Chapter-1

INTRODUCTION

1.1 General Introduction

Agriculture, Agroforestry not only are the factors affecting the food industry and the feeding habits of the people they also play a crucial role in escalating the economy of the country. In India, the economy predominantly (70%) depends on agricultural productivity [7]. The diminishing agriculture and plant diseases could actually cause economic, social, environmental losses. Traditionally, the plant investigation took place manually by certain plant experts or trained people. Most of the plant diseases lay on leaves, fruits and buds. The variations in plant diseases made it really difficult for even the experts to manually detect them, this can be solved using the machine learning algorithms considering the similarity between plant species. So due to the technological advancements, automated, timely and accurate detection of plant diseases has become an important field of study to increase the production and improve the quality of the supplies from the agroindustry of any country [6]. Flower forms the crucial part of the agriculture as they are the feed to the small species completing the food cycles and maintaining the balance. These flowers can be used as drugs for many diseases that why having knowledge about them plays a crucial role. The techniques, based on colour and texture failed due to the countless species availability but machine learning integrated with mobile technologies gives effective results.

The detection of the plant disease methods is divided into phrases like image acquisition, image processing, feature extraction and image classification, on various datasets or the knowledge datasets captured using scanners, videos and cameras. Flower detection has remained a challenge using image processing and computer vision so the potential solution to the issue addressed has been sorted out by working on machine learning models such as Multilayer Perceptron, Support Vector Machine, K-Nearest Neighbors on flower-based datasets.

1.2 Problem Statement

Agricultural fertility being the salient feature of the economy helps in providing pillars of strength for the building of a strong country. The proper study of these plants and the diseases can lessen the crop yield loss and can increase the quality, quantity and productivity of the crops. Earlier the identification and detection were performed by the experts based on their experience and the efficiency of the results would be still doubtable therefore automated detection for the disease and flowers proved to be a requirement. This project can predict the disease just by scanning the plant leaf, making it easier as well as cheaper. The technique can work hand in hand with large crop-producing areas and fields.

Flower identification, part of plant identification plays an important role in the pharmaceutical industry as these are the drugs and cure for many diseases. The flowers are also used in horticulture appreciation, flower fair exhibition and so on, people wish to obtain the type of certain flowers and information. Thus, identifying the flowers and knowing about them is crucial. The challenges here are the wide variety of these flowers and their similarity in colour and features. An automated system can help in turning down these problems and creating the models which can become a tool for farmers to increase the yield and contribute to the economy of the country.

1.3 Significance of the Problem

Plant disease detection attracts attention in the field of agriculture. The plants which die due to some issues lead to a decrease in yield loss and can severely affect the country. If the detection of these diseases is done on time it could escalate the economy and farmers can also get benefited. The technology has become a boom with convenient and high-speed internet.

The automated model prepared using machine learning integrated with mobile technology can efficiently help the farmers to detect the disease immediately. The biggest advantages of having an automated system are that they are fast, light weighted easily available, and have high accuracies for the prediction. Thus, the systems can be easily directed to agricultural industries.

As indicated in an examination by the Related Offices of Trade and Industry of India, yearly yield misfortunes because of nuisances and infections add up to Rs.50,000 crore, which is huge in a nation wherein any event 200 million Indians hit the sack hungry consistently [20][21]. Therefore, there is a requirement for successful early illness identification procedures to control plant infections for food security and supportability of agro-environment. This new continuous application will endeavour to take care of the issue of recognizing sicknesses in plants utilizing AI.

Innovation today is available in all pieces of our lives, and it is broadly accessible to everybody. Practically 94% of ranchers utilize mobile phones, particularly in developing nations, where cell phones might be the main accessible across the board figuring and correspondence innovation [22]. This number leads us to the end that we can utilize versatile innovation to assist farmers with improving their agrarian creation.

1.4 Empirical Study

Plant Disease

Plant disease is any kind of damage or lesion observed in a plant by an agent that intervenes in its growth, structure, function or any other activity. The occurrence and duration of the diseases vary from season to season depending upon the environmental conditions and type of cultivation. In History, large plant disease outbreaks like the late blight of potato in Ireland, coffee rust in Central and South America have affected huge masses of people [19]. Loss of crops due to plant disease causes hunger, starvation and economic loss to the crop producers and suppliers.

To prevent the above-mentioned crisis, there is a need to detect the plant diseases as soon as possible so that measures for healing the plant or precautions to prevent the spread of the disease can be taken at an appropriate time.

Some of the approaches used for the detection of plant diseases are:

- Keep a simple naked eye on the observation by the experts through which identification and detection of plant disease are done. This is highly costly and time-consuming.
- Image processing was the second step towards processing, but analysing each pixel on your own is tiresome and recurrent.

- With advancements in technology, computer vision and deep neural networks played a major role in plant disease detection. Convolutional Neural Nets are known to learn the pixel settings of the images and the weights can be stored to test other samples with the same accuracy. Also, it ensures high accuracy and performance in classification.

Convolutional Neural Network

It is a class of deep neural networks commonly used for classification of visual data. It consists of an input, an output layer and multiple hidden layers. The hidden layer includes a convolutional, pooling, fully connected and normalization layer. The basic CNN structure is shown in Fig 1.

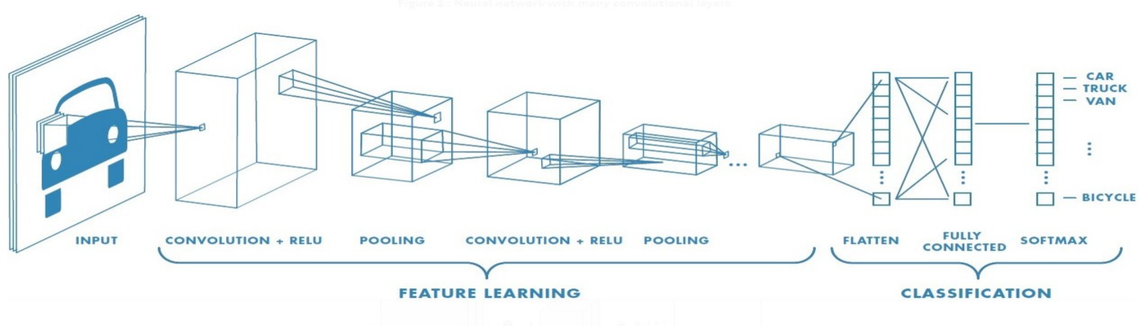


Fig 1: CNN Architecture

1.5 Comparison of The Existing Approaches To The Problem Framed

Plant disease detection has become an important part of agriculture making it a field of research. In 2015, S. Parnell worked on a model to predict the emerging plant pathogens as these pathogens can be a threat to food in the model they have used a simple rule of thumb to determine, surveillance that helps in taking a fixed number of samples at regular intervals. Then another research is done in 2015, which was Advanced methods of the plant disease detection here they have used DNA based and serological methods that are an efficient way to detect the disease [23][24]. In 2017, Alexander Johannes developed an automated plant disease using mobile devices on a wheat use case, the analysis here was done using 7 mobile devices and more than 3500 images captured. In this model, the image processing was done using global colour constancy and leaf segmentation using a binary mask and the results were obtained [3].

Significance of the Problem-Resolution Hyperspectral UAV Images was the detection of yellow rust in wheat that has resulted in significant yield losses respectively. Here the model uses a deep convolutional neural network approach for automated crop disease detection (DCNN).

In our approach, we have developed an automated system for plant disease detection. The dataset used for plant disease detection consists of 73,515 plant images taking into consideration 18 plant species and 47 plant diseases covering a large variation existing in the plant disease and then classified them using CNN.

Chapter-2

LITERATURE REVIEW

2.1 Summary of Papers Studied

In this section, we present a review of the most significant research approaches on flower identification including the disease which forms the major milestone of the crop industry. Many of the techniques were useful including flower identification based on Deep Learning [15] which is performed on the dataset consisting seventeen types of flowers, these are classified using deep networks combined with softmax classifier. A research study on machine learning techniques using current trends and challenges [12] where relevant characteristics of a leaf are calculated using Leaf shape, leaf venation. The further work includes Automatic plant disease diagnosis using mobile capture devices in wheat use cases [3] in this leaf segmentation is used to extract one or more portions from prior colour images. Deep learning-based approach for yellow rust disease [1] uses a deep convolutional network (DCNN) used to detect the yellow rust.

In India, agriculture has always been the most important aspect. The effect on any crop can directly link with the decrease in the economy. The yield of the crop can escalate by the plant disease detection models. [2] Machine Learning Classification Techniques for Plant Disease Detection is one of the many types of research done recently. It consists of the steps such as image acquisition which is gathering of images using cameras, scanners, drone. Then processing the acquired images to improve its features, followed by feature extraction where colour, shape and texture of the disease is evaluated. Finally, technique like SVM is used for citrus fruits, grapes, oil palm; ANN for cucumber; KNN for sugarcane and cotton; Fuzzy classifier for wheat and deep learning for peach, apple, cherry is applied.

A Deep Learning-Based Approach for Automated Yellow Rust Disease Detection [1] is the detection of yellow rust that is a serious fungal disease resulting in yield loss. The model here works on a deep convolutional neural network (DCNN). The dataset consists of four plots, two healthy crops and the other two with yellow rust. The researchers has proposed a model that works on wheat fields with four steps that is data pre-processing, feature extraction, classification, post-pre-processing and the results. 15000 blocks have been extracted out of which (80%) that is 10000 is chosen for training and rest for validation.

Alexander Johannes [3] in 2017 developed a model taking into respect three diseases septoria, rust and tan spot. A database was generated and image prepossessing performed where a leaf is segmented using binary leaf mask. The identification in the model was done in disease region where the hotspot is extracted and was processed in meta classifiers. The results obtained in ROC is 0.80 for all analyzed data. Detection of plant leaf diseases using image segmentation and soft computing techniques [4]. Singh, Vijai, and Ak K. Misra proposed an algorithm for image segmentation in leaves. Disease here is detected using genetic algorithms, which is an evolutionary technique that is generally used for optimization. The methodology is based on machine learning detection techniques. The experiments are performed on a banana leaf to find scorch disease, in bean leaf to find bacterial fungal disease, rose leaf suffering from bacterial leaf spot and lemon leaf agonized by sunburn. The accuracies for the above are 86.54% using K-means and 95.71 using SVM.

In plants, the sign of diseases usually occurs on the buds, leaves, fruits and new branches [6]. Deep Convolutional Neural network has recently gained popularity to analyse and understand the pixels in the images efficiently. Türkoğlu, Muammer, and Davut Hanbay proposed to extract the features from images of plants using 9 different deep neural networks and then applying machine learning algorithms for the classification of the diseases. The dataset used consists of images of plant diseases commonly found in Malatya, Bingöl, and Elazığ regions of Turkey. The feature vectors of sizes: 1024, 1000, 1000, 1000, 4096 were obtained from certain layers of GoogleNet, ResNet50, ResNet101, InceptionV3, InceptionResNetV2, and SqueezeNet model. These feature vectors were then used as inputs to the traditional machine learning classification techniques as SVM.ELN and KNN. The highest accuracy of 97.86% was obtained with the ResNet50 model and SVM classifier.

Plant Disease Detection: A review [7]. Sandhu, Gurleen Kaur, and Rajbir Kaur gives a brief description of the importance and the methodology of plant disease detection using automated techniques. It states that 70% of the Indian economy depends on agriculture and any damage to the crops can have a huge impact on the economy. It divides the plant disease detection process into 4 methods: Image acquisition, Image Segmentation, Feature Extraction and Classification. Image Acquisition can be done by taking pictures of plants using a camera or a mobile phone and adjusting it to the needed resolution and size. This is then followed by image segmentation which simplifies the representation of the image so that it is easier to analyse. Feature Extraction as the name suggests uses some techniques to pull out the features that are important for classification like colour, texture etc. After going through all these phases our data is ready to be fed into a classification algorithm like SVM, Naive-Bayes etc. that would help us to determine the disease.

Just like any other disease, every plant disease has a different type of symptom. The paper [8] briefly discusses the various kinds of symptoms encountered while studying plant diseases and then classifying them accordingly. The first kind of symptom known as scattered small comprises of a huge number of small lesions or spots spread over the leaf surface. The second kind, scattered large, comprises of large lesions spread over the leaf surface. The third kind of symptom called isolated consists of single spots. The fourth kind is widespread, which means large lesions covering the leaf surface completely. The first and the second kind deals by dividing the images with single and clustered lesions separately. While the fourth kind needs to be analysed firstly by the original image and then by analysing the sub-images. For the implementation, a dataset of plant leaves images of 1-24 mega-pixels taken from smartphones and cameras and was formulated. Transfer learning was applied on pre-trained GoogleNet CNN. The first experiment dealt with the classification of the symptom, due to which in this case the images of healthy samples were not included. The second experiment dealt with the detection of the disease and then its classification. The best results were given for severely diseased plants. Though the detection rate was too low.

A lot of emphases has been given on machine and deep learning techniques for plant disease detection due to improved levels of recognition and disease identification [9] in comparison to the traditional approaches like chromatography, thermography etc. It proposes to use Random forest as a classifier because firstly unlike decision trees it overcomes the problem of overfitting. Secondly, it is capable of achieving high accuracy with a small dataset. The dataset is split into training and testing samples. HoG feature extraction is applied to obtain a feature vector. The feature vector of the training dataset is then provided as an input to the random forest classifier. The trained classifier is then saved and is then fed with the test samples to find the prediction accuracy.

CNN’s have been known to show high performance and accuracy in tasks like crop identification, fruit counting, yield prediction and disease detection [10]. This paper proposed to use a large self-constructed dataset, through augmentation. The augmentation was done by another neural network known as a GAN. This helped to solve the problem of overfitting, that is prominent in cases of small datasets. The generated images combined with the original images were then fed into a convolutional net, the convolutional layers extracted the important features. The feature vector then obtained was send as an input to the fully connected 32 layer classifier via a pooling layer. The highest accuracy after training was calculated to be 93.67%.

S. Anubha Pearline proposed a model that is plant recognition using conventional image processing and deep learning approaches [11]. Here the plant species is recognized using cameras, videos following 2 approaches. First is traditional approach identifying species with texture using a local binary pattern (LBP), shape using Hu moments which is then classified using naive Bayes, K-nearest neighbour, random forest and bagging having an accuracy of 82.38%. The second approach is done on three datasets Folio, Swedish leaf and Flavia, pre-trained on VGG16, VGG19, inception V3 and inception ResNet V2. The datasets statistics are different Swedish leaf consists of 15 different classes and 75 images each while Flavia and Folio have 1907 and 637 images respectively. The accuracies vary for the different datasets with 96.53, 96.25 and 99.41 for Folio, Flavia and Swedish leaf using CNN.

A study on the machine learning techniques for automated plant species identification: current trends and challenges [12] talks about the early works for plant identification that were based on parameters like area, length, breadth, perimeter. The challenges in identifying the plant are also depicted by Bojamma & Shastry. The challenges include vast varieties of flora and fauna are present and many of them hold a lot of similarities which makes them difficult to get identified. The leaf identification here is done using the leaf shape linear, oblong, Rhombic etc. Leaf venation is also one of the factors to identify the leaf using the primary veins which run on the centre of the leaf. Furthermore, the classification techniques used were KNN, INN, and SVM.

Identifying the plant with help of its leaf, bark or needle structure is of utmost importance to the foresters and the botanists. That is why the automated identification of plants using its visual features has become a popular area of study [13]. Artificial Intelligence combined with computer vision techniques can help to increase the accuracy of the assessment. A dataset of 637 healthy leaves with 32 different species of plant was taken into account. Using image processing techniques, 22 visual features were extracted from each leaf and divided into 4 groups: dimension, pattern, colour and texture. The RGB scale images were used to identify the colour of the leaf, these images were then converted to grayscale to understand the texture of the leaf, and then the images were converted to binary for the pattern and dimension identification. The dataset was then divided into a training set consisting of 510 images and a testing set of 127 images. Fifteen combinations of the above- mentioned groups were trained and tested on Classification algorithms like ANN, RF, KNN, SVM and Naive Bayes. SVM outperformed all other AI algorithms in each combination. The best accuracy of 94.2% was obtained with the SVM classifier when all the groups were taken into consideration while making the combinations for train and test datasets.

A huge improvement in the performance of the plant image classification can be observed due to the evolution of deep learning techniques [14]. This paper proposes a model that combines a linear and deep learning technique to address the fine-grained plant image classification problem. The dataset has

100-Korean plant information divided into metadata embodying the GPS, location, month of flowering, colour of flower and no. of petals, that is used as an input for the linear model and the image data that is used as an input for the DCNN. The results show that DCNN with a linear model (Gps+Date) gives the highest accuracy of 96% on the top 5 plants. The main drawback of this paper is that it gives a low accuracy score of 78% when all the plants are taken into consideration and hence not reliable for generalisation.

Flower identification based on Deep Learning [15], research that is based on the agroforestry Production and management which forms the basis of the plant identification and flower identification being an important part of it. The dataset consists of seventeen types of flowers by oxford university having different characteristics to improve the quality of sample data is augmented. Then the deep neural network has a depth of 16 is applied which increased the accuracy for the model. This model is different and has advantages over the other, as it is optimised accurate classification identification. The issue with their model is that the availability of the flower datasets is comparatively less adding more complex flower datasets could help them in future and integration with the mobile would also be one of the factors. The issue with their model is that the availability of the flower datasets is comparatively less adding more complex flower datasets could help them in future and integration with the mobile would also be one of the factors.

Flower Identification is a combination of object recognition and image classification, as the system needs to detect the flower as well as classify to which species it belongs [17]. Recognition of the species uses supervised learning algorithms for classification. This paper uses CNN along with the concept of transfer learning as a self-learning technique to detect and identify the species of the flower. The Flower17 and Flower102 dataset from the visual geometry group Oxford University was merged with images of some more species expanding the dataset to 28 species and 102 species respectively. It states that there are three most important attributes to analyse a flower species: colour, texture and shape. These attributes are extracted using handcrafted computer vision techniques.

Colour Histogram is the most reliable feature descriptor to learn about the distribution of each colour. Gray Level Co-occurrence Matrix (GLCM) is commonly used as a texture descriptor. Hu moments and Zernike moments are the descriptors used for shape. LBPs and HoG are two more feature extractors used before the formation of the final feature set. The reduced dataset is then segmented into training and testing sets. The Modelled CNN is then trained using the training set. To reduce the training task a method known as “transfer learning” is used. In this method, a pre-trained network on a large dataset is used as a feature extractor by keeping all the pre-trained layers except the fully connected that are specifically used for classification.

Flower detection has been a difficult task, unlike the object classification due to inter-class similarity and large intra-class variation [18]. The paper proposes to make a large scale flower dataset by collecting 63,442 images of flowers belonging to 79 different species. It also proposes to identify the regions in the image that are most significant in identification by understanding the spatial organization. It uses CNN for feature extraction as they are capable to learn multiple instants of a uniform feature for a particular task. The dataset is initially down-sampled to a resolution of 100 for uniformity. The image is then made to pass through stacked convolution layers that help to extract the beneficial features. This feature vector is then made to pass from 3 fully connected layers with the last one having 79 channels that is one for each class.

2.2 Integrated Summary of the Literature

The integrated summary of the literature studied is shown in Table 1.

Table 1: Summary of Literature Review

SNo.	Title	Database Used	Features Extracted	Classification Technique	Paper Type conference/ Journal
1.	Deep learning-based approach for automated yellow disease from VAV images [1]	4 plots from the station of China academy. (2 Healthy, 2 with yellow rust)	DCNN (Deep convolutional neural network) Using high resolution UAV images.	DCNN	Remote sensing paper - Journal (2019)
2.	Review on ml classification techniques for plant detection [2]	Knowledge based dataset to be created for captured images with different classes.	Colour, shape and texture are extracted using grey level Co-occurrence Matrix (GLCM), Blend vision and machine intelligence etc.	5 different classifiers and their different accuracies are calculated. ANN, KNN, SVM, fuzzy, Deep learning	ICACCS-Conference paper (2019)
3.	Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. [3]	An extensive image database has been created for wheat crops. 987-Rust images 2505-septoria 657-tanspot	Image is processed by means of colour constancy algorithms to minimize the natural illumination variability effects. Leaf segmentation is a module to extract one or more portions from the prior colour-constant image.	Random-Forest based classifier And then in the second step a meta-classifier is used, to compute a confidence score for the particular disease.	Elsevier - Journal (2017)

4.	Detection of plant leaf diseases using image segmentation and soft computing techniques [4]	Image acquisition here is the very first step that requires capturing an image with the help of a digital camera.	Image segmentation is used which is the process of separating or grouping an image into different parts. Here genetic algorithm is used which is an evolutionary algorithm.	Two classifiers K-Means and SVM are used for the classification purposes	Elsevier - Journal (2017)
5.	Factors influencing the use of deep learning for plant disease recognition [5]	The database consists of 50000 images of 171 diseases affecting 21 plant species.	NONE	Transfer learning applied to pre trained CNN model	Elsevier - Journal (2018)
6.	Plant disease and pest detection using deep learning-based features [6]	A self- made dataset of plant diseases common to Malatya, Bingöl, and Elazığ regions of Turkey.	The feature vectors of sizes :1024, 1000, 1000, 1000 ,4096 were obtained from certain layers of GoogleNet, ResNet50, ResNet101, InceptionV3, InceptionResNetV2, and SqueezeNet models.	The traditional classification algorithms like SVM, ELM and KNN were used on the extracted feature set.	Turkish Journal of Electrical Engineering & Computer Sciences – Journal (2019)
7.	Plant Disease Detection Techniques: A Review [7]	Discussed about the formation of dataset	Discussion about phases of plant disease detection that include: Image Acquisition, Image Segmentation, Feature Extraction and Classification.	It recommends to use SVM, Naïve Bayes and CNN for classification.	International Conference on Automation, Computational and Technology Management (ICACTM) – Conference (2019)

8.	Plant disease identification from individual lesions and spots using deep learning [8]	The database was captured using various sensors like cameras, smartphones etc. with resolution 1-24 MPixels.	NONE	GoogLeNet CNN was used for classification as well as detection.	Elsevier - Journal (2019)
9.	Plant Disease Detection Using Machine Learning [9]	Self-made dataset of leaves was used for disease detection.	HoG Feature Extraction technique.	RF, CART, logistic regression, SVM, KNN, Naive Bayes Were the machine learning algorithms used for classification.	International Conference on Design Innovations for 3Cs Compute Communicate Control – Conference (2018)
10.	Solving Current Limitations of Deep Learning Based Approaches for Plant Disease Detection [10]	Self-made large dataset combining the original images and images generated through GAN was used.	Convolution layer Of CNN were used for feature extraction.	32 layer fully connected classifier was used for classification.	Department of Industrial Engineering and Management, Faculty of Technical Sciences, Serbia-Journal
11.	A study on plant recognition using conventional image processing and deep learning approaches [11]	3 Standard datasets are used in here Folio Swedish Leaf Flavia	Local Binary Pattern and haralick are used for texture information. Pretrained models on VGG16, VGG19, inceptionV3 and inception Resnet-V2	CNN architecture with logistic regression as a classifier.	Journal of Intelligent & Fuzzy Systems 36 – Journal (2019)

12.	A study on the machine learning techniques for automated plant species identification: current trends and challenges [12]	The images can be captured using cameras inbuilt into mobile phones or using a digital camera.	The extraction can include both local and global descriptors. It can be extraction of colour descriptors, texture descriptors or contour based descriptors.	Numerous classification algorithms like k-NN, 1-NN, fuzzy k-NN, and SVM to name a few have been proposed.	Bharati Vidyapeeth's Institute of Computer Applications and Management 2019 - Journal
13.	A study on visual features of leaves in plant identification using artificial intelligence techniques [13]	Dataset of 637 healthy consisting of 32 different plant species is used.	22 visual features of each leaf are extracted and divided into 4 groups: dimension, color, texture and pattern	5 different algorithms were used for classification namely: ANN, Naive Bayes Algorithm, RF, KNN and SVM	Elsevier - Journal (2019)
14.	Fine-Grained Plant Identification using wide and deep learning model [14]	The Dataset of 100-Korean Plants is divided into two parts: Metadata that is used as an input in the linear model and image data used as input in the DCNN.	NONE	Uses combination of linear model and DCNN for classification	International Conference on Platform Technology and Service (PlatCon) – Conference (2019)

15.	Flower identification based on Deep Learning [15]	The seventeen flowers dataset by oxford university.	The feature here are extracted manually by normalising image into the same size	Deep convolutional neural network	IOP Conf. Series: Journal of Physics. - Journal (2019)
16.	Flower Classification with Deep CNN and Machine Learning Algorithms [16]	It uses two dataset oxford-17 and oxford-102 flowers	Deep CNN for extracting the features.	Various machine learning classifiers SVM, random forest and Multi-level perceptron (MLP)	Computer Engineering Department Istanbul University - Cerrahpasa Istanbul, Turkey- Conference
17.	Flower Species Recognition System using Convolution Neural Networks and Transfer Learning [17]	Oxford-17 flower dataset is extended with more images to a flower-28 dataset. Oxford 102 flower dataset is also used.	Multiple hand extracted techniques like colour histogram, GLCM, Hu and Zernike Moments, LBPs, HoG are used, followed by convolution layers of CNN.	The fully connected layers of CNN help in classification of the species.	4th International Conference on Signal Processing, Communications and Networking (ICSCN) 2017 - Conference
18.	Flower Classification via Convolutional Neural Network [18]	Online images of 79 species of flowers were collected. The dataset consisted of 63,442 images of flowers.	Stacked Convolutional Layers are used for feature extraction.	3 fully connected neuron layers, with the last one having 79 channels each corresponding to one class was used for classification.	IEEE International Conference on Functional-Structural Plant Growth Modelling, Simulation, Visualization and Applications-Conference

Chapter 3

REQUIREMENT ANALYSIS AND BRIEF DESCRIPTION

3.1 Overall Description of The Project

Technology today is present in all parts of our lives, and it is widely available to everyone. Almost 94% of farmers use mobile phones, especially in developing countries, where mobile phones may be the only available widespread computing and communication technology [18]. This number leads us to the conclusion that we can use mobile technology to help farmers improve their agricultural production. According to a study by the Associated Chambers of Commerce and Industry of India, annual crop losses due to pests and diseases amount to Rs.50,000 crore ($500 billion), which is significant in a country where at least 200 million Indians go to bed hungry every night [19][20]. The value of plant science is therefore huge. Therefore, there is a need for effective early disease detection techniques to control plant diseases for food security and sustainability of agroecosystem. This new real-time app and the model will attempt to solve the problem of identifying plants, the diseases in plants and the types of flowers using machine learning

Without proper identification of the disease, control measures can be a waste of time and money and can lead to further plant losses. So, the farmers can use the app to identify the diseases and could be cautious beforehand. The common diseases occurring in the plant could help the farmers in increasing the yield. Some of the plant diseases are due to unfavourable oxygen level, unfavourable water levels or rusts, fungus, smuts and white blisters can be responsible for degrading the plant quality, the best way to control it is early detection for the same and then eradicating the disease. Our real-time app could work well and detect these diseases and can help the farmers to attenuate the crop losses and maximise the yield.

Dataset:

1. Plant disease detection: The dataset used for plant disease detection consists of 73,515 plant images taking into consideration 18 plant species and 47 plant diseases

2. Flower Recognition: The dataset for this task includes 4242 images of 5 different flowers.

3.2 Requirement Analysis

Software Requirements:

- Operating System: Windows 7/8/10, Ubuntu
- Platform: Anaconda 5+ / Python 3.5+/ Jupyter Notebook
- Server Environment: NodeJS
- Text Editors/IDE: VS Code /Sublime Text/ PyCharm
- NoSQL document database: Google’s Firestore
- Mobile Application Framework: React-native

Python Libraries used:

1. **Keras:** Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, Theano, or PlaidML. We have worked with Keras in our previous projects.
2. **TensorFlow:** TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks.
3. **Numpy:** NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.
4. **Matplotlib:** It is a library used for plotting in python. It is used for static, animated and interactive visualization in python.
5. **Firebase-admin:** The Firebase Admin Python SDK which empowers backend (server-side) python engineers to coordinate Firebase into their administrations and applications.
6. **Firebase:** Python interface to the Google's Firebase REST APIs
7. **Pillow:** Python Imaging Library is a free and open-source additional library for the Python programming language that includes support for opening, manipulating, and saving the image file in different formats.

React-Native Libraries used:

1. **Expo:** Expo is an open-source platform for making universal native apps for Android, iOS, and the web with React and Javascript.
2. **React Native Firebase:** Collection of packages that brings React Native support for all Firebase services on both Android and iOS apps.
3. **React Navigation:** Routing and navigation for your React Native apps.
4. **Uuid:** Simple approach for the creation of creation of RFC4122 UUIDs.

Functional Requirements:

1. Collection of the dataset of plant diseases and flower species
2. Train the proposed CNN Model using this image dataset.
3. Building an application using react-native for user interaction.
4. Store the image inputted by the user in the Google Firestore Database.
5. Building Python scripts to download the image from Google firestore and apply the proposed model for disease detection.
6. Updating the predicted result back to the react-native application.

Non-Functional Requirements:

1) Accuracy: The CNN models must have good accuracy in predicting the diseases.
2) Time: The time taken by the model to predict the result should not be large enough as the user is waiting for the result.
Therefore, a perfect balance between the accuracy and the time needs to be maintained.

3.3 Solution Approach

Developing a cross-platform native mobile application for the detection of plants diseases and identification of the flower species by preparing a bridge between the front-end side of the application and the server-side where all the CNN models will be running, using Google's Firebase and Firestore.

Advantages of using CNN:

- CNN's use a black box technique to understand the pixels of an image. This technique is really useful to differentiate the diseases in plants as most of them look alike and have lesions.
- It is capable of giving a result with a high accuracy and in less time in case of image classification and hence was recommended for plant identification.
- A lot of work on Object detection and Identification has been done using a CNN we combine this approach for flower identification.

Our work is divided into three phases:

1. **Phase 1:** Building an App using react-native to input the image data from the user.
2. **Phase 2:** Store the data in the Google's database – Firestore.
3. **Phase 3:** Building a python server that downloads the image and feeds it into the CNN to predict the result.

Chapter-4

MODELING AND IMPLEMENTATION DETAILS

4.1 Design Diagram

The Design Diagram of the project is shown in Fig 2.

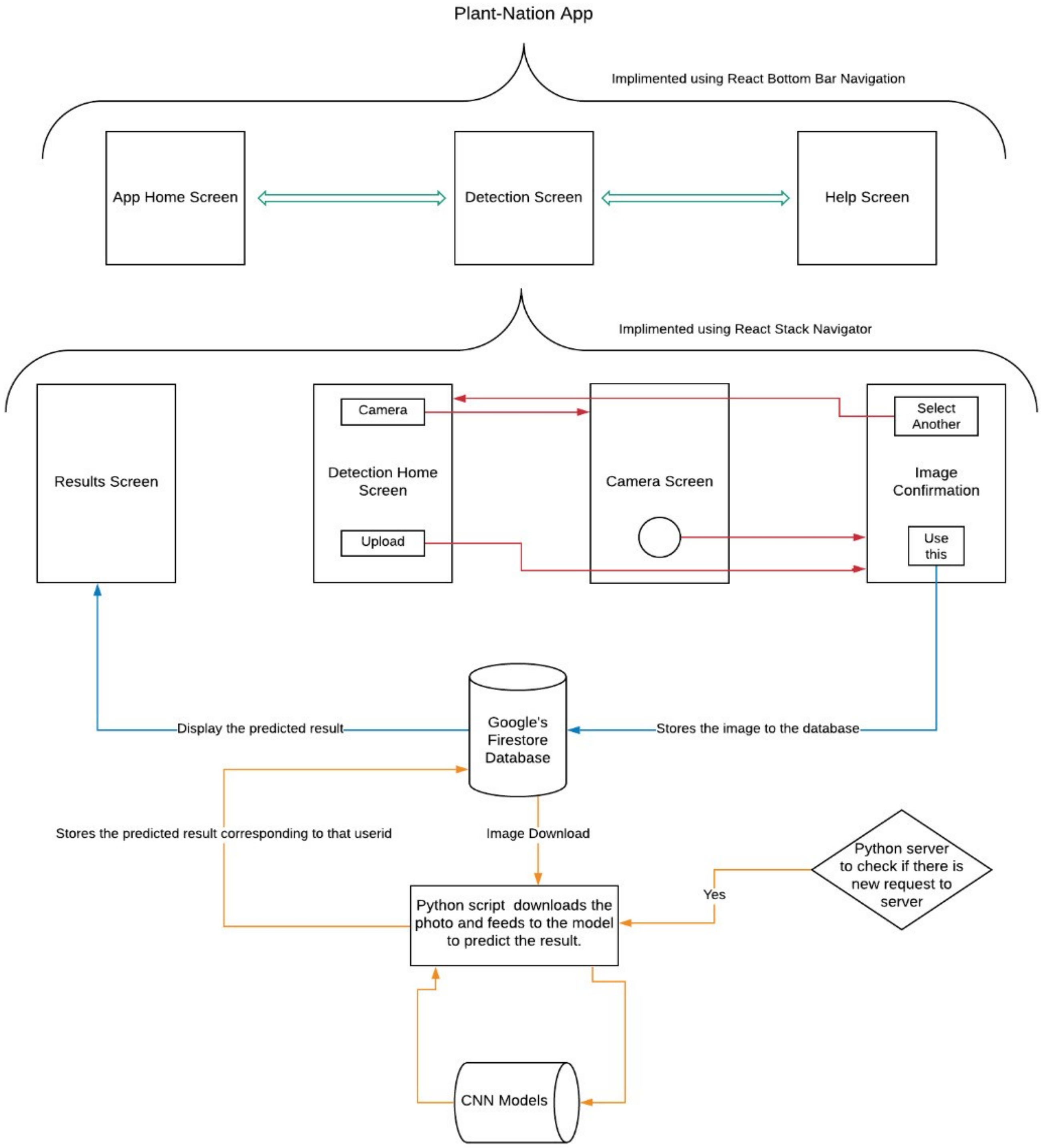


Fig 2: Design Diagram of the proposed model.

4.2 Implementation details and issues

4.2.1 Collection of Dataset

As a very first step, we started collecting the dataset of infected plant leaves. The most common dataset, on which most of the current disease detection apps are based on is the Plant-Village dataset. The plant-village-v1 dataset consists of 15 leaf diseases in 4 plants. Considering the vast variety of crops used in agriculture, this seems to be less. So, we decided to increase our dataset by combining more datasets into it. The second version of the dataset which consists of 38 leaf diseases found in 14 plant species is used. To further extend the dataset rice leaves dataset is added which consisted of 5 rice diseases. Also, the images of banana leaves are scrapped from bing images for 3 different diseases using a python script and added to the dataset. Therefore, a custom dataset is prepared which comprises 73515 images in a total of 46 diseases found in 16 plant species - Apple, Banana, Blueberry, Capsicum, Cherry, Corn, Grapes, Oranges, Peach, Potato, Raspberry, Rice, Soybean, Squash, Strawberry, Tomato.

For flower detection, we used the “5 category flower dataset” and added more scrapped images to get a total of 4242 images of 5 flowers which includes daisy, roses, tulip, sunflower and dandelion. The information assortment depends on sources like Flickr, Google pictures, Yandex pictures.

4.2.2 Proposed CNN Model

One significant favourable position of utilizing CNNs over NNs is that you don't have to level the info pictures to 1D as they are equipped for working with picture information in 2D. This aide in holding the "spatial" properties of pictures. A 3-layer CNN architecture is used for generating the pre-trained model for most of the major plant species individually and also a general single-layered pre-trained model with all the diseases of all the plants as the classes. This has been done to save time at the time of prediction because if the user has the knowledge about the type of plant then running a specific pre-trained model of that plant for the disease prediction will save time and give more accurate results.

4.2.3 Frontend

Using “react-native” as a framework and “expo” as a platform for making native apps, we started developing our app screen-by-screen. Firstly, the camera screen, then the image confirmation screen, then the detection home screen is prepared. After adding a few more screens, things started to get complicated and an easy way to navigate between the screens is needed. Therefore, React Navigation, a routing and navigation package for the react native apps, is used to switch between the screens. A bottom tab navigator is used at the upper layer of the app which includes the home screen, all the detection screens (disease, flower, rotten) and the help screen. The lower layer uses the react stack navigator to navigate from the camera screen to image confirmation screen to the results screen when initiated from any detection screen.

4.2.4 Backend

So when the user confirms the selected image from the image confirmation screen then the unique id is generated and a request is generated which uploads the image to the Firestore's database along with information like the image height, image width, type of detection and the unique prediction id.
A python server is running at the backend. Whenever a new photo gets uploaded at the Firestore, it immediately downloads the photo and feeds it to the CNN pre-trained model and the predicted result gets updated corresponding to that id's entry in the Firestore. As soon as the result entry modifies at the database, the result screen at the frontend displays the predicted disease and delete that entry from the database.

It is very important to find the **right balance between the accuracy and the time** a model takes to predict, as the user is waiting at the frontend for the result and he/she wants the result to be accurate but doesn't want to wait long.

Issues:

1. Our baseline CNN model follows a very simple architecture.

2. For further predicting the result with improved accuracy in much less time, a machine with better hardware to host the server is needed.

4.3 Risk Analysis and Mitigation

The risk analysis and mitigation can be seen in Table 2.

Table 2: Risk Analysis and Mitigation

Risk ID	Description of risk	Risk Area	Probability	Impact	Risk selected for mitigation	Mitigation on plan
1.	Software not responding or system shuts down	System and software failure	H	H	YES	Worked on a computer with better RAM
2.	Database File gets corrupt	Losing all data stored in the database .	M	H	YES	Always keep the backup of the database.
3.	Overfitting of model happens	Training model using early-stop function	M	H	YES	Applied feature generation Technique
4.	Feature vector can be corrupted	Software	L	H	YES	All the vectors are stored in .pkl file
5.	Issue in loading data	Project development	M	H	YES	Several time model has been loaded
6.	All paths not tested	Inefficient path recommendation	L	H	YES	Setting the counter to minimum 10.

Chapter-5

TESTING

5.1 Testing Plan

Since the project is about improving the performance of the CNN base model with newly generated images, hence the focus of testing will be on tuning the hyperparameters and obtaining results with different sets of images containing an increasing number of images.
The testing plan is shown in Table 3.

Table 3: Testing Plan

Type of Test Will	**Will Test Be Performed?**	**Explanations**	**Software Component**
Requirement Testing	Yes	To check the feasibility of our project in terms of budget, requirements, etc. this will be done.	Anaconda
Unit	Yes	Individual units of source code will be run with operating procedures to see if they are fit to use.	Anaconda
Integration	Yes	The built test cases and test data are integrated and then predictions are made. The bugs (if found) are fixed, model is re-tested.	Anaconda

Performance	Yes	To test whether our project will work well under expected workload, this is a must.	Browser, data fields etc.
Stress	Yes	Stress Yes To compare what our model predicts and what is ground truth, this test has to be performed.	Tables in database with heavy data files.
Security	No	Not required	Not Attempted
Compliance	No	Not required	Not Attempted
Load	Yes	In future, to check the performance of our project under real-life load, this will be done.	Not done.
Volume	No	Not required	Not Attempted

5.2 Component Decomposition and Type of testing required

The components involved in the project is shown in Table 4.

Table 4: Component Decomposition and Type of Testing.

S.No.	List of various functions that require testing	Type of testing required	Technique for writing cases
1.	Validation	Integration	White box
2.	Loss vs Validation loss	Integration	White Box
3.	Testing Predictions generated	Integration	White Box
4.	System Testing	System	Black box
5.	Data Storage and Transmission	Unit, Security, Performance	Black Box
6.	Application Interface	Integration	Black Box

Types of testing:

- Ad-hoc testing: The purpose of this testing is to find any faults, errors, and defects in our Neural Network Models.
- Black Box testing: Since we don’t know the exact higher-order representations of images in deeper levels of CCN, so we use this type of testing to check how a given input performs with a particular output. Also, in GAN we can’t tell for sure when and where the Adversarial network counteracts the Generator Network, so in that case, also we test the model by input and output images.
- Comparison Testing: The entire results and analysis of our model are dependent upon the comparison of base case results with Augmented results, so by comparing them both we can make sure that the model is working fine.
- Unit Testing: Since our Model consist of multiple Neural Networks, we have to test each sub-network individually.
- Component Testing: After testing all the single components we tested all the models together to make them work synchronously and get the results.

5.3 Testing Results

Fig 3 and Fig 4 shows the plot of loss function and accuracy with each epoch respectively.

```
plt.plot(History.history['loss'])
plt.plot(History.history['val_loss'])
plt.title('Model Loss')
plt.ylabel('Loss')
plt.xlabel('Epochs')
plt.legend(['train', 'test'])
plt.show()
```

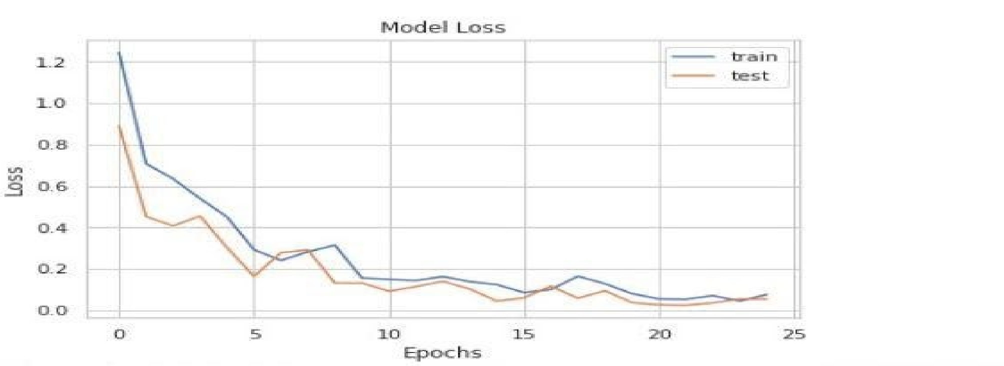


Fig 3: Plot of Loss Function with each epoch

```
plt.plot(History.history['accuracy'])
plt.plot(History.history['val_accuracy'])
plt.title('Model Accuracy')
plt.ylabel('Accuracy')
plt.xlabel('Epochs')
plt.legend(['train', 'test'])
plt.show()
```

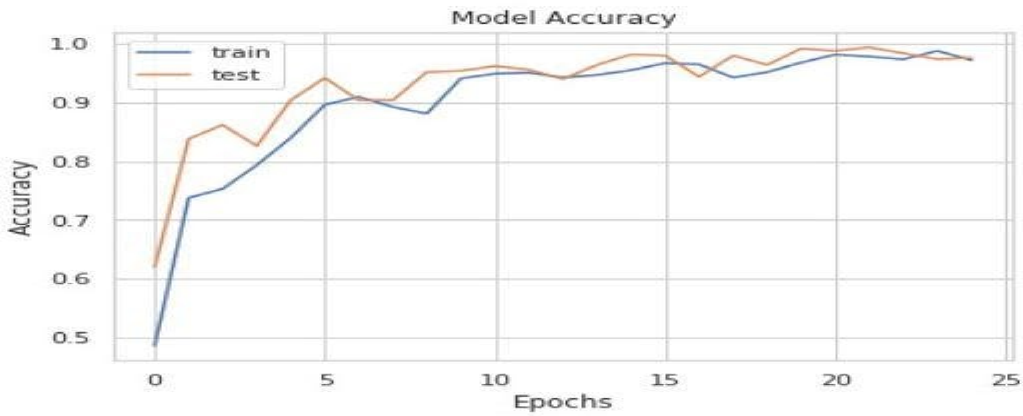


Fig 4: Plot of accuracy with each epoch

5.4 Error and Exception Handling

The errors that we encountered were in following domains:

- Image conversion to numpy files
- Image resolution(shape)

Although we solved all these errors, we did encounter some exceptions such as a few multichannel images

5.5 Limitations

There are certain limitations and challenges of the solution as below:

- A vast variety of flora and fauna are present therefore, it becomes difficult to identify all the flowers accurately.
- The variations in diseases in the plant also make it difficult to effectively predict it.
- Some large similarities that occur in nature make it tough to maintain efficacy in the model.
- There are very few datasets available in this field. This hinders our model to be rigorously tested upon various categories of people.

Chapter-6

FINDINGS, CONCLUSION AND FUTURE WORK

6.1 Findings

The accuracy of diseases for each of the species is shown in table 5.

Table 5: Accuracy of detection of disease in each plant

Apple	95.2 %
Corn	94.4 %
Potato	98.2 %
Grapes	93.4 %
Corn	93.9 %
Rice	98.2 %
Tomato	93.6 %
Banana	85.6 %
General Model including all	89.8 %

Here the general model including all the diseases of all the plants is trained only on the single-layered CNN model for quicker results. The accuracy of the classification of the flowers is 90.3 % when trained over the 3 layered CNN model.

The training of model for the grapes diseases is shown in Fig 5. Fig 6 and Fig 7 shows how the data and the images get stored on the Google Firestore cloud. Working of application on the android device is shown in Fig 8 (i-vi).

```
5408/5408 [==============================] - 3s 576us/step - loss: 0.1961 - accuracy:
0.9329 - val_loss: 0.2089 - val_accuracy: 0.9164
Epoch 7/10
5408/5408 [==============================] - 3s 600us/step - loss: 0.1658 - accuracy:
0.9456 - val_loss: 0.2115 - val_accuracy: 0.9112
Epoch 8/10
5408/5408 [==============================] - 3s 580us/step - loss: 0.1562 - accuracy:
0.9462 - val_loss: 0.1893 - val_accuracy: 0.9283
Epoch 9/10
5408/5408 [==============================] - 3s 581us/step - loss: 0.1355 - accuracy:
0.9562 - val_loss: 0.1865 - val_accuracy: 0.9223
Epoch 10/10
5408/5408 [==============================] - 3s 584us/step - loss: 0.1238 - accuracy:
0.9610 - val_loss: 0.1767 - val_accuracy: 0.9349
```

In [8]:

```python
print("[INFO] Saving model...")
# pickle.dump(model,open('plant-village-model.pkl', 'wb'))
model.save('model-grape128.h5')
```

```
[INFO] Saving model...
```

Fig 5: Snapshot of training the model

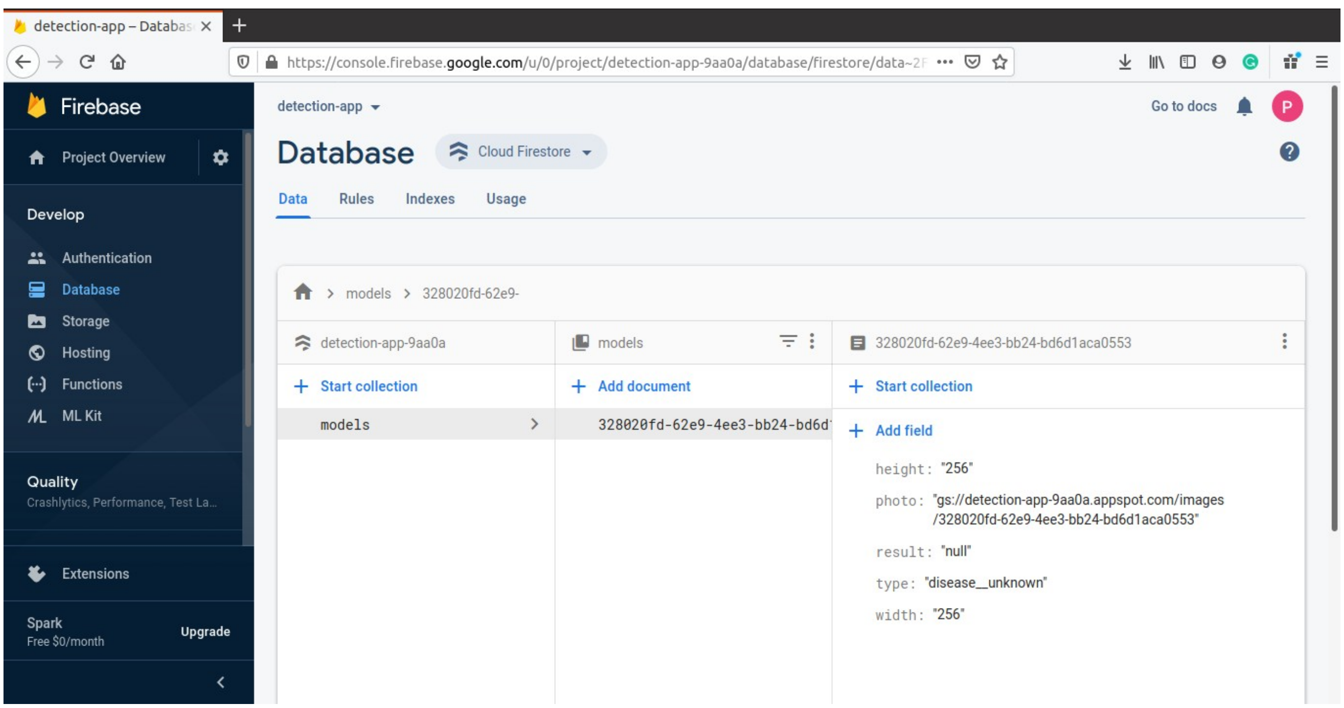


Fig 6: Snapshot of the Google Firestore database

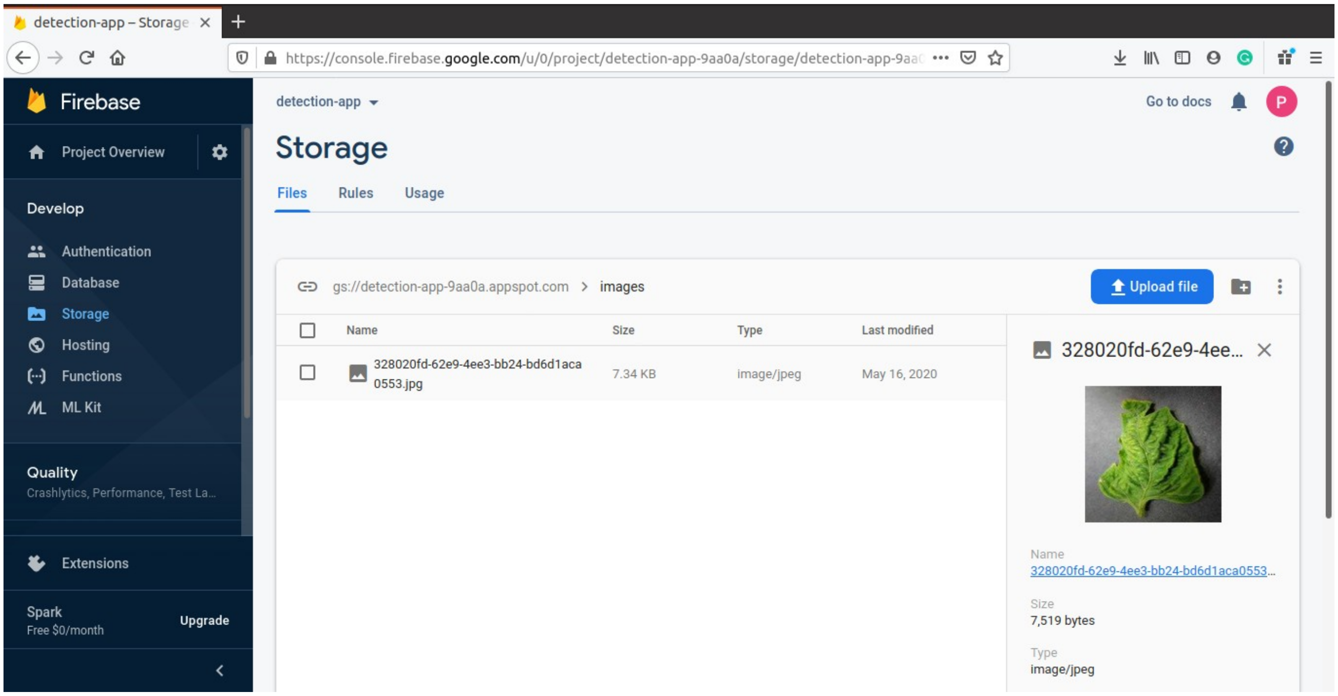
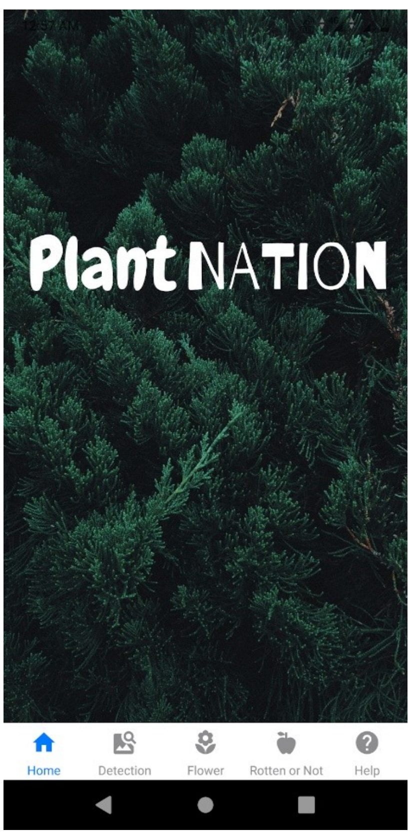
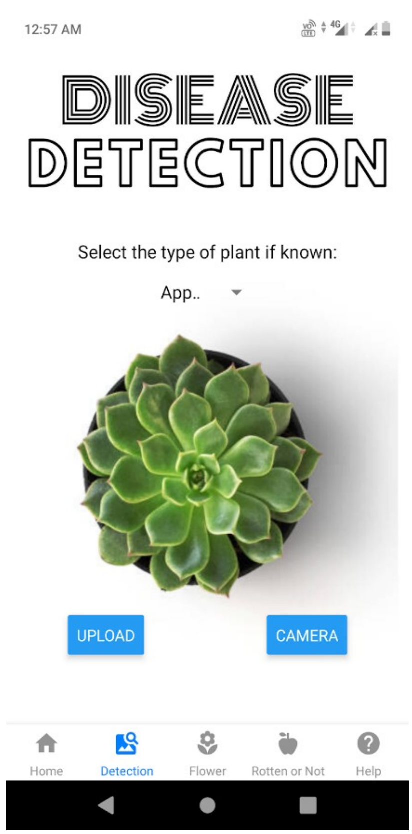


Fig 7: Snapshot of the Google Firestore storage

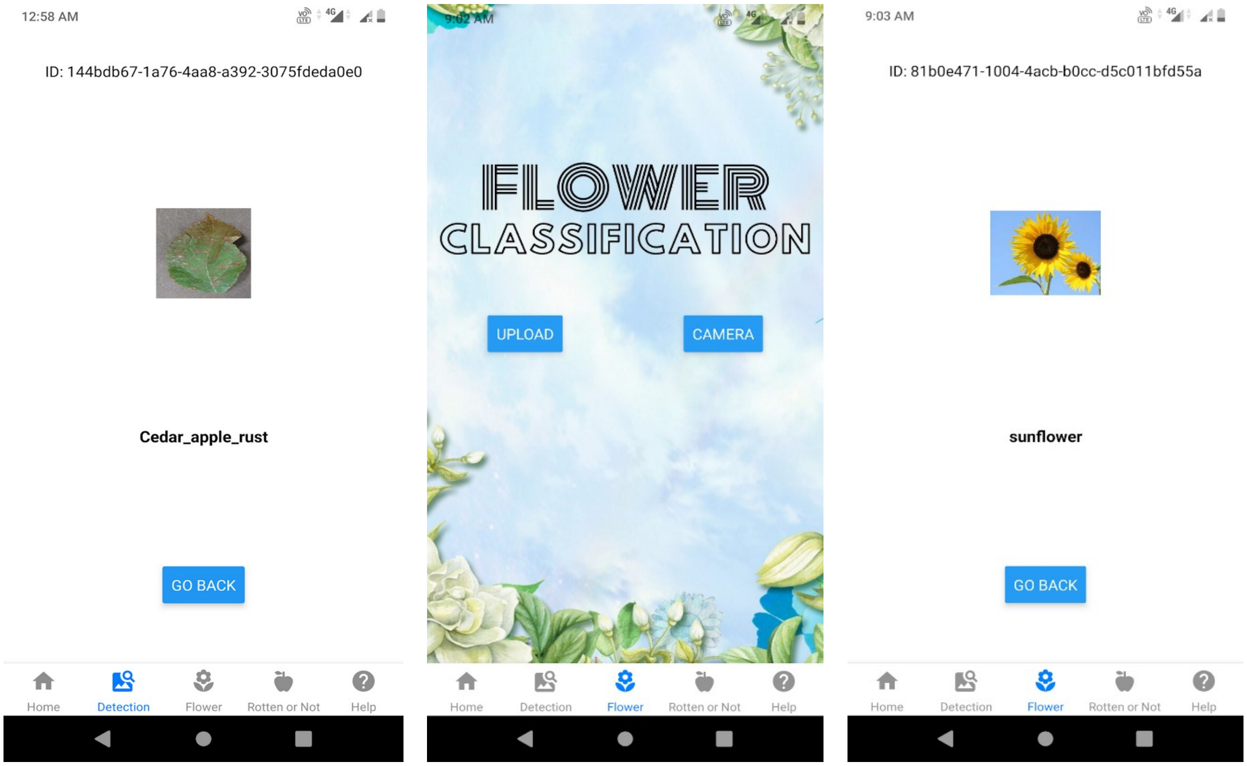


i) App's Home Screen



ii) Detection Screen 1

iii) Image Confirmation Screen



iv) Results Screen 1 **v) Detection Screen 2** **vi) Results Screen 2**

Fig 8: Screenshot of working of application (i-vi)

6.2 Conclusion

In this project, an automated system is developed with react-native that predicts the diseases in plants. The algorithm is validated for 46 diseases in 16 plant species Apple, Banana, Blueberry, Capsicum, Cherry, Corn, Grapes, Oranges, Peach, Potato, Raspberry, Rice, Soybean, Squash, Strawberry, Tomato. The general accuracy for the disease is 89.8 %. While the reported results are very encouraging. The flower detection works on “5 category flower dataset ” total of 4242 images of 5 flowers which includes daisy, roses, tulip, sunflower and dandelion. The algorithm has been deployed on a real smartphone application. The results obtained until now are encouraging with good accuracy but still, there is a scope of increasing the accuracy by using various other machine learning models.

6.3 Future Work

Traditionally, methods of plant disease identification were developed using different models, but these models could not cover all the disease in the species. In our model, we have tried to cover major plants and their diseases but the number of flora and fauna species is vast and developing a model for all is very difficult but we'll try to add new models for different plant species or flower species in the near future. Secondly, the backend python server which runs CNN models to predict the result needed to be hosted online. Currently, the backend python server is running on a local device.

REFERENCES

[1]. Zhang X, Han L, Dong Y, Shi Y, Huang W, Han L, González-Moreno P, Ma H, Ye H, Sobeih T. A deep learning-based approach for automated yellow rust disease detection from high-resolution hyperspectral UAV images. Remote Sensing. 2019 Jan 11:1554.

[2]. Shruthi, U., V. Nagaveni, and B. K. Raghavendra. "A review on machine learning classification techniques for plant disease detection." In 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), IEEE.281-284.

[3]. Johannes, Alexander, Artzai Picon, Aitor Alvarez-Gila, Jone Echazarra, Sergio Rodriguez-Vaamonde, Ana Díez Navajas, and Amaia Ortiz-Barredo. "Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case." Computers and electronics in agriculture 138(2017). 200-209.

[4]. Singh, Vijai, and Ak K. Misra. "Detection of plant leaf diseases using image segmentation and soft computing techniques." Information processing in Agriculture 4, no. 1 (2017): 41-49.

[5]. Barbedo, Jayme GA. "Factors influencing the use of deep learning for plant disease recognition." Biosystems engineering 172 (2018): 84-91.

[6]. Türkoğlu, Muammer, and Davut Hanbay. "Plant disease and pest detection using deep learning-based features." Turkish Journal of Electrical Engineering & Computer Sciences 27, no. 3 (2019): 1636-1651.

[7]. Sandhu, Gurleen Kaur, and Rajbir Kaur. "Plant Disease Detection Techniques: A Review." In 2019 International Conference on Automation, Computational and Technology Management (ICACTM), IEEE.2019 pp.34-38

[8]. Barbedo, Jayme Garcia Arnal. "Plant disease identification from individual lesions and spots using deep learning." In 2019 Biosystems Engineering 180: 96-107.

[9]. Maniyath, S. R., Vinod, P. V., Niveditha, M., Pooja, R., Shashank, N., & Hebbar, R. Plant disease detection using machine learning. In 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), IEEE. 2018 pp. 41-45

[10]. Arsenovic, M., Karanovic, M., Sladojevic, S., Anderla, A., & Stefanovic, D. Solving current limitations of deep learning based approaches for plant disease detection. In 2019, Symmetry, 11(7), 939.

[11]. Anubha Pearline, S., Sathiesh Kumar, V., & Harini, S. A study on plant recognition using conventional image processing and deep learning approaches. In 2019, Journal of Intelligent & Fuzzy Systems, 36(3), 1997-2004.

[12]. Bojamma, A. M., & Shastry, C. A study on the machine learning techniques for automated plant species identification: current trends and challenges. In 2019, International Journal of Information Technology, 1-7.

[13]. Yigit, E., Sabanci, K., Toktas, A., & Kayabasi, A. A study on visual features of leaves in plant identification using artificial intelligence techniques. In 2019, Computers and electronics in agriculture, 156, 369-377.

[14]. Lee, J. W., & Yoon, Y. C. "Fine-Grained Plant Identification using wide and deep learning model". In 2019 International Conference on Platform Technology and Service (PlatCon)IEEE. 2019 pp. 1-5

[15]. Tian, Mengxiao, Hong Chen, and Qing Wang. "Flower identification based on Deep Learning." In Journal of Physics: Conference Series, vol. 1237, no. 2, p. 022060. IOP Publishing, 2019.

[16]. Mete, B. R., & Ensari, T. Flower Classification with Deep CNN and Machine Learning Algorithms. In 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)IEEE,2019 pp.1-5

[17]. Gogul, I., & Kumar, V. S.. Flower species recognition system using convolution neural networks and transfer learning. In 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN) IEEE,2017 1-6

[18]. Liu, Yuanyuan, Fan Tang, Dengwen Zhou, Yiping Meng, and Weiming Dong. "Flower classification via convolutional neural network." In 2016 IEEE International Conference on Functional-Structural Plant Growth Modeling, Simulation, Visualization and Applications (FSPMA), IEEE 2016 pp. 110-116.

[19] Malcolm, Shurtleff-Michael, Pelczar-Rita, Pelczar-Arthur, Kelman. “Plant Disease,” April, 2020. [Online]. Available: https://www.britannica.com/science/plant-disease .

[20] The Economic Times, “Crops worth Rs 50,000 crore are lost a year due to pest, disease: Study,” Feb, 2014. [Online]. Available: https://economictimes.indiatimes.com/news/economy/agriculture/crops-worth-rs-50000-crore-are-lost-a-year-due-to-pest-disease-study/articleshow/30345409.cms

[21] Croplife International, “India's Farmers Fighting Pests,” May, 2015. [Online]. Available: https://croplife.org/news/keeping-indias-pests-in-line/

[22] Tanja Folnovic. “Smart Agriculture on Smartphones,” [Online]. Available: https://blog.agrivi.com/post/smart-agriculture-on-smartphones

[23] Parnell, S., Gottwald, T. R., Cunniffe, N. J., Alonso Chavez, V., & van Den Bosch, F. (2015). Early detection surveillance for an emerging plant pathogen: a rule of thumb to predict prevalence at first discovery. Proceedings of the Royal Society B: Biological Sciences, 282(1814), 20151478.

[24] Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., ... & Davis, C. E. (2015). Advanced methods of plant disease detection. A review. Agronomy for Sustainable Development, 35(1), 1-25.