National
College *of*
Ireland

An Implementation of Transfer Learning & Deep Learning
Techniques to Detect Tomato Leaf Diseases

MSc Research Project

MSc in Data Analytics

Manikanta Dinesh Gudivada

Student ID: x18191851

School of Computing

National College of Ireland

Supervisor: Vladimir Milosavljevic

Student Name:	Gudivada Manikanta Dinesh
Student ID:	x18191851
Programme:	MSc Data Analytics
Year:	2019 - 2020
Module:	Research Project
Supervisor:	Vladimir Milosavljevic
Submission Due Date:	17th August 2020
Project Title:	An Implementation of Transfer Learning & Deep Learning Techniques to Detect Tomato Leaf Diseases
Word Count:	9480
Page Count:	25

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

I agree to an electronic copy of my thesis being made publicly available on TRAP the National College of Ireland's Institutional Repository for consultation.

Signature:	G.M. DINESH
Date:	17th August 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	Q
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	Q
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	Q

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

An Implementation of Transfer Learning & Deep Learning Techniques to Detect Tomato Leaf Diseases

## Manikanta Dinesh Gudivada

X18191851

Abstract

From the past few decades, Agriculture is playing an indispensable role in the survival of humans. In agriculture, the most cultivated crops are potato and sweet potato. According to the statistics of FAO organization, the India ranks 3 for their immense production in tomato (**Kaur and Bhatia, 2019**). Across the world, the tomato crop is enormously popular for their growth as in every kitchen people use tomatoes massively. However, the growth is high there is a huge loss to the farmers due to an increase in massive Tomato Leaf Diseases. So, to detect leaf diseases the majorly used techniques in this paper are Deep Learning techniques where the image data can be trained and modelled perfectly using various Neural Network models. By this research study, every individual can gain knowledge over different methodologies involved in Deep Learning, to encounter those methodologies, this research had experimented 4 models which are Le Net, Dense Net-121, Mobile Net and CNN. Once after building the model to verify the predicted results and this research consists of different evaluation metrics such as Confusion Matrix, Accuracy, Precision, MSE, Recall and F1-Score. To develop, all these models the coding implementation was done in python using Anaconda Navigator and Google Collab software's. Finally, all the 4 models have gained very good accuracy which is around 85-95%. Out of all the models, Dense-Net model had gained the best accuracy compared to all the models with 97%. All the other models' Le Net, Mobile Net and CNN have gained 86%, 88% and 88% respectively with 50 Epochs. This research had also verified by using 25 and 100 Epochs as well, but Dense Net have given the best results.

Keywords: *Deep Learning, Transfer learning, Leaf Disease Detection, Le-Net, Mobile-Net, Dense-Net-121, Convolutional Neural Network.*

1. Introduction

Tomatoes biologically called as Solanum Lycopersicum provides principle antioxidants such as Vitamin C and A along with beta-carotene. The production of tomato is in an increasing trend where the consumption is high as well the production is also revamped. These tomato crops are being pollinated on the wet soil in the sunlight, water and using different pesticides. Consumption of tomatoes improves the immunity in the body and gives better health. Even for some of the disease's doctors use to prescribe the tomatoes as the ingredients to develop antibodies in the body for various diseases. Majorly the tomato plants were being diseased because of misinterpretation between the farmers by over usage and less usage of pesticides which makes the soil week. In the same way, using so much of water and using very less water also destroys the crop along with these factors the sunlight, rainfall and other climatic

conditions also should be validated at the time of pollinating the crops. In addition to this, if people plant the crop on the soil which is reused for the potato and brinjal plants within a year for four to five times then the soil gets week and have chances in destroying the crop. We can also verify whether a crop is good or bad based on the leaf's which develops a black and brown patch along with the small holes on the leaf (**Puno** *et al.*, **2019**). All these are different types of problems being faced by farmers for many decades. So, the only hope for the farmers is giving a promising solution by providing a simple method in detecting the disease on their own by sorting the problems without going to scientists. So, in providing the promising solutions for the farmers this research is integrated with the Deep Leaning which will help in detecting and classifying the Tomato Leaf Diseases which will make farmers life easier in cropping the tomato plants. **Figure 1** helps in understanding the statistics of the losses occurred in agriculture caused due to different infectious diseases and medicines.
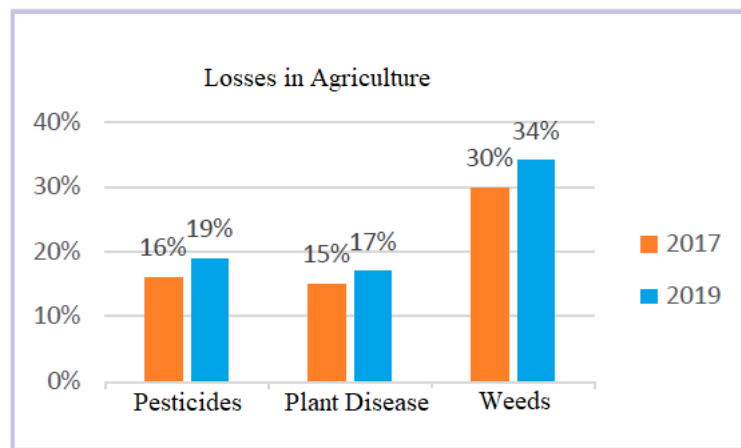
Figure 1: Graph for losses in agriculture due to pesticides, plant disease and weeds (**Bapat** *et al.*, **2020**).

1.1. Background Scope

Enhancing the crop is the major issue facing by the farmers there are few techniques which are coming in the farming. The major scope for this topic is recently a well-known person have started a new ideology in farming where the farmers will be benefited with an immense economy. The ideology which implemented is that he started an Indore farming where the plant growth will be done inside the house by taking primitive measures such as providing the artificial light to the efficiency of crop requirement. In the same way, he also provided the water-based on automatic generating motor systems where the system gives the water of how much the crop is required if once the soil becomes wet the system automatically understands that the crop requires water and provides water automatically which is connected to the motor and power supply is required for this process. So, basing his interest even the research was been developed using new aged science so this research considered tomato leaf disease detection as input which will help people to understand in a better way in predicting the diseases etc.,

1.2. Motivation

The major incident which motivated in considering this research study is that India is having a major part as land. So, the agricultural lands are having very high in demand where the people were producing 70% of the food on their own and India is independent in producing artificial vegetables. In some of the countries, the vegetables were being produced using some chemicals which are hybrid, not much good for health. As the growth is high there will be more losses as crops were being diseased and due to some floods and different effects, the crops were being destroyed for 5-6 decades. The major thing which motivated in choosing this topic as research study is that because of huge loss in crop the economy drops a lot and so that the farmers were committing suicides. According to the global survey, the major part of the suicides in the countries was being happened because of agriculture in last 10 years over 2 lakh people were died. So, this scenario must be changed in future by introducing Nano Technology for protecting the crop and robotic technology should be invented where the machines should have the power to do farming.

1.3. Providing a Brief on Deep Learning Models and Proposed Research Study

The research done by many scientist's states that for performing Image Processing massively used technique is Deep Learning. The important thing is that the Deep Leaning functions based on AI where using AI, works can be done automatically with no human interaction. The models are trained in such a way that they will help in detecting the speeches, images, translation of various languages and help in taking the decision in this way there are many more advantages of using AI. So, coming to this project study, the goal is to detect the diseased leaves based on the images of tomatoes so this research requires the models which can read the images and train the models. In this case, the research had collected the dataset around 17000 images with 9 Tomato diseased leaves classes and one healthy tomato leaves class from the online repository and 4 different DL models had performed by following few steps: Collecting Data, Processing the Data and detecting the diseased leaf. In different models, different convolutional, pooling, dense and activation functions are used according to the input shape of the model which are pre-defined and evaluation metrics had been used for verifying the accuracy results.

1.4. Problem Statement

The Major problems according to this research were facing by farmers in identifying what kind of diseases were being affected by the crops and what are the primitive measures to be taken in safeguarding the crops.

1.5.Research Question

RQ: "How accurately the identification of seed diseases can be enhanced by Deep Learning models in Tomato Plants by adding external factors such as Data Augmentation and EDA analysis?"

Sub RQ: "What are the evaluation results obtained in comparing the (Le-Net, Dense-Net-121, Mobile-Net and CNN) models?"

1.6. Research Objectives

Table 1 provides information about the objectives of the research work followed and worked in achieving these objectives and the table columns. The values were clearly explaining about how the models and evaluation techniques have used and evaluated in this research.

Table1: Research Objectives followed in using Deep Learning for Tomato Leaf Disease

Main Objective	Description	Metrics Involved
Objective 1	Critically reviewing the research papers of Tomato Leaf Detection and identifying the graphs.	--
Objective 2	Exploratory data analysis for knowing the feature insights involved for Tomato Leaf Detection.	--
Objective 3	Implementing the model and Evaluating the Results of Le-Net.	Precision, Recall, F1-Score, R2 Score, MSE, MAE.
Objective 4	Implementing the model and Evaluating the Results of Dense-Net-121.	Precision, Recall, F1-Score, R2 Score, MSE, MAE.
Objective 5	Implementing the model and Evaluating the Results of Mobile-Net.	Precision, Recall, F1-Score, R2 Score, MSE, MAE.
Objective 6	Implementing the model and Evaluating the Results of CNN.	Precision, Recall, F1-Score, R2 Score, MSE, MAE.
Objective 7	Finally, comparing the performed models based on their results obtained.	--

1.7. Format of the Paper

For an easy way of understanding the paper, the organization of the paper is being explained here: Section II gives a review on what previous authors have done in their research studies. Section III provides a brief view of the methodology which this research study implemented for detecting tomato leaf diseases. Section IV gives a specification diagram. Section V helps in understanding the approach of implementation and obtained results clearly explained in section VI. Section VII gives a discussion about the comparisons of the models and the future work is explained in the section VIII.

2. A Critical Review on Tomato Leaf Detection (2015-2020)

2.1. Leaf Diseases

2.1.1. Importance of Leaf Diseases

The plants play an indispensable role in the society by providing all kinds of vitamins, proteins and different types of calcium carbonates which are required for the body. The plants are the human's source of living in a way that they are providing various kinds of vegetables which are been growing by the farmers. In one word if there are no plants there is no survival for this universe. In some of the areas due to lack of water, sunlight and in using different types of pesticides the plants are getting affected by different diseases. So, in this literature survey, I am going to provide information about different authors and scientists how they identified the diseases in plants using Deep Learning and Machine Learning techniques. This survey helps many farmers in a way that they can safeguard their crops and revamp their economy.

2.2. List of various approaches in detecting tomato diseases using Deep Learning and Machine Learning techniques

2.2.1. Convolutional Neural Network (CNN)

In this paper (**Tm** *et al.*, **2018**) have discussed a deep learning model which is Le-Net comes under Convolutional Neural Network (CNN). This model is helpful for image detection. To perform this model authors have considered dataset consisting of 18000 tomato leaf diseased leaves which are classified in 10 classes are taken from plant village data. After collecting the data, they have reshaped all the images into 60 x 60 resolution which helps for performing better results. Later they have performed the model evaluation with basic convolutional, activation and various pooling layers. They have performed a basic CNN model which provides an accuracy of 94.8% after using 30 epochs the accuracy had raised to 99%. After all the model evaluation techniques it is proved that the Le-Net model performed well in detecting the images for leaf diseases. In the same way, in the year 2020, (**Nithish** *et al.*, **2020**) have performed a Deep Learning Model which is ResNet-50. For the better performance of this model, they have implemented a technique of Data Augmentation which increases the data to the 4 times of existing data. The methodology they have mentioned is that they collected the plant village data where 80% is divided for testing and remaining for validation which is the major part for the evaluation of any model. Once the data is divided later step is to increase the data which helps in increasing the accuracy and provide better results. The third step is the major step which they performed in this paper is transfer learning using ResNet-50 in this paper they have changed the final layer of Res-Net model initially it had to be as fully connected layers but using transfer learning they modified to six SoftMax layers and finally, they performed the model. Results were evaluated with the same methods of accuracy and f1-score along with ROC curves. Using the same topic of detecting leaf diseases in tomato plants author (**Agarwal** *et al.*, **2020**) have proposed a

CNN model and did model comparisons with VGG16, MOBILE-NET one of the efficient models and finally INCEPTION model. All these are the transfer learning models which have already performed in the previous papers. So, according to this paper author have collected the same plant village data with only 9 diseases including healthy leaves. After collecting data, they have performed the necessary steps for the image to be trained properly and finally they performed their proposed model. The model comparisons were made using different graphs and presented in the results where the proposed model had given an accuracy of 91.2% which is the highest accuracy compared to all the already trained models concerning the models with 77.2% accuracy for VGG model and 64 and 65 per cent of accuracy for MOBILE-NET and INCEPTION models respectively. Even in this paper, they mentioned about the layers that they have used for the proposed models. Finally, the author concluded that the CNN model is the best compared to all the three models performed in the paper for analysis.

2.2.2. Deep Neural Network (DNN)

This paper majorly mentioned about a DNN technique which is ALEX-NET used for the detection of images and KNN is used for classification purpose. For performing the model, they have used 450 images of segmented data. For detecting the images, they have used 5 convolutional layers. The codes have been executed using MATLAB and the accuracy given by the ALEX NET is 76%. KNN model helped for the classification of leaves and helped the model to get good accuracy. As mentioned in this paper this is the best accuracy, they got compared to all different models mentioned in the paper. Finally, they discussed the increasing the data and considering different leaf plants for performing different more models as a future work **(Batool *et al.*, 2020)**. In extension to the same topic **(Fuentes *et al.*, 2018)** have mentioned that rather than using basic CNN models it is better to use most complicated models such as F-RCNN and R-FCN which gives the extraordinary results. By proving the methods gives best results he proposed a system architecture known as filter bank detection in this they have used three different steps as units such as primary, secondary as well as integration unit each unit have its own importance in detecting the leaves disease by creating the bounding boxes around the images with the help of these boxes in the image the model by not only spotting the disease spot it also can spot the other relevant images such as a person, bike etc kind of things by mentioning what is what. This model gave a new extension to all the Deep Learning models with proper output results. By the end, they mentioned that this works helps in finding the False positives in the very early stage of the model which helps the model to perform in a very gigantic way and provide the proper results. The future work they included in the paper is they can extend this model to all the other following crops which are being suffered from the diseases.

2.2.3. Artificial Neural Network (ANN)

This model is one of the Deep Learning models used for the classification of leaf diseases as the author mentioned in the paper **(Usha *et al.*, 2020)** in the year 2019,

author detected the disease in the various plants which will benefit the farmers a lot and improve the cultivation. According to the author they have considered two different modern models which are known as classifiers one is Neural network this helped the author in detecting the disease and other is K-means which is also a classifier used for the segmentation of images. The steps for the architecture are Image data collection from the plant village containing 4-5 classes of diseases and 20 various cotton folders were also included in the process. After the data collection, they have segmented as the conversion of images from Grayscale to RGB using K-means which helps in improving the accuracy. In the later stages, the feature extraction for the data had been done whereas the matrices for the images were created. Once after the completion of pre-processing steps, the model ANN was created using MATLAB and the results were evaluated using evaluation steps and final accuracy obtained from the model is about 92.5%.

2.2.4. Deep Convolutional Neural Networks

The paper discusses the most effective way of the Deep Learning as far we haven't seen is that the Faster R-CNN and Mask R-CNN together performed for detecting the tomato diseases whereas the FR-CNN is used for detecting what kind of tomatoes and locating where the tomatoes were present. Instead of using this using Mask R-CNN, it gives us the exact shape of how the tomato looks like and locates the diseased area of tomato in the image with proper bounding boxes around the image. So, to perform this model the author had considered 4 various combinations of models such as VGG, MOBILE NET, Res-Net as considering RCNN as a backbone all these models were performed and all the necessary time information was mentioned briefly in the paper. The major methodology used here is edge box detection models which helps to images with proper resolution. The methodology they came up with is collecting data, pre-processing the collected data and later the main part of the project is to label images using LABELIME tool. Once the images were created the dataset is divided into three parts for training purpose, validation, and testing purposes. Later using the models, the results were obtained. As all the papers were considered the same evaluation models were performed to verify the results whether the results were correct or wrong **(Wang *et al.*, 2019)**[.

2.2.5. Image Based System to Capture the Tomato diseased leaves

(De Luna *et al.*, 2019) explained about a farming system which works very smartly for detecting diseased images in the fields. They have implemented a box which can cover all the four sides of tomato and can capture the images of all sides. For the model performance of this model, they have used an F-RCNN model to make use of the system. The main important thing is that they have developed a website where a farmer can log in to the website and can observe the diseased plants which are affected by various reasons. To perform this model, they have considered around 5000 images and trained 80% of the images using Alex-Net. Finally, the model development was taken place using F-RCNN with 50 epochs which helped the model to get an accuracy of 95%.

2.2.6. Smart Mobile Application

This is one of the most efficient models this research had ever come across where the author introduced a new method of the methodology of smart mobile application which it helps in finding tomato affected leaves. This model is built on Mobile Net where this can help the farmers in a way that using mobile phones, they can detect the disease of tomato plants with an embedded system. For this model, they have considered 8000 images and trained the model based on D-CNN and gained an accuracy of 93.2%. This kind of applications helps people to make their work easy and in a very simple way they can find the diseases **(Elhassouny and Smarandache, 2019)**. **(Hidayatuloh *et al.*, 2018)** have discussed the same kind of technique which is performed using a different efficient model which is a SQUEEZE-NET model this is one of the CNN models which helps in building the smartphone applications as well as in this paper they also discussed about the robot usage in green house places in the crop fields and this is also involved in server-side computing along with the micro-processing controller systems techniques. Majorly, a dataset was considered from the fields of plasma around 1400 images and pre-processed the image sizes to 224 x 224 and performed a method called as normalization and then they performed the Squeeze Net model. Using evaluation metric technique such as confusion matrix they build all the information of epochs and the accuracy of each image class in a matrix format which helps the individuals to understand the process happening in each step. In the same way, the other paper, have discussed the same thing of Squeeze Net but they compared the model with Alex Net and considered a huge amount of plant village data and performed the models in the same pattern and gained accuracies and finally concluded with the Squeeze Net which is better for mobile applications usage compared to Alex Net **(Durmus *et al.*, 2017)**.

2.2.7. CNN and LVQ Model

(Sardogan *et al.*, 2018) used a basic CNN with limited layers but he implemented a new feature in this paper to detect tomato diseased leaves is LVQ methods which helps in calculating the weights of each class in training purpose. So, he collected a very small amount of data around 500 images and divided the data for training and testing and after dividing the data he reshaped all the images to the same size and performed the model using LVQ vectorization technique which produced a normal result compared to all the different techniques in deep learning.

2.2.8. Performing the models of PCA and SVM for Tomato Detection

According to **(Heltin Genitha *et al.*, 2019)** I scrutinize that the paper is a different approach for image detection whereas author worked on PCA and SVM models by collecting 1000 sample images where the images were divided for training and testing with the help of PCA where this plays a major part in extracting the features of images and these features will be forwarded to SVM with the help of this method the diseases get detected with a proper match of supporting vectors. If once the data is divided then they performed around 5-6 techniques such as histogram equalization, Covariation of

matrices etc., in this way once the pre-processing is done in the final stage they performed the model and got an accuracy of 88.67% which provides the cappa value of 0.82. In this way, the whole process is completed in a very different method and help the farmers in detecting the tomato diseased plants.

2.2.9. Image Processing for Detecting Tomato Diseases in Plants

Image Processing techniques are also one of the Deep Learning processes which use the same kind of algorithms. According to **(Kumar and Vani, 2019)** they have considered the data around 15000 images and performed four different models such as VGG, Res-Net, Inception and Le-Net by using same king od methodology for all the models and finally mentioned that the VGG model gave about 99.87% accuracy compared to all the models performed the author concluded that the VGG is the best Deep learning model for considered dataset and the topic of tomato detection. In addition to this other author **(Bapat *et al.*, 2020)** have mentioned about the Image Processing techniques which are implemented using MATLAB and finally gained an accuracy of 98% by comparing with different CNN models. In the same manner, the techniques of image processing are used in a different way in **(Jasim and AL-Tuwaijari, 2020)** where the dataset is considered was much bigger by combining potato, tomato, and pepper diseased leaves. The whole dataset is finally with 20000 images and the CNN technique is been implemented using different backbone techniques such as Google Net and Whe Net have been used differently by linking to convolutional layers and finally they have obtained an accuracy of 98%. Where even these techniques proved to be the best models of CNN in detecting any kind of plant diseases. In **(Khirade and Patil, 2015)** the author mainly focused on segmentation method which comes under the ANN model. Firstly, the author considered the plant village dataset as everyone considered and then he divided the images into different images with the same similarity by matching to the same images. This method segmentation is done using an otsu method which is a feature extraction in a way that the image is first segmented into a blurred format and then it is converted into a binary format and finally the model gets trained based on the binary format of the image. With this model, the image can be read by a system more efficiently. In the other way, the k means clustering algorithm is also been performed by the author to increase the pixel value where the k means works in a way that the image is divided into clusters and based on the model the cluster will be identified in the form of pieces and later the image will be created automatically in the system using these techniques. Finally, this model is helpful in providing the data for the feature extraction phase. Once the feature extraction phase is done, they have performed the model which is ANN for the classification purpose. The author also mentioned that the SVM's and different models can also be helpful for the model evaluation purpose to improve accuracy. In the same way, there is one more author who discusses the image processing in identifying the tomato images using Raspberry Pi software which helps for performing the software models to detect images they have collected through online and they used OpenCV which is a python library for manipulating the images and they have

used the python code for the model to run. The methodology which they have mentioned is that they augmented the data and then they did fine-tuning and the transfer learning is applied for the model evaluation which is a CNN architecture model helps in integrating the system with the IoT server which will help the rural and remote areas for the better detection **(Adhikari *et al.*, 2018).**

Table 2 provides information about the comparisons in previous works of what methodologies authors have implemented using various Deep learning & Transfer Learning models and how much of accuracy had been gained in using different techniques, gives a brief of the previous research studies.

Table 2: Provides the information of different authors work and their approaches and accuracies of the models.

| Authors	Proposed Methodology	Models Adapted for Classification	Results Achieved
(Bapat, Sabut and Vizhi, 2020)	Extraction, Segmentation, Clustering	CNN	96%
(Nithish *et al.*, 2020)	Extraction, KNN, RBF, SVM	Alex-Net	76.1%
(Batool *et al.*, 2020)	Collection, Extraction, Segmentation and Classification	K-Means and ANN	92.5%
(De Luna, Dadios and Bandala, 2019)	Scaling, Normalization, K-Fold	Squeeze-Net	86.92%
(Durmus, Gunes and Kirci, 2017)	Edge Extraction, Histogram Equalization, Threshold Segmentation	Principal Component Analysis and Support Vector Machine	88.67%, Kappa value 0.82

2.3. Conclusion

According to the survey done in this research study from different IEEE and journal papers, it is clear that they have used different Deep Learning models and various applications and also have done various comparisons using different datasets of plant village along with some of the MNIST datasets. Here, to inform that the model comparisons which performed in this research were not been in any paper to the knowledge of this research study, but models might have performed using different plant disease datasets. So, hereby this study makes sure that the model comparisons which were performed are Le-Net, Dense-Net-121, Mobile-Net and CNN.

3. Methodology

3.1. CRISP-DM Methodology

The Methodology which was implemented in this research study is based on the CRISP-DM methodology. This methodology is user friendly and can be understood very easily. It also proved as one of the best methodologies compared in data mining research studies. The whole process had been explained in the same methodology process by a process to get a clear understanding for readers. The major issue of this project is detecting diseases in plants and the stakeholders are farmers who will be benefited because of detecting Tomato Leaf Diseases. The overall methodology is followed in the same way mentioned in the step by step process in **Figure 2**. In below sections gives detailed information about the steps involved in this research work.
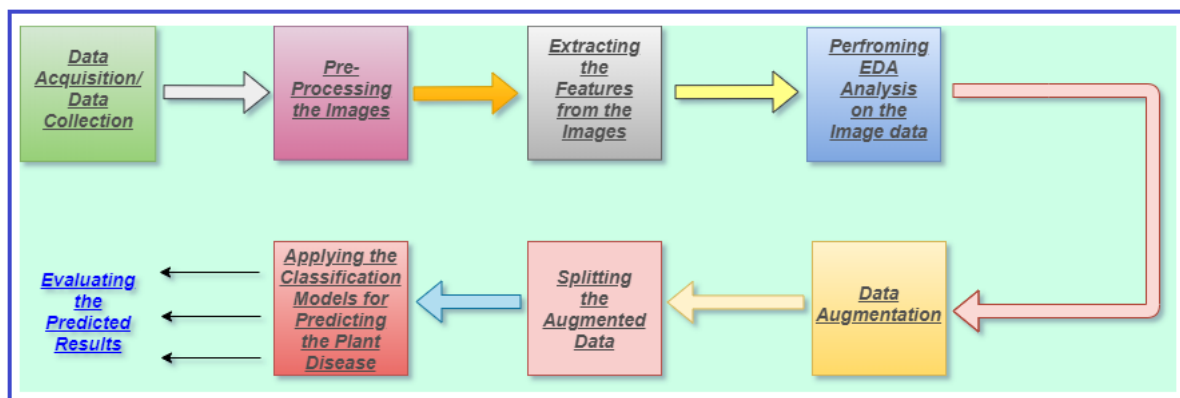
Figure 2: Proposed Methodology for Detecting Tomato Leaf Diseases basing CRISM-DM.

3.2. Business Understanding

In this step, we can come to know about the project how well it is useful in the real-time for world markets. As the stakeholders are farmers, they are facing enormous issues in selling their crops so in this perspective this research goal is to help them in finding the disease and how can it be cured might be analysed in the next. For any problem first thing is to identify, once after the identification of the problem then we can set up some plan. In this scenario, this research planned to provide hope for the farmers and by reducing the number of diseases the environmental conditions also change accordingly which helps in providing the major benefits for the fertilizers. As previously mentioned in the background scope main intention is to plan indoor farming which is already implemented by a person and some of the scientists. If we consider this research study, in this case it helps the lot of people in a way that by identifying what kind of disease. We can easily have an idea of what kind of remedies need to be taken to eradicate those diseases. In this way, the crop economy increases farmers will be satisfied and the fertilizers economy increases and there will be no chance of artificial vegetables which helps in maintaining proper hygiene. How can others trust this project is by checking the revamped statistics of the sales of the vegetables in future and the new agricultural system will come into the light. To make this business goal successful only thing to be followed in identifying the symptoms in the early stage using this learning algorithm

and develop the antibodies related to the disease. So, this can be successful in the market. With the help of this project, goal of saving the lives of farmers is been satisfied.

3.3. Dataset Summary

The most important part of any research study is to perform any model is acquiring Data. This research had also collected the data from KAGGLE website containing Tomato diseased leaf images of 17000 containing 9 diseased leaf classes and 1 Healthy leaf class. In each class, Tomato disease leaves are around 1700. The classes which are present in the **Figure 3** are some of the sample leaves of different diseased tomato classes. So, based on this data research was performed using all the Deep Learning models and identified the plant diseases.
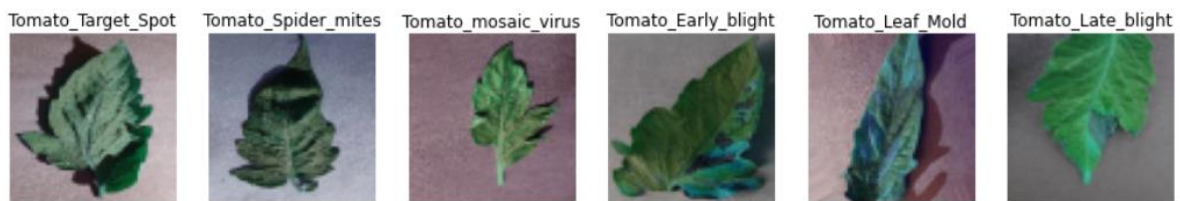
Figure 3: Sample Images for the Dataset driven from the Dense Net model.

✓ **Dataset Link:** https://www.kaggle.com/kaustubhb999/tomatoleaf?

3.4. Steps Involved in Processing the Data

To perform the model firstly, everyone must check for artefacts as we are collecting the data from the outer source. Once the data is collected we have to check the image resolutions, pixels and labels etc., based on this we have to prepare the data in a way that the model can understand the images very clearly and the images were labelled and subcategorized according to their own class and finally the data is divided into 2 different folders for testing and training.

3.5. Exploration of Data

In exploring the data, research study had identified that all the classes were divided equally, and it is ready for splitting according to the model. Hereby to inform individuals that the analysis gives detailed information about the dataset and its count.

3.6. Incrementation of Data

This step is not performed on all the models it is performed only on the models which are not giving good accuracy results. Where is it used in Mobile-Net model wherein this step the dataset is augmented, and the images had increased in its number which gives the proper accuracy. Remaining models provided good accuracy which helped in predicting the disease.

3.7. Initializing the Deep Learning Models

This is the main step for methodology where in this step research will define all the parameters and input shapes required for the models. In this case, research was performed using 4 different models and used the trained data for training the model.

3.8. Evaluation of the Model

The final step is verifying the results which are obtained from the test data from the models based on different constraints such as (Precision, Accuracy, Recall, F1-Score, R2-Score, MAE and MSE) which already pre-defined in Keras and Tensor Flow Libraries.

4. Design Specification
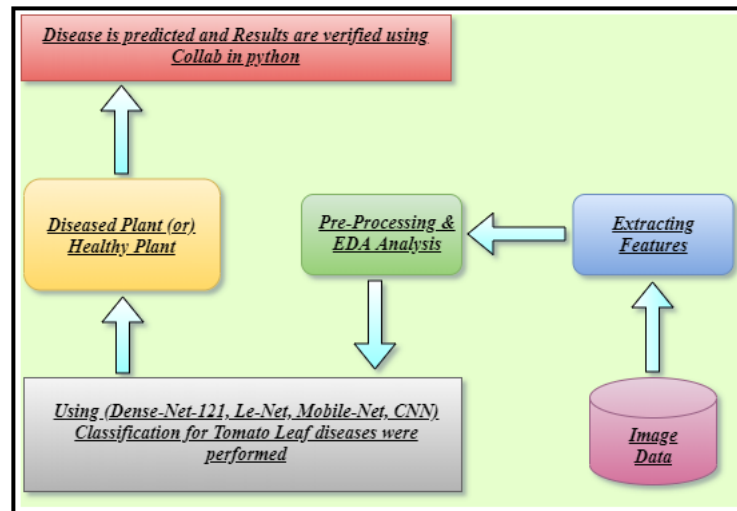
Figure 4: Design Specification Flow Diagram for Detecting Tomato Leaf Diseases

Before going to know about the models, readers must know that, the dataset used in this research study was in two different ways. Where in the model Mobile Net the data images were divided into train and test folders and 10000 images were stored in train and 7000 in test and then used for performing the model. In the other case, the data images which are 9 diseased images and one tomato healthy images were all stored in a single folder which is train and while modelling the data it is divided in the ratio of 1:8 where 13600 images for training and 3400 images for testing the models Dense Net, Le Net and CNN.

5. Implementation

The Major part of whole research is explained in this step, till now, the research came to know about the information regarding previous works, data collection and processing. After all the steps the main step is to train the images with the help of model. In this step, the model reads all the images based on the layers involved in the model and finally the model predicts the disease. **Figure 4** gives an idea of the whole architecture followed in this research study.

5.1. Project Setup by Hardware & Software

Deep Learning models can be performed only on preferred platforms where all the codes cannot run on all the platforms. As the libraries and packages were well defined in python. The research was developed using python language to implement the codes and to execute those codes Anaconda Navigator and Google Collaboratory was used in this research. Whereas the Google Collab is a cloud platform used by everyone to run the codes using GPU

online and it is 5 times faster than local level platform Jupyter notebook. To implement all the codes in Deep Learning the python had provided Keras, Tensor Flow and Pytorch pre-defined libraries for image data. For plotting, there are pandas, Matplotlib and seaborn etc., But many of the people had done their research studies of tomato detection using MATLAB. Comparatively, the Collab gave better results in this research study.

5.2. Dense NET-121

The Dense Net is known as Densely Connected Neural Networks which is a CNN model used as an extension to Res-Net model and 121 is the output size of the model there are 169, 201 and 264 sizes as well. In Dense-net the major useful thing is that, it creates dense blocks, based on the dense blocks created the neural network can easily understand the file what people are passing in and helps in giving proper accuracy result values **(Huang *et al.*, 2017)**.

Step 1: The very first thing after loading the data is to initialize the model using transfer learning from the Image Net database.

Step 2: Now, the model reads the data and identifies the input as an image and then passes the input shape of the image to the model in this step.

Step 3: Convolutional 2D is defined with 3 kernel size and with 3x3 Conv.

Step 4: Hereby, the max Polling layer is used. Because this layer is the most important layer where if the feature maps are high in number then this layer helps in reducing the size of those maps.

Step 5: This step is the major difference in all the CNN models where in this step dense blocks were created according to the model. This dense block multiplies the values of the previous block and adds to the next block and same happens till the end of the last block.

Step 6: In this step, Global Average pooling for the model will be performed where it helps in calculating the feature maps. Simply, it will keep on reducing the data till it gives the data to the final classified layer.

Step 7: For any model which contains huge amounts of data runs very slowly to read whole data. So, in this case, batch normalization helps in revamping the speeds and stability of the model.

Step 8: In this step, the drop out layer comes into the picture where to avoid overfitting we use drop out function in deep learning.

In this way, the whole Dense Net 121 model is performed and predicted the diseases very much accurately. The model had given the best accuracy with 97% with 100 Epochs and 96% for 25 & 50 Epochs. This is the best model accuracy compared to all the other models.

5.3. Le NET

This model is very easy and simple model in CNN where it had only 7 layers. Out of that, only one is fully connected layer and 3 are convolutional layers and one is sampling layer.

Step 1: The first step is to input the image and define an input shape for the model.

Step 2: Next step is defining the convolutional layers based on the input shape. So, 20 kernels with 5x5 size is utilized in this research study for this model.

Step 3: In this step, the model had used the drop out function for overfitting of data.

Step 4: Once after dropping out the data then we need to activate the layer. So, to activate the layer RELU function had been used in this research.

Step 5: The feature maps which are created in the very early layer this helps in balancing those maps by increasing or decreasing.

In this way, the whole model was been performed where in the next steps the same repetition of layers for the 2nd convolutional layers was implemented and finally dense and flatten functions are used for reshaping and inputs the images in sequential order. Le Net provided about 91% accuracy with 100 Epochs and 86% and 85% with 25 and 50 Epochs, respectively.

5.4. CNN

1. CNN is the head of all the algorithms where all the above-mentioned algorithms have been derived and backboned to CNN by using at various stages of CNN. So, here on CNN, the layers relate to different parameters by adding the weights and biases functions. In the later stages, CNN is a basic convolutional structural layer where there is no kind of internally activated layers. In CNN if the layers were internally connected is by saying that the neuron present in the first layers will be connected to all the neurons in the second layer this all happens in CNN this is called densely connected Neural Network where in this way the models will be activated at different parts and also the dense block will be created across the Conv layer.

2. In the same way, all the other models were also connected in different ways to the CNN. So, to give a difference in ratios of the models this research study was been performed along with the CNN and compared the results with the other models. Based on the input data provided the CNN model also worked very well compared to other models.

3. In this scenario, CNN was implemented by using 5 Conv layers and for every layer, the max-pooling layer is also included. In addition to this, to increase the speed of the model, Normalization is performed. Finally, the drop out layers and dense layers were used for good model accuracy. This had provided 95% with 100 Epochs and 88% and 89% accordingly with 25 and 50 Epochs. (**Khan *et al.*, 2020**).

5.5. Mobile Net

1. The Mobile Net model is a less weight process and it can use depth wise separable layers where it can read a single convolution for every colour channel which can help the model in filtering the input channels. To combine the outputs of the model the depth-wise convolution had implemented a new technique known as 1x1 convolution. This convolution helps in combining the filters and divides into two layers. To reduce the model size Mobile Net introduces a new technique called as Factorization.
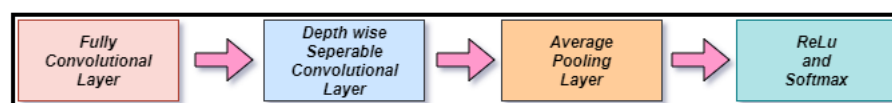

Figure 5: Basic Mobile Net Architecture (**Sinha and El-Sharkawy, 2019**).

2. There are two different versions in Mobile Net which are Mobile Net and Mobile Net v2 which has different formats in their internal layers. If we observe **Figure 5** it gives information about a basic architecture of the mobile net where all the mobile net models follows the same architecture with small changes **(Howard *et al.*, 2017)**.

3. In the same way, the model is also developed based on the same architecture where the model is a predefined model which had been implemented in this research study based on the Image Net hub. As mentioned in the above models the Mobile Net is also having the input shape with different layers. In this way, the whole Mobile Net model is implemented. This model had given an accuracy of 91% with 100 Epochs and 88% and 87% with respect to 50 and 25 Epochs.

6. Evaluation Results

6.1. Dense Net
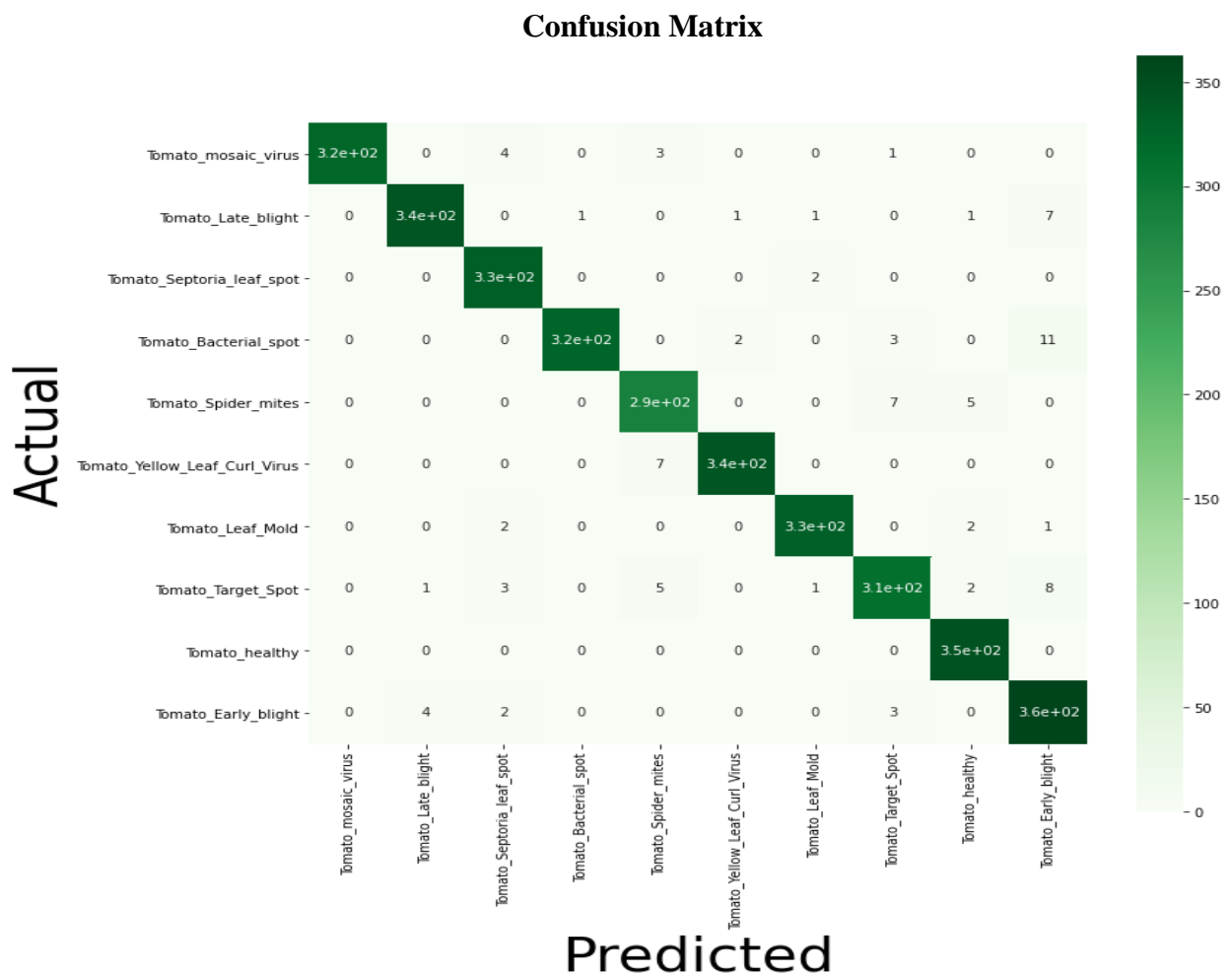
Confusion Matrix

Figure 6: Confusion Matrix for Dense Net Model

Figure 6 provides the information about the actual and predicted values of the dataset where all the green labelled values are the values which are predicted correctly for each class. If we observe the above figure all the values are around 320-360 images were predicted

perfectly in all the classes. Where for the training 1360 images from each class was trained and remaining all the images were given for the testing and almost all images had predicted the correct results.

Loss and Accuracy
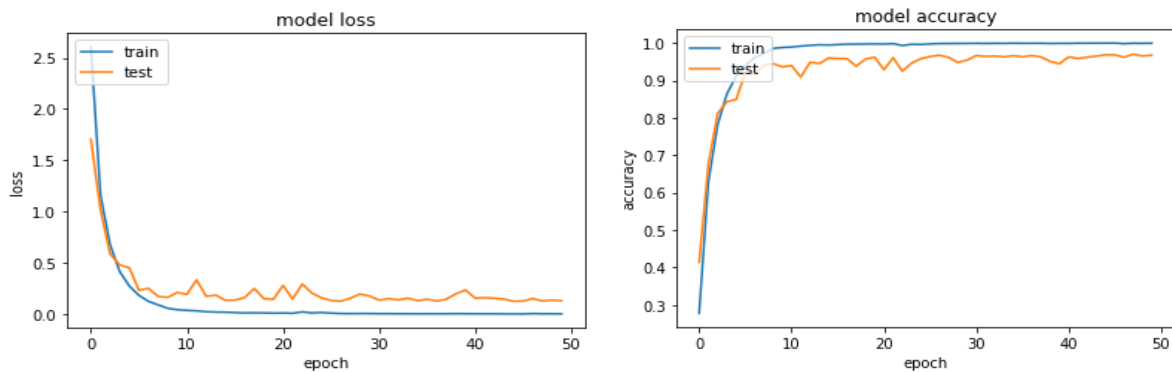
Figure 7a&7b: Val Loss & Val Accuracy Graphs for Dense Net

The **Figures 7a&7b** provided the graphs with accuracy and loss values raised for each epoch till 50 epochs. Where the data is not having any over-fitting, the model had given proper graphs. By observing **Figure 7a** we can understand that the loss value is in decreasing order wherein vice versa the Model accuracy values had raised nearly to 0.9 for training and testing. In this way, the model accuracy had increased and raised to 0.9.

6.2. Le Net

Confusion Matrix
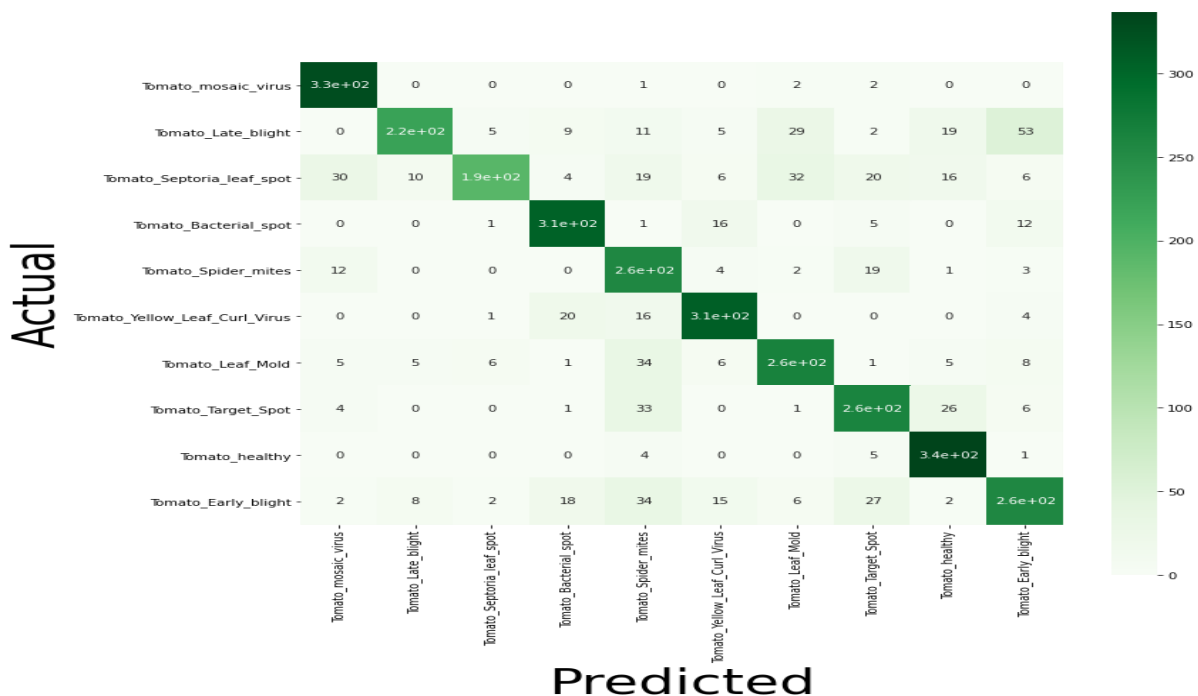
Figure 8: Confusion Matrix for the Model Le Net.

17

Figure 8 explains the predicted values of the model Le Net almost all test values were predicted correctly. Out of 1700 images from each class 1360 were given for training and remaining 340 were given for testing and all the images from each class have given good predictions which means the model is performing the perfect results.

Loss and Accuracy
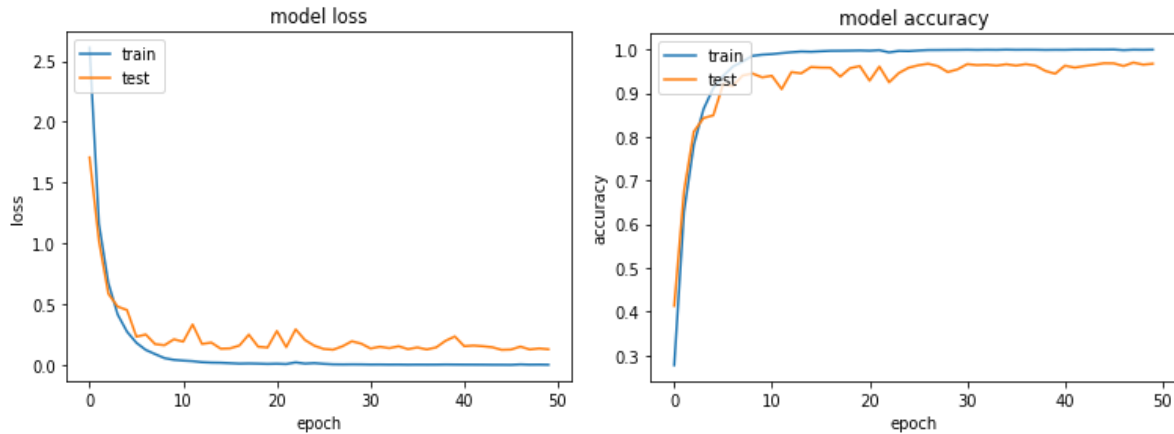
Figure 9a & 9b: Val Loss & Val Accuracy Graphs for Le Net Model.

The above graphs represent the values and accuracies of the model Le Net. In **Figure 9a** the loss value is decreased gradually from 2.5 to 0.1 which is shown for train and test in blue and orange respectively and in **Figure 9b** the Model accuracy had gradually increased from epoch to epoch and by the end of the epoch 50, it is around 0.9.

6.3. CNN

Loss and Accuracy
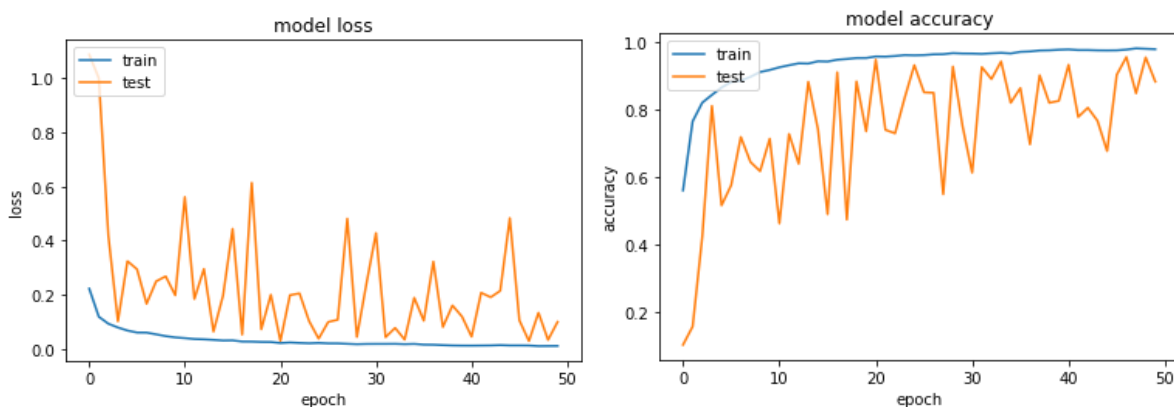
Figure 10a & 10b: Val Loss & Val Accuracy Graphs for the Model CNN

The accuracies and loss values of the CNN model was mentioned above in the form of graphs where the loss values from **Figure 10a** for both train and test have gradually decreased from 0.9 to 0.1 and in **Figure 10b** the model accuracy values for both train and test is increased from 0 to 0.9 and 0.8 respectively.

Loss and Accuracies of all the Models
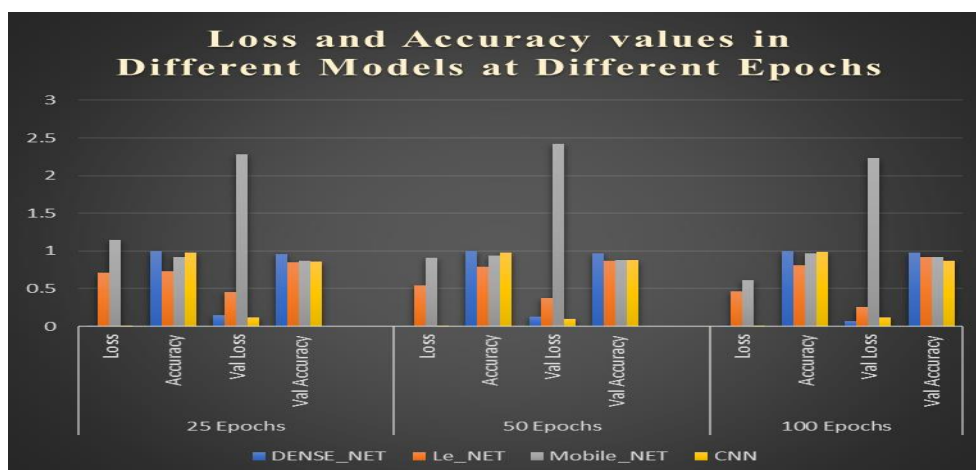
Figure 11: Graph indicating Loss & Accuracy Values for all the Models.

Figure 11 indicates the Loss, Accuracy, Val Loss and Val Accuracy values for all the models and been compared with the accuracy which is highest, and loss which is lowest. In detail, if we observe the graph at 25 epochs the accuracy and Val accuracy is high in Dense Net and very less in CNN compared to all other models. Coming to 50 epochs again Dense Net stood in the first position with the highest accuracy and less loss. In the same way, with 100 epochs with an 0.9 accuracy value compared to all the models and very less in CNN compared to other models. In this way, to conclude the graph at each epoch value, the Dense Net is having the highest value for both accuracy and Val accuracy.

Evaluation Metrics Results Comparision
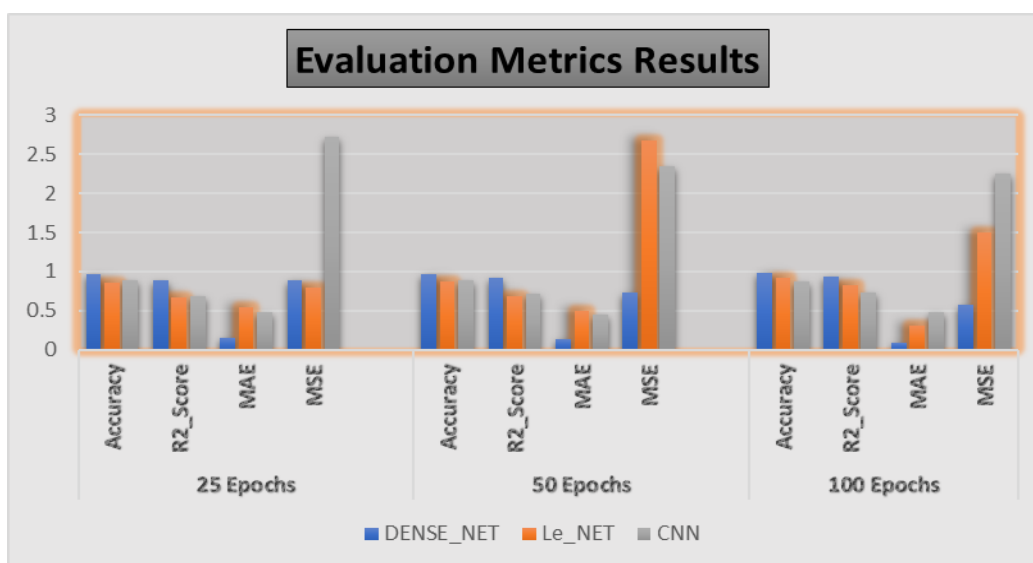
Figure 12: Evaluation Metrics Results for all the Models.

Considering the **Figure 12** it is clear that graph is mentioning about the results of evaluation metrics of Accuracy, R2_Score, Mean Absolute Error and Mean Square Error where the mean error values should be nearer to zero which is performed in Dense Net, Le Net and CNN. Comparing these three models the Accuracy value is high in Dense Net compared to all the models mentioned in blue colour. In the same way, the R2-Score is also very high in Dense Net compared to all the models. Considering the MAE values, it is very low for Dense Net and is very nearer to zero which is having very less error value which means the model is performing well compared to all the other models. Finally, the MSE is inconsistent due to the epochs values where it is very less at 25 Epochs and very high at 50 epochs. This is how all the evaluation methods had performed and Dense Net gave the ultimate results and proved as the best model by predicting the disease with good accuracy compared to all the models.

7. Discussion

The research study is completely related to the implementation of Deep Learning and Transfer Learning models. The insights which are implemented in this research study were obtained from the Deep Learning models and are very innovative and productive. For the model performance, a dataset of Tomato Leaves was considered in this research which are injected by 9 various diseases and one healthy class is considered. In previous works, there are many authors who had considered the data, but which are about 500-1000 samples of images. So, there is no problem regarding ethical concerns. The main aim is to help farmers in detecting diseases easily and grow their crops far better than before. So, 4 different models to detect Tomato Leaf Diseases are implemented in this research. The model comparisons which done in this research was not in any part of the research to this extend of study.

In this research, Dense Net 121, Le Net, Mobile Net and CNN model comparisons which are implemented are not in any previous works of what mentioned in previous references. From the results evaluated the Dense Net 121 gave the best results with 97% accuracy. To initiate this model the pre-processing, Feature Extraction and EDA analysis were done as a part of this research and finally implemented the Dense Net 121 using Transfer Learning. Other model Mobile Net is also implemented using Transfer Learning, but the CNN gave the least accuracy when model performed with 100 Epochs. In the same way, the Dense Net 121 is in first and Le Net, Mobile Net gave 91% accuracy with 100 epochs and while training the model with 50 Epochs Le Net gave the less accuracy with 86% and Mobile Net and CNN are with 88%. Model training with 25 Epochs Dense Net 121 is in the first position with 96% and Le Net and CNN are with 85%. In final comparing all the Epochs with 25, 50 and 100 Dense Net 121 gave the best accuracy comparing with other models because of the Data images which are very perfect for understanding.

Comparing this research to the previous research studies the model gave the best predictions in identifying the Tomato Leaf Diseases. To evaluate all the results, evaluation metrics such as Accuracy, Precision, Recall, F1-Score, R2-Score, MSE and MAE were used in this study.

Based on these results Dense Net Model results which were obtained from the model were perfect in each class and gave the best predictions.

8. Conclusion and Future Scope

The whole research is totally helpful for the farmers. To improve the agricultural sciences, this research might help scientists in elaborating the process of finding the plant diseases. In this process, the 4 different models performed have detected the Tomato Leaf diseases which will help the farmers to understand the kind of disease occurred in leaves. So, based on this disease identification process farmers can implement some major precautions to get rid out of this disease and get a good crop which helps in revamping their economy. Out of all the 4 models, the proposed model Dense Net 121 had given a high accuracy of 97% with 100 Epochs. In future research, individuals can consider the same data and can combine the Deep Learning models such as CNN, DNN, Alex Net, Le Net and Dense Net algorithms and create a hybrid model and compare the results with existing models.

9. Acknowledgement

References

Adhikari, S., KC, E., Balkumari, L., Shrestha, B. and Baiju, B. (2018) 'Tomato Plant Diseases Detection System Using Image Processing', *Researchgate.Net*, I(September), pp. 81–86. Available at: https://www.researchgate.net/profile/Saban_Kc2/publication/327930730_TOMATO_PLANT _DISEASES_DETECTION_SYSTEM_USING_IMAGE_PROCESSING/links/5bada156458 51574f7ebdd24/TOMATO-PLANT-DISEASES-DETECTION-SYSTEM-USING-IMAGE-PROCESSING.pdf.

Agarwal, M., Singh, A., Arjaria, S., Sinha, A. and Gupta, S. (2020) 'ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network', *Procedia Computer Science*. Elsevier B.V., 167(2019), pp. 293–301. doi: 10.1016/j.procs.2020.03.225.

Bapat, A., Sabut, S. and Vizhi, K. (2020) 'Plant leaf disease detection using deep learning', *International Journal of Advanced Science and Technology*, 29(6), pp. 3599–3605.

Batool, A., Hyder, S. B., Rahim, A., Waheed, N., Asghar, M. A. and Fawad (2020) 'Classification and Identification of Tomato Leaf Disease Using Deep Neural Network', *2020 International Conference on Engineering and Emerging Technologies, ICEET 2020*. IEEE, pp. 1–6. doi: 10.1109/ICEET48479.2020.9048207.

Durmus, H., Gunes, E. O. and Kirci, M. (2017) 'Disease detection on the leaves of the tomato plants by using deep learning', *2017 6th International Conference on Agro-Geoinformatics, Agro-Geoinformatics 2017*. doi: 10.1109/Agro-Geoinformatics.2017.8047016.

Elhassouny, A. and Smarandache, F. (2019) 'Smart mobile application to recognize tomato leaf diseases using Convolutional Neural Networks', *Proceedings of 2019 International Conference of Computer Science and Renewable Energies, ICCSRE 2019*. IEEE, pp. 1–4. doi: 10.1109/ICCSRE.2019.8807737.

Fuentes, A. F., Yoon, S., Lee, J. and Park, D. S. (2018) 'High-performance deep neural network-based tomato plant diseases and pests diagnosis system with refinement filter bank', *Frontiers in Plant Science*, 9(August), pp. 1–15. doi: 10.3389/fpls.2018.01162.

Heltin Genitha, C., Dhinesh, E. and Jagan, A. (2019) 'Detection of leaf disease using principal component analysis and linear support vector machine', *Proceedings of the 11th International Conference on Advanced Computing, ICoAC 2019*, pp. 350–355. doi: 10.1109/ICoAC48765.2019.246866.

Hidayatuloh, A., Nursalman, M. and Nugraha, E. (2018) 'Identification of Tomato Plant Diseases by Leaf Image Using Squeezenet Model', *2018 International Conference on Information Technology Systems and Innovation, ICITSI 2018 - Proceedings*. IEEE, pp. 199–204. doi: 10.1109/ICITSI.2018.8696087.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H. (2017) 'MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications'. Available at: http://arxiv.org/abs/1704.04861.

Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K. Q. (2017) 'Densely connected convolutional networks', *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. IEEE, 2017-Janua, pp. 2261–2269. doi: 10.1109/CVPR.2017.243.

Jasim, M. A. and AL-Tuwaijari, J. M. (2020) 'Plant Leaf Diseases Detection and Classification Using Image Processing and Deep Learning Techniques', pp. 259–265. doi: 10.1109/csase48920.2020.9142097.

Kaur, M. and Bhatia, R. (2019) 'Development of an improved tomato leaf disease detection and classification method', *2019 IEEE Conference on Information and Communication Technology, CICT 2019*. IEEE, pp. 1–5. doi: 10.1109/CICT48419.2019.9066230.

Khan, A., Sohail, A., Zahoora, U. and Qureshi, A. S. (2020) 'A survey of the recent architectures of deep convolutional neural networks', *Artificial Intelligence Review*, pp. 1–70. doi: 10.1007/s10462-020-09825-6.

Khirade, S. D. and Patil, A. B. (2015) 'Plant disease detection using image processing', *Proceedings - 1st International Conference on Computing, Communication, Control and Automation, ICCUBEA 2015*, pp. 768–771. doi: 10.1109/ICCUBEA.2015.153.

Kumar, A. and Vani, M. (2019) 'Image Based Tomato Leaf Disease Detection', *2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019*. IEEE, pp. 1–6. doi: 10.1109/ICCCNT45670.2019.8944692.

De Luna, R. G., Dadios, E. P. and 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*. IEEE, 2018-Octob(October), pp. 1414–1419. doi: 10.1109/TENCON.2018.8650088.

Nithish, E., Kaushik, M., Prakash, P., Ajay, R. and Veni, S. (2020) 'Tomato Leaf Disease Detection using Convolutional Neural Network with Data Augmentation', (Icces), pp. 1125–1132.

Puno, J. C. V., Bandala, A. A., Dadios, E. P. and Sybingco, E. (2019) 'Vision System for Soil Nutrient Detection Using Fuzzy Logic', *IEEE Region 10 Annual International Conference, Proceedings/TENCON*. IEEE, 2018-Octob(October), pp. 617–622. doi: 10.1109/TENCON.2018.8650242.

Sardogan, M., Tuncer, A. and Ozen, Y. (2018) 'Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm', *UBMK 2018 - 3rd International Conference on Computer Science and Engineering*. IEEE, pp. 382–385. doi: 10.1109/UBMK.2018.8566635.

Sinha, D. and El-Sharkawy, M. (2019) 'Thin MobileNet: An Enhanced MobileNet Architecture', *2019 IEEE 10th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2019*, pp. 0280–0285. doi: 10.1109/UEMCON47517.2019.8993089.

Surya Engineering College and Institute of Electrical and Electronics Engineers (2020) 'Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020', *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020*.

Tm, P., Pranathi, A., Saiashritha, K., Chittaragi, N. B. and Koolagudi, S. G. (2018) 'Tomato Leaf Disease Detection Using Convolutional Neural Networks', *2018 11th International Conference on Contemporary Computing, IC3 2018*, pp. 2–4. doi: 10.1109/IC3.2018.8530532.

Wang, Q., Qi, F., Sun, M., Qu, J. and Xue, J. (2019) 'Identification of Tomato Disease Types and Detection of Infected Areas Based on Deep Convolutional Neural Networks and Object Detection Techniques', *Computational Intelligence and Neuroscience*, 2019. doi: 10.1155/2019/9142753.