A Disease Prediction and Rectification System for Banana Plant

1Prof.D.D.Pukale, 2Gunjan Kokru, 3Sneha Nadar, 4Sanskriti Dhar, 5Shivangi Singh

Abstract—Banana is one of the major and economically important fruit crop in India. Banana occupy 20% of the total area of crop in India. In India banana is grown under diverse conditions and production systems. This system focuses to identify, detect and rectify the diseases in banana plant and also continuously providing updates about the diseases in the leaves of banana plant to the farmer. Here, the system will be provided with the input as regular images of banana plant captured through different image capturing medias, and the system will further process those images to detect the disease if any and then notify the farmer as well. The system will also guide the farmer about the further actions to be taken such as suggesting him with the right pesticides, fertilizers to use and farming techniques so that diseases will be cured and would not corrupt the crops nearby. Therefore, the further yield of his crop will reach the maximum level and the disease would not replicate in the future.

Keywords— Convolutional Neural Network (CNN), K- NN

1. INTRODUCTION

Banana plant have a wide range of applications because they are large, flexible, waterproof and decorative. They are used for cooking, wrapping and food serving in a wide range of cuisines. Banana leaf is full of antioxidants and eating hot, freshly cooked foods on the Banana leaf is one good way to get all the antioxidants easily. Therefore its of very high importance to detect diseases in Banana plant through an automatic system .Some of the Banana plant diseases are Panama disease, Wilt disease, Sigatoka disease, Aphid, banana streak virus and banana mosaic virus diseases. Hence it would be very beneficial to use a automatic system that will detect diseases of Banana leaves and provide with remedial actions to be taken.

A. Problem Statement

To evolve an efficient system for banana leaf disease detection, after procuring the images from a banana farm as well as providing remedial measures for the detected disease using CNN. The system will also notify the farmer about any infected plant in the crop

B. Objective

To help farmers to keep their banana crop healthy and continuously provide then updates about any disease that occurred in the banana leaves. Also provide farmers with remedial actions against the disease predicted. So that the further yield of his crop will reach the maximum level and the disease would not replicate in the future.

2. METHODOLOGY

This project has been divided into 2 phases. First, literature study is conducted, followed by system development. Literature study involves conducting studies on various disease detection techniques for banana plant and method that currently in use. In phase 2, application requirements and functionalities are defined prior to its development. Also, architecture and interface design of the program and how it will interact is also identified. In developing the disease prediction system, several tools are utilized, such as Eclipse and MYSQL Workbench.

3.LITERATURE SURVEY

a. A Deep Learning-based Approach for Banana Leaf Diseases Classification:

Description: Apply deep neural networks to detect two famous banana diseases which are banana sigatoka and
Banana speckle in real scene and under challenging conditions such as illumination, complex background,
Different images resolution, size, pose and orientation.

b. Banana Plant Disease Detection and Grading Using Image Processing :

Description: Software solution for automatic plant disease detection and finally the percentage infection using an image processing technique. The proposed work uses Artificial Neural Network to classify the Banana plant diseases. The proposed system involves several steps, which include- dataset creation, image pre-processing, HOT feature extraction and artificial neural network based training and classification.

c. Detection and Prevention Of Banana Leaf Diseases From Banana Plant Using Embedded Linux Board

Description: The detection and prevention of banana streak viral disease is carried out using an Embedded Linux development board interfaced with a camera, which is used to capture the leaf of banana plant. The captured image is to be processed by an algorithm called Economic Threshold Level (ETL). This algorithm is used to set the threshold value of a healthy banana leaf of the captured image.

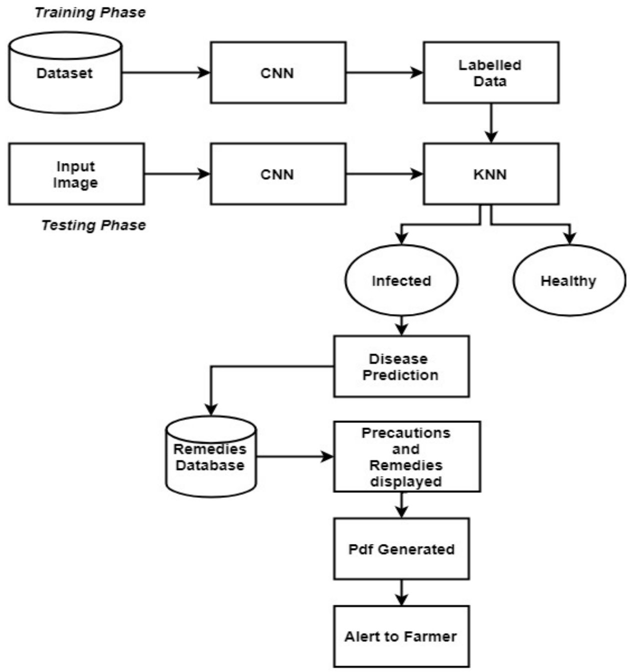
d. A Survey on Methods of Plant Disease Detection

Description: Agriculture is important sector in Economy and Social life. Earlier unscientific methods were in existence. Study and evaluation of existing techniques for detection of plant diseases to get clear outlook about the techniques and methodologies followed.

e. Study on Banana Leaf Disease Identification Using Image Processing Methods

Description: Gives a brief review about major banana plant diseases that show symptoms in leaves and explains in detail the image processing techniques that are involved in the process of disease identification in banana leaves.

4. SYSTEM ARCHITECTURE



5. IMPLEMENTATION

1:DataSet Our dataset is 800 banana plant images Reference: www.Github.com , Google images.

APHID:

APHID

Damage plants in three ways: (i) they take nutrients and water from the plants by feeding on their sap, (ii) by producing honeydew that attracts dark fungi called sooty moulds, and (iii) by spreading viruses.

BSIGATOKA:

Bisigatoka

The first visible symptom is a slight discoloration between the leaf's secondary veins (A). Over time, these points develop into pale yellow streaks, brown streaks and elliptic necrotic spots arranged parallel to the secondary veins.

YSIGATOKA:

Bacteria:

Bacteria

In early stage of infection dark brown or yellow water soaked areas are more in the cortex area When affected plants are cut open at collar region yellowish to reddish ooze is seen.

Mosaic:

Mosaic

The virus is known to infect only banana and can be transmitted in a nonpersistent manner

Streak:

Streak

A prominent symptom exhibited by BSV is yellow streaking of the leaves, which becomes progressively necrotic producing a black streaked appearance in older leaves.

Panama wilt:

Panama Wilt

The disease is soil borne and the fungus enters the roots through the fine laterals.

2: Phases during system implementation

1] Training:

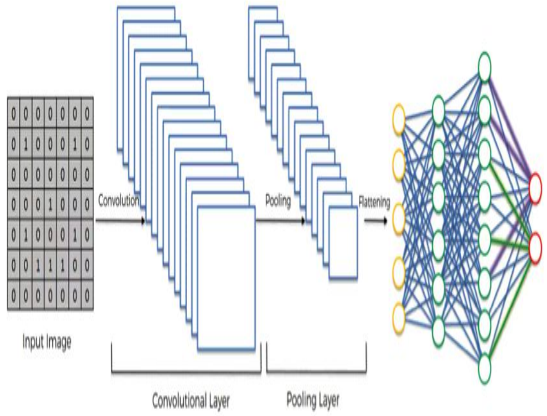
A file named train.java is executed for training the dataset.

Steps in train.java file:

1: During training time images label are known to check the accuracy of the images after cnn layer, so we store the labels of the images with each image id.

2: We convert the image in byte format to buffered images and resizing is performed.

Steps in CNN file:

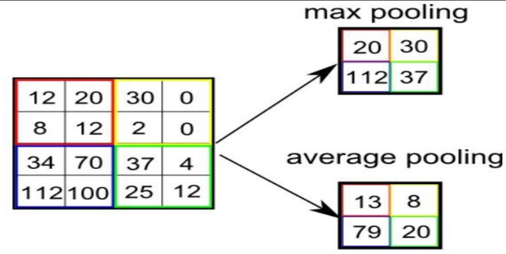


CNN

1: We resize the image width and height and pass to the convolution layer.

2: **Convolution map**: The goal of convolution is to extract features from the input image. It consists of a set of learnable filters. Each filter is applied to the raw pixel values of the image taking into account the **red, green and blue** color channels in a sliding window fashion, computing the dot product between the filter pixel and the input pixel. This will result in a 2-dimensional activation map of the filter called the feature map. Hence, the network learns filters (ie. edges, curves, veins, color, intensity) that will activate when they find known features in the input. **The CNN learns the values of these filters on its own during the training process**. The Convolution operation is presented in Equation 1. A convolution layer is configured by the number of convolution maps it contains Mi , the size of the filters which are often squared kx ⋆ky (3X3). The feature map Mi is computed as follows: **Mi = bi +∑ k Wik ⋆Xk (1)** where ⋆ is the convolution operator, Xk is the k th input channel, Wik is the sub kernel of that channel and bi is a bias term(ReLU). In other words, the convolution operation being performed for each feature map is the sum of the application of k different 2D squared convolution features plus a bias term. Moreover, the rectified nonlinear activation function (ReLU) is performed after every convolution to introduce nonlinearity to the CNN. The **ReLU** is a very popular activation function which is defined as f(x) = max(0, x) where x is the input to a neuron [1].

3. Max-pooling map: In the architecture of convolutional neural network, convolution layers are followed by sub-sampling layers. Each sub-sampling layer reduces the size of the convolution maps, and introduces invariance to (low) rotations and translations that can appear in the input. A layer of max-pooling is a variant of such layer that has shown different benefits in its use. The output of max-pooling layer is given by the **maximum activation value** in the input layer over sub windows within each feature map. The max-pooling operation reduce the size of the feature map[1].

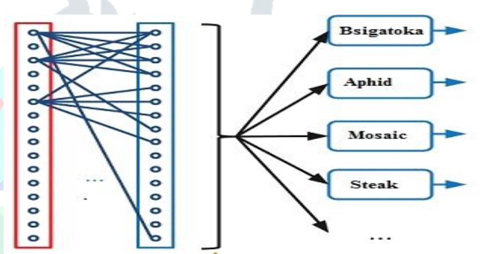


CNN layer 2

4. **Classification model**: Within the classification step we use fully connected layers where each neuron provides a full connection to all learned feature maps issued from the previous layer in the convolution neural network. These connected layers are based on the **SIGMOID** activation function in order to compute the classes scores by converting the 3d matrix to 1d feature vector. The input of the softmax classifier is a vector of features resulting from the learning process and the output is a probability that an image belongs to a given class.

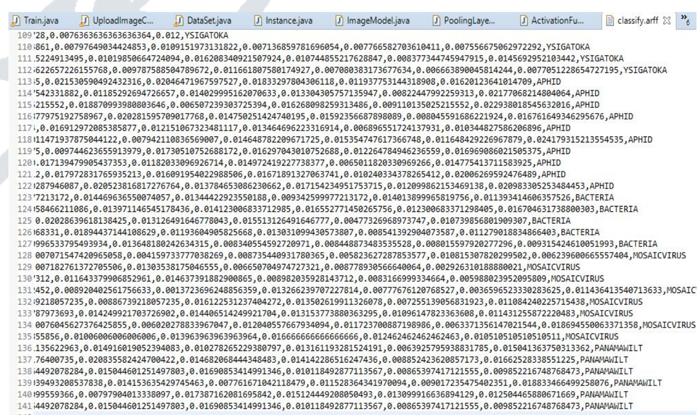
$$h_\theta(x) = \frac{1}{1 + \exp(-\theta^T x)},$$

The features are saved in the classify.arff file with the categories(Streak, Aphid, Healthy .. etc). The actual labels are known they are compared with the classify.arff. file.



CNN layer 3

Classify.arff file:




```
'28,0.007636363636363636364,0.012,YSIGATOKA
861,0.00797649034424853,0.0109151973131822,0.007136859781696054,0.007766582703610411,0.007556675062972292,YSIGATOKA
.5224913495,0.01019850664724094,0.016208340921507924,0.010744855217628847,0.008377344745947915,0.014569295210344 2,YSIGATOKA
62265722615576 8,0.009787588504789672,0.011661807580174927,0.007080383173677634,0.006663890045814244,0.007705122865472719 5,YSIGATOKA
5,0.021530590492432316,0.020464719675975 27,0.018332978043061 18,0.011937753144318908,0.0162012364101470 9,APHID
'542331882,0.011852926947266 57,0.014029995162070633,0.013304305757135947,0.008224479922593 13,0.021770682148040 64,APHID
215552,0.018870993980803646,0.006507239303725394,0.016268098259313486,0.009110135025215552,0.022938018545632016,APHID
77975192758967,0.020281595709017768,0.014750251424740195,0.015923566878980 89,0.008045591686221924,0.016761649346295676,APHID
,0.016912972085385877,0.012151067323481117,0.013464696223316914,0.006896551724137931,0.010344827586206896,APHID
114719378750441 22,0.007942110836569007,0.014648782209671725,0.015354747617366748,0.011648429226967879,0.024179315213554535,APHID
'5,0.009744623655913979,0.017305107526881 72,0.016297043010752688,0.012264784946236559,0.016969086021505375,APHID
.01713947990543735 3,0.011820330969267 14,0.014972419227738377,0.006501182033096926 6,0.014775413711583925,APHID
.2,0.017972831765935213,0.016091954022988506,0.016718913270637 41,0.010240334378265412,0.020062695924764 89,APHID
287946087,0.020523816817276764,0.013784653086230662,0.017154234951753715,0.012099862153469138,0.020983305253484453,APHID
'7213172,0.014469636550074057,0.013444229235501 88,0.009342599977213172,0.014013899965819756,0.011393414606357526,BACTERIA
58466211086,0.013971146545178436,0.014123006833712985,0.016552771450265756,0.012300683371298405,0.016704631738800303,BACTERIA
0.020286396181384 25,0.013126491646778043,0.015513126491646777,0.004773269689737 47,0.010739856801909307,BACTERIA
68331,0.018944371441086 29,0.011936049058256 68,0.013031099430573807,0.008541392904073587,0.011279018834866403,BACTERIA
996533795493934,0.013648180242634315,0.008340554592720971,0.008448873483535528,0.008015597920277296,0.009315424610051993,BACTERIA
0070715474209650 58,0.004159733777038269,0.008735440931780365,0.005823627287853577,0.010815307820299502,0.006239600665557404,MOSAICVIRUS
0071827613727055 06,0.013035381750465 55,0.006650704974727321,0.008778930566640064,0.002926310188880021,MOSAICVIRUS
'312,0.011643379906852961,0.014637391882900865,0.008982035928143712,0.0083166999334664,0.005988023952095809,MOSAICVIRUS
452,0.008920402561756633,0.001372369624885635 9,0.013266239707227814,0.007776761207685 27,0.003659652333028362 5,0.011436413540713633,MOSAIC
9218057235,0.00886739218057235,0.016122531237404272,0.013502619911326078,0.007255139056831923,0.011084240225715438,MOSAICVIRUS
'87973693,0.014249921703726902,0.014406514249921704,0.013153773880363295,0.010961478233636 08,0.011431255872220483,MOSAICVIRUS
00760456273764258 55,0.006020278833967047,0.012040557667934094,0.011723700887198986,0.006337135614702154 4,0.018694550063371358,MOSAICVIRUS
55856,0.01006006006006006,0.013963963963963964,0.016666666666666666,0.012462462462462463,0.010510510510510511,MOSAICVIRUS
.135622963,0.014916019052394083,0.010278265229380797,0.013161193281524191,0.006392579593883178 5,0.015041363750313362,PANAMAWILT
.76400735,0.020835582424700422,0.014682068444348483,0.014142286516247436,0.008852423620857173,0.016625283385512 25,PANAMAWILT
4492078284,0.015044601251497803,0.016908534149913 46,0.010118492877113567,0.008653974171215 55,0.009852216748768473,PANAMAWILT
39493208537838,0.014153635429745463,0.007761671042118479,0.011528364341970094,0.009017235475402351,0.018833466499258076,PANAMAWILT
99559366,0.007979040133380 97,0.017387162081695842,0.015124449208050493,0.013099166368941 29,0.012504465880671669,PANAMAWILT
4492078284,0.015044601251497803,0.016908534149913 46,0.010118492877113567,0.008653974171215 55,0.009852216748768473,PANAMAWILT
```

Arff file

2] Testing:

Steps:

1: Login or Register: User can register and use the system. The user is given the unique id for maintaining his session in the system, so that he can perform prediction many times.

2: Image Acquisition: The images of the plant are captured through image capturing media. This image is saved in the database with unique id which in byte format, For further processing we can get the image and resize the image into 32 bit.

3: Then it is passed to same CNN layers as used in training phase.

4: But, here no classify.arff is generated.

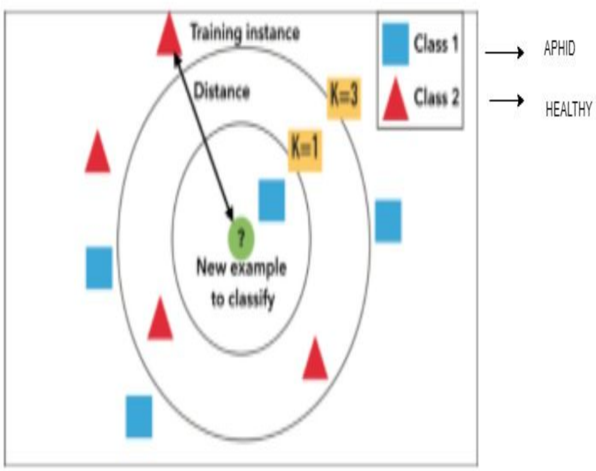
5: The extracted features goes to KNN layer.

6: KNN: In this module we calculate the Euclidean distance between the features of Input Image and Trained dataset and allot them their probable classes based on minimum distance.

$$d(\mathbf{p},\mathbf{q}) = d(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + \cdots + (q_n-p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}.$$

q(1…n)=Trained Images features one vector
p(1)=Tested feature vector

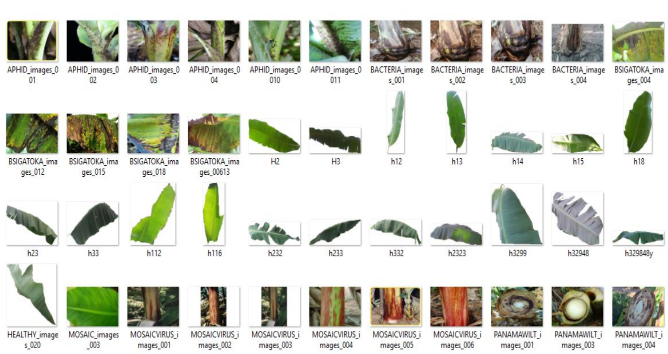


KNN

Prediction of banana leaf and stem disease and suggest the symptom and remedies base on predicted category of disease finally generate and PDF report send it through mail on user Mail-Id.

6. RESULT

Training Dataset:



Confusion Matrix:

- **true positives (TP):** These are cases in which we predicted and actual yes (they have the disease), and (they do have the disease).
- **true negatives (TN):** We predicted no, and they don't have the disease.
- **false positives (FP):** We predicted yes, but they don't actually have the disease.
- **false negatives (FN):** We predicted no, but they actually do have the disease.

Confusion Matrix Calculation:

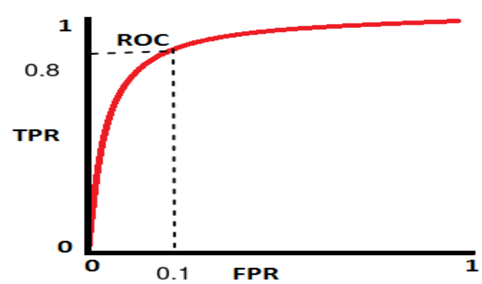
Total images=55,Healthy=19,Infected=36
True Positive=(15+27) =42 True Negative=(3+2) = 5
False Positive=(1) =1 False Negative=(7)

	predicted no	predicted yes	
actual no	TN=5	FP=1	6
actual yes	FN=7	TP=42	49
	12	43	

Confusion Matrix

Accuracy:
(TP+TN)/total = (15+27+3)/55= 85.4545

ROC Curve:
FPR=FP/ (TN+FP) =1/6=0.16
TPR=TP/ (TP+FN) =42/(49)=0.85



Misclassification Rate:

- (FP+FN)/total = (1+7)/55 = 14.5
- equivalent to 1 minus Accuracy
- also known as "Error Rate"

Precision:

- TP/predicted yes = 42/47 = 76.66

Prevalence:

- actual yes/total = 49/55 = 89.09

Comparison:

- Our system covers the diseases which are not in the previous papers and also with better accuracy i.e. 85.5%

- We are alerts to the farmers via email and text messages, This functionality was not present in the earlier papers

predicts banana diseases like mosaic, black Sigatoka, yellow Sigatoka, Panama wilt, streak etc. The system provides the user with an analysis report that consists of the symptoms and remedies of predicted disease which will be passed on to the farmers for further actions

7. CONCLUSION

The system works for the identification, detection and providing remedial action(rectification) of the Disease's in the Banana leaves with accuracy of 85%. This system This system predicts banana diseases like mosaic, black Sigatoka, yellow Sigatoka, Panama wilt, streak etc. The system provides the user with an analysis report that consists of the symptoms and remedies of predicted disease which will be passed on to the farmers for further actions.

8. REFERENCES

[1]Jihen Amara, Bassem Bouaziz, Alsayed Algergawy "A Deep Learning-based Approach for Banana Leaf Diseases Classification", Lecture notes in Informatics (LNI) 2017.

[2]Basavaraj Tigadi, Bhavana Sharma "Banana Plant Disease Detection and Grading Using Image Processing", IJESC,Vol.6, Issue No.6, 2016

[3] Karthik .G, Praburam.N "Detection and Prevention of Banana Leaf Diseases From Banana Plant Using Embedded LinuxBoard", International Conference on Green Engineering and Technologies, 2016.

[4] Basavaraj Tigadi, Bhavana Sharma "A Survey on Methods of Plant Disease Detection", IJSR, Vol.4, Issue No.2, Feb 2015

[5] D. Surya Prabha & Dr. J. Satheesh Kumar "Study on Banana Leaf Disease Identification Using Image Processing Methods", IJRCSIT, Vol.2, Issue No.2A, March 2014.