

After collecting and segregating the dataset into train and test directories, we have done pre-processing to enhance the quality of our dataset. We have first renamed the image names in the test directory of all the three classes, viz., Healthy, Invalid and Rice Blast with their respective class names for our pre-planned evaluation in further steps. After renaming test images, we have resized all images of the whole dataset into the desirable input size, i.e., (224, 224), as shown in Figure 5(a) and 5(b), for the transfer learning methodologies opted by to enhance its training. We have saved the resized images into another directory known as output having train and test directories inside it. So, after pre-processing, we have created two sets of dataset, one original and one resized for our training. We have also applied data augmentation by rescaling the data through image data generator function for enhancing the length of the dataset by normalizing the pixels for stable training as well as passing variations of the same image through augmentation.

5(a) Image before pre - processing

5(b) Image after pre - processing

Figure 5. Dataset images, leaf having cotton blast (a) before pre-processing, and (b) after pre-processing

8.4. Deep Learning Architectures/Transfer Learning Methodologies

In this research, we have used MobileNetV2 architecture, a convolutional neural network known for its efficiency and light weight design . We have performed a comparative analysis on training capabilities by keeping the training parameters constant to evaluate which algorithm is best .We have tried to select the best set of CNN parameters & kept them constant throughout the research for uniformity. The parameters include batch size, epochs, activation function, loss function, optimizer, metrics, early stopping & image data generator. We have selected the parameters w.r.t. to small length of our dataset and evaluation aspects which primarily aims to avoid over-fitting.

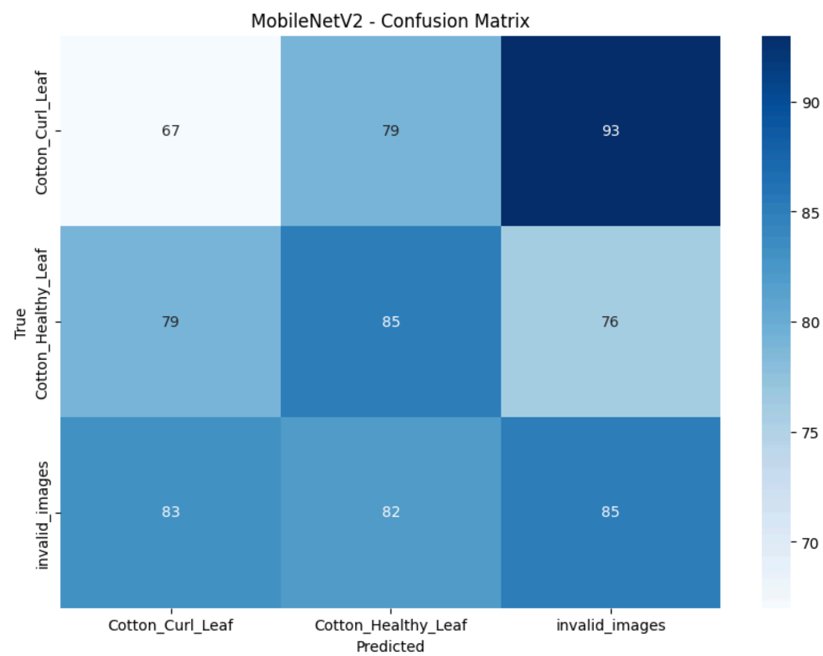
8.5. Model training with MobileNetV2

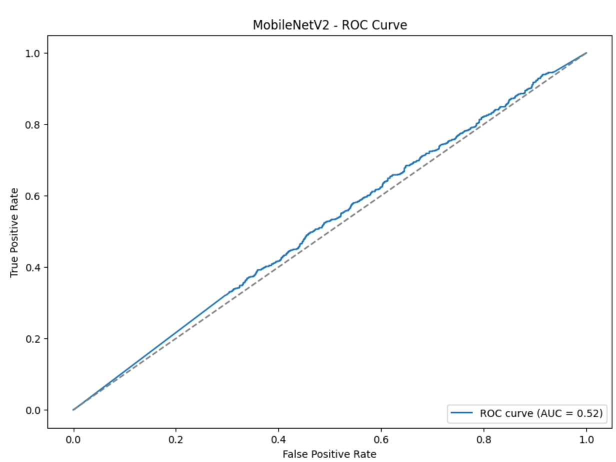
After setting the optimal parameters we have trained our model through MobileNetV2.
Firstly through a Python script , which resizes images directly during training to desirable input and target sizes of (224, 224, 3). This generated output files including model.h5 files, accuracy and loss curves, ROC curves, Confusion Matrices, Classification Reports. The test accuracy obtained in MobileNetV2 is 98.13, whereas the test loss is 0.27 respectively.

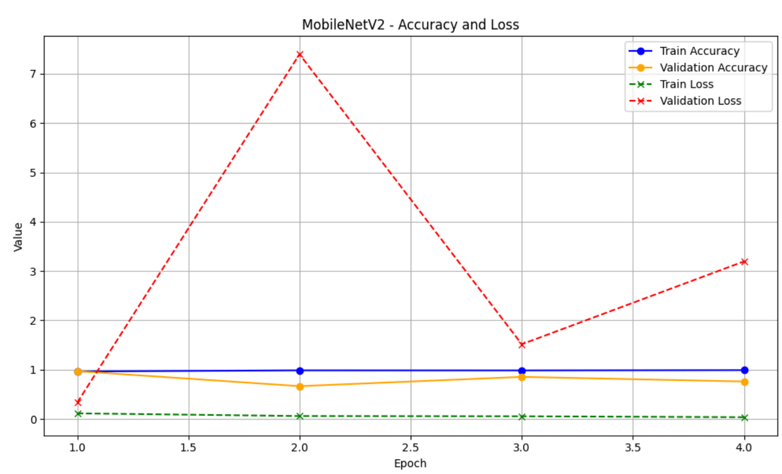
We have also tested this with our 640 images (in test dataset) belonging to the three classes by passing all of them one by one through a python script and saving the prediction results in an excel workbook. This is why we have renamed the test images earlier in order to reflect the prediction along with their file names to evaluate whether the prediction is correct or not .

8.6. Final Prediction

Based on the performance metrics , MobileNetV2 emerged as the best performing architecture . To maintain the model consistency , we kept the model parameters unchanged and only replaced the original training dataset with a preprocessed version containing the same number of images . After training the accuracy was 97.16 , the validation accuracy was 91.86% and training loss to be 0.14 and the validation loss to be 0.1416 which meet our theoretical expectations . We test the model with test images which belong to three classes: Healthy , Invalid and the one which are suffering from cotton curl disease . While testing , 241 images were classified as healthy test images , for invalid images 200 images are classified as correctly invalid and 297 images are suffering from cotton curl disease


MobileNetV2 - Confusion Matrix
Cotton_Curl_Leaf
Cotton_Healthy_Leaf
invalid_images
True
Predicted
67
79
93
79
85
76
83
82
85
90
85
80
75
70


MobileNetV2 - ROC Curve
True Positive Rate
False Positive Rate
ROC curve (AUC = 0.52)


MobileNetV2 - Accuracy and Loss
Train Accuracy
Validation Accuracy
Train Loss
Validation Loss
Value
Epoch

9.References

9.1 https://texmin.gov.in/sites/default/files/Note%20on%20Cotton%20Sector.pdf
9.2 https://www.nfsm.gov.in/BriefNote/BN_Cotton.pdf
9.3 https://sathee.prutor.ai/article/social-science/cotton-producing-states-in-india/
9.4https://www.texmin.nic.in/sites/default/files/STATEWISE%20YEAR-WISE%20AREA%2C%20PRODUCTION%20AND%20YIELD%20IN%20INDIA%20AS%20PER%20COCPC_0.pdf
9.5 https://www.cotton.org/pubs/cottoncounts/story/
9.6 https://www.cotton.org/tech/ace/soil-fertility.cfm
https://en.wikipedia.org/wiki/Cotton_leaf_curl_virus#:~:text=In%20Asia%20and%20Africa%20the,and%20yield%20is%20reduced%20drastically.