PLANT DISEASE IDENTIFICATION USING DEEP LEARNING CLASSIFICATION MODEL: CNN

Sakshi Mangal[1] and Pratiksha Meshram[2]

[1]Department of Information Technology, Narsee Monjee Institute of Management Studies, Shirpur, Maharashtra, 425405, India

[2]Department of Information Technology, Narsee Monjee Institute of Management Studies, Shirpur, Maharashtra, 425405, India

Abstract: *Agriculture in India is not all about getting food for livelihood from agriculture. Large portion of Indian economy comes from agriculture as India is one of the largest exporters of food, grains and other agricultural products. More than 70% of rural population of India is dependent on agriculture of their livelihood. 42% of crop gets damaged every year which leads to huge loss to Indian farmers, out of which 15.7% of crop is damaged due to pest. Thus, early detection of plant disease is very important to protect complete plant from getting damaged. The older technique in which diseases are detected when changes are noticeable on plant leaf is not efficient as till that time almost whole plant gets damaged thus new technique like image processing and computer vision algorithms should be used to detect disease at early stage. The technique which is implemented should provide accurate and precise detection of disease, so that the pesticide and insecticide which are being used does not damage quality of soil and also spreading of large amount of pesticides and insecticide damages directly or indirectly the health of crop. Damage in crop quality or quantity leads to losses thus early and accurate detection of plant disease are very important.*

Thus, for early disease detection we used Laplacian filter and Unsharp masking technique as image processing technique and Canny edge detection as image segmentation technique. The classification model which we are implementing in this project is convolution neural network which is a deep learning classification model. [1][6][7][8][10][12][14]

Keywords: **CNN, deep learning, GLCM, canny edge detection and Gaussian blur.**

1. INTRODUCTION

India is a developing country with world's second largest population. The demand for food increases day by day with the increase in population and on the other end increasing population leads to bad climate which not only leads to global warming but also a bad environment for plant to grow and survive with new diseases and infection it gets. Traditional methods like naked eye detection is not only time consuming but also it needs huge man power to continuously monitor plants in farms. Thus, alternative methods were developed by scientist which could not only help farmer to detect disease at early stage but also need less time and be user friendly as most of the Indian farmers are not highly educated, so now image processing technique was used with machine learning algorithms but provide better result. Plant disease management is very important for rising Indian economy and to provide better lifestyle to hardworking farmers, many of whom had committed suicide as they were in huge dept. Thus, quality and quantity assurance of crop are necessary as economic growth of farmers and agriculture-based industries depends on it. [1][3][4][6][7][9][11].

In this paper we will implement a model for Pepper bell, Potato and Tomato plant which are mostly cultivated in Punjab, Maharashtra, Karnataka, Gujarat, West Bengal and Andhra Pradesh.

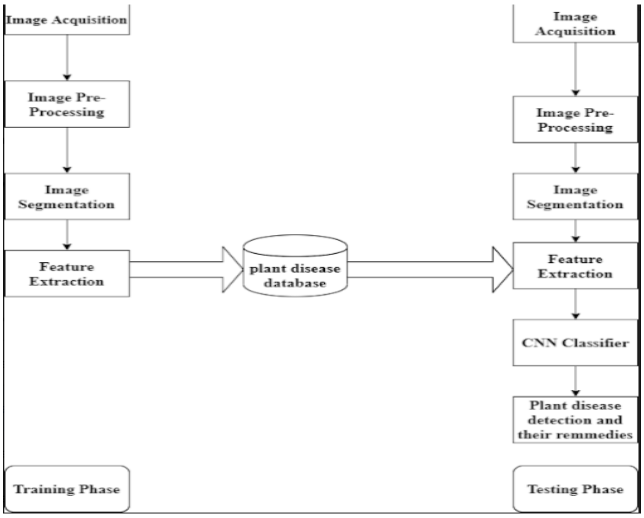

Figure-01: classification Process. [1]

This figure gives the brief of how this model is going to operate we in first step will collect all images using camera or from already data-set. Then in the second step we will remove extra noise from image using low pass filter and sharpen image. Then in third step we will segment the image using canny edge detection for knowing actual edges of leaf under detection. Later in fourth step we calculate the feature of leaf and if then feature of the test image is similar to any leaf in training database then it is estimated that the leaf under detection have same disease in step five using deep learning convolution neural network.

2. Types of Diseases

Early Blight: Its actually name is Alternaria solani which is a fungus. In this disease a 'bullseye' patterned spot is been observed on plant leaf. Initially it is 1-2 mm black or brown lesion on leaf which later turns into yellow halo when whole plant gets damaged. In tomato it is mostly seen on fruit, stem and foliage. In potato it is seen commonly on stem, foliage and tuber.

Late Blight: It is caused by oomycete pathogen which is a fungus. It infects fruits, stem, leaves and tubers of tomato and potato plant. The symptoms of this disease is that there are light to dark green color circular to irregular shape water spots which are mostly present at leaf tips. This disease spreads very fast and damages whole farm.

Mosaic virus: this virus gets spread by insect and pest in whole farms. In this disease the leaf of tomato plant turns yellow, white and light green by showing speckled appearance.

Yellow leaf curl virus: it is a DNA virus from Geminiviridae family. This virus causes the most destructive damage to tomato plant. It is mostly found in tropical and sub-tropical region. In this disease the leaf gets curled upwards or cupping of leaf occurs also there is reduction in leaf size and production of fruit.

Bacterial Blight: This disease is caused by bacterial pathogen. This disease is found in pepper bell plant and it mainly occurs because of soil as the bacteria can survive in soil for long period without host plant also. In this disease the plant leaf turns extensively dark. This disease decreases plant growth up to 65-75 %. [8][9]

Leaf Mold: This disease is caused by Passalora fulva which is a fungus. This disease only attack tomato plant and especially its foliage. The symptoms found in this disease are pale green to yellowish spots on leaf and foliage with curled leaves at times.

Spider Mite: this disease is commonly found in tomato plant. This virus makes a web like structure on both side of leaf but mostly on underside of leaf veins. Feeding of plant food to virus causes yellow whitish and molted leaves.

3. Methodology

3.1. Image Acquisition

To detect disease the plant is having either one can see the plant by naked eyes and by using previous experience detect the disease but is a time taking process thus the second alternative is been introduced which is by implementing deep learning and machine learning algorithms in agriculture for which we first need a image for both training the model and testing the model. For which we can either use previously created data-set for training model or create our own data set by taking photos using camera. Mostly disease's symptoms of plants are been found on leaf, stem and foliage so this should be taken into consideration while taking images. Here for implementing this paper we used already created dataset from plant village which have images of pepper bell, potato, tomato having the above-described diseases. [15]

Figure-02: image of pepper bell plant having bacterial spot from plant village dataset.

3.2. Image Pre-Processing

The images which are been collected in acquisition step are being pre- processed in this step as the original image has background noise and irrelevant information. Also, there is a need to sharpen the image for further processing.

Here in this project, we use Laplacian filter to sharpen the image using kernel. Kernel is a 3*3 matrix which is convoluted with image matrix to get desired sharpen output image. For which all we do is take the sum of second derivative of function with respect to x and second derivative of function with respect to y and sum them together to get Laplacian result.

Figure-03: result of Laplacian filter on pepper bell leaf having bacterial blight.

Similarly, we also applied unsharp mask technique on our original image to get an image without noise. In this technique we use Gaussian filter to remove extra noise as according to central limit theorem, if there is an unknown noise in your image then it must be Gaussian noise thus, we applied Gaussian filter. Now there is a duality in Gaussian filter size because small size filter gives sharp image with proper localization but with incorrect peaks and large size filter gives us proper peaks with smoother edge but with improper localization. So here according to our images we used a kernel sixe of (5, 5) with standard deviation as 1. In this whole process the low pass filter blurs the background of image and foreground is focused more because of which edges of leaves gets more visible which helps in segmentation.

In this we get an image by Gaussian filter and we subtract this metrics from image metrics now this resultant image is our Unsharp mask image to which we multiply weighted fraction to amplify image, now we added this weighted image with original image to get desired output. [1][4][8][9][10][15]

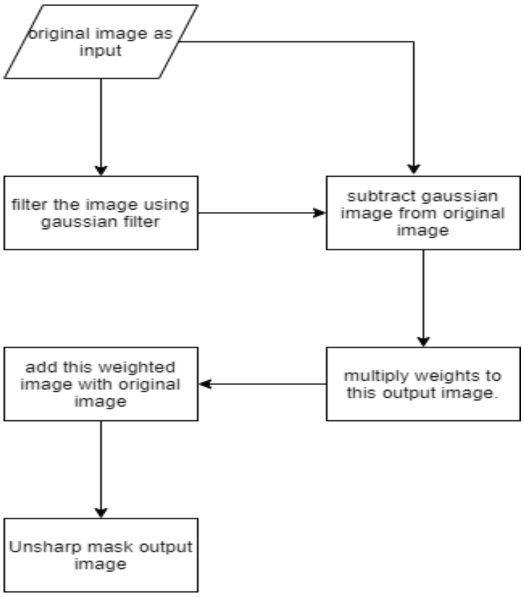

Figure-04: procedure for image pre-processing using unsharp mask technique.

Figure-05: result of unsharp mask technique on pepper bell leaf having bacterial blight.

3.3. Image Segmentation

An image has many pixels grouped together to form a complete colored picture. So, in the process of image segmentation we implement a technique through which we could segment those pixels into different segments on the basis of similarity of those neighboring pixels.

In this paper we implement canny edge detection technique which classify and segment the complete image on the basis of edges. Canny edge detection technique mainly involves five major steps:

Filter noise: filtering of noise is needed to blur the background so that the main infected region can be focused more. For which we apply Gaussian blur filter, which is a low pass filter and removes all unnecessary noises. After this step the edges of leaf are more visible and image becomes easy to work on for further use.

Find gradient: In the second step we can use Sobel operator or prewitt operator to find gradient. Prewitt and sobel operator work mostly same, the only major difference is that sobel operator is asymmetric around the center while prewitt operator is symmetric around the center. In this paper we used prewitt operator to implement canny edge detection technique. To find gradient we convolute the image pixel matrix with prewitt operator kernel matrix for both Gx and Gy. In this step we will find gradient magnitude to know the exact value by taking the under root of Gx square with Gy square sum

i.e. $\sqrt{(Gx^2 + Gy^2)}$

And gradient angle to know the direction of edge pixel which helps us in third step of this procedure. For which we will take tan inverse of Gy/Gx.

i.e., Angle = *tan-1* (Gy/Gx)

Non maxima suppression: In this step of canny edge detection, we draw a tangent to a point of edge whose gradient direction we know due to previous step and then we find gradient magnitude at that point and other two points which are closer to edge direction on both of the side to know the exact edge point. The point which has most gradient is considered as edge point.

Double thresholding: In this step we find two thresholds for complete image and then the threshold of pixels is compared with those two thresholds. This is done for edge thinning to know actual width of edge. In this the pixels with magnitude above the maximum threshold are discarded and pixel with magnitude less than minimum threshold is discarded.

Edge tracking and Hysteric thresholding: In this step edges are been tracked as many relevant medium edges also get discarded in double thresholding process.

In this paper we used automatic thresholding technique called p-tile technique for double thresholding and Hysteric thresholding. [1][3][8][9][10][13][15][17][18]

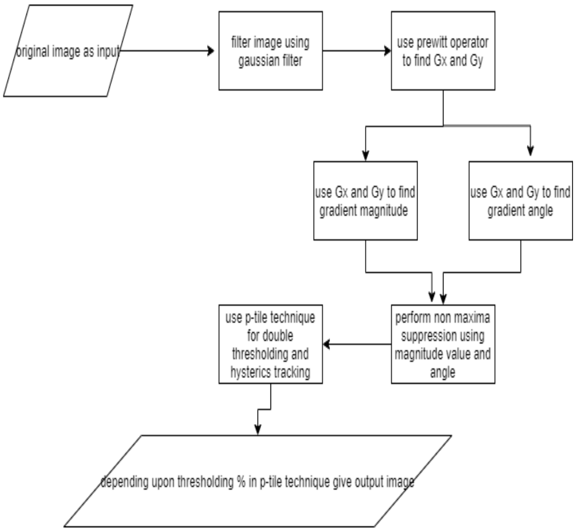

Figure-06: procedure for canny edge detection as image segmentation technique.

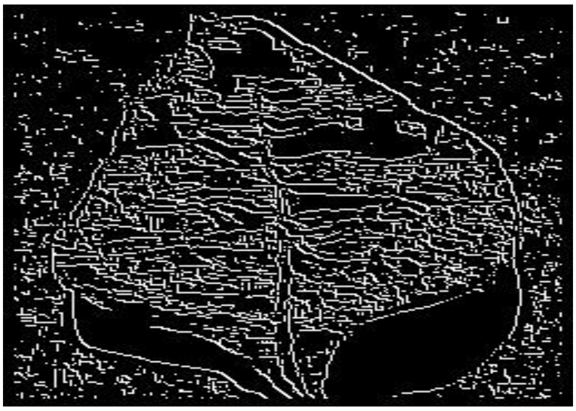

Figure-07: result of canny edge detection on pepper bell leaf having bacterial blight.

3.4. Feature Extraction

Every leaf has some features which make it easy to identify the kind of leaf it is. Plant disease also show some symptoms on plant leaf which is needed to be identified so that when in future there comes a leaf with same disease for testing in model could identify that which disease does that leaf have. [8][10][14]

Here in this paper, we use GLCM technique for extracting features on leaves. GLCM is grey level co-occurrence matrix which helps to find texture and spatial feature of images. GLCM have four types of transformation:

Point transformation: In this kind of transformation, we use 1*1 window to perform masking to increase or decrease contrast of output image.

$S= T (r)$

Where S is the output image for which T is the transfer function which is applied to r, the input image.

Linear transformation: It has negative and identity transformation. Negative transformation converts white pixel to black and black pixel to white.

$S = (L-1) - r$

Where S gives output image, L is the grey levels which is [0, 255] and r is input image.

Logarithmic transformation: In this a log function is plotted on graph and if the log output curve is higher than identity curve then output is wide which means output have high contrast and input is narrow with low contrast. If the output log curve is lower than identity curve then input image have high contrast and output image is having low contrast.

$S = c \log (1+ r)$

Where S is output image, which we get through logarithmic transfer function. C is the constant and here one is added with r because if r=1 is there then log 1 gives error. By this kind of transformation, we get enhanced and di-enhanced image according to spread and compression of grey level.

Power law transformation: This kind of transformation is called as gamma correction.

$S = C r^{\delta}$

Where increasing and decreasing gamma (δ) value, increase or decreases contrast. If gamma is more than one them contrast decreases and if gamma is less than one increases contrast.

Here in this paper, we calculate following texture feature using GLCM:

Mean: This feature helps to find the mean of feature as we don’t find feature for single pixel but we find feature for complete image. It let us know about coarse feature, about the size of image.

Standard deviation and variance: It help us to find texture of neighboring pixels by finding the square of difference of central pixel intensity and neighboring pixel intensity.

Contrast: This feature let us know the local variation is the image. If the value of p[i][j] increases away from main diagonal then image have high contrast.

Max: This feature let us know the maximum value in matrix.

Homogeneity: In this the small grey level will be present on the main diagonal of grey level co-occurrence matrix.

Entropy: It is measure of randomness of intensities in image. When all value in grey level co-occurrence matrix is similar then entropy is high and when values are unequal then entropy is low. [1][3][7] [15][16][18][19]

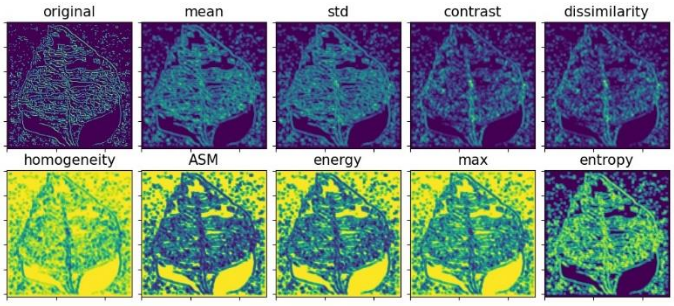

Figure-08: result of GLCM on pepper bell leaf having bacterial blight.

3.5. Classifier

Classification model is needed to train our data-set and test new infected leaf image accurately and precisely so that correct insecticides and pesticides could get spread on farm to cure disease and prevent further growth of any pathogen, fungus, bacteria etc. in farm.

In this paper we implemented convolution neural network to train our data-set. Convolution neural network is a deep learning model and its working have been shown in Fig-07 below:

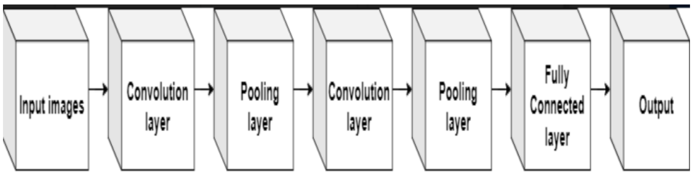

Figure-09: operation of convolution neural network layers

The original image goes as the input image in CNN model for which it firstly goes through convolution layer. In convolution layer all features are extracted by making a feature map in which a image matrix is convoluted with 3*3 filter matrix to get a feature map as a result. Convolution layer helps in doing edge detection and image sharpening depending on the kernel of filter used for convolution. Now many times it happens that the filter does not fits perfectly with image matrix for which there comes a concept of padding. In this we can either add zeros with image matrix which is called zero padding or we can remove or drop out the remaining part of image matrix which is called valid padding, in this only infected area is focused and irrelevant area is dropped out. Now we use ReLU to bring non- linearity in image ReLU is rectified linear unit. There are other functions also to bring non - linearity like sigmoid but ReLU performance is much better.

After getting through convolution layer resultant image goes through pooling layer. In pooling layer, the irrelevant parameters of image are been reduced. For spatial pooling, Max pooling is used. In max pooling, maximum value of rectified feature map is been taken and matrix size is reduced in resultant matrix which represent its input image.

Now the output image of Max pooling layer is been given to fully connected layer in which first the image matrix is been flatten into vector and goes through fully connected layer as neural network then model is created after which activation function like softmax and sigmoid are been called to perform logistic regression and classify the image as infected or healthy image. [4][5][10][11][12][15][19]

Figure-10: result of convolution neural network on the dataset.

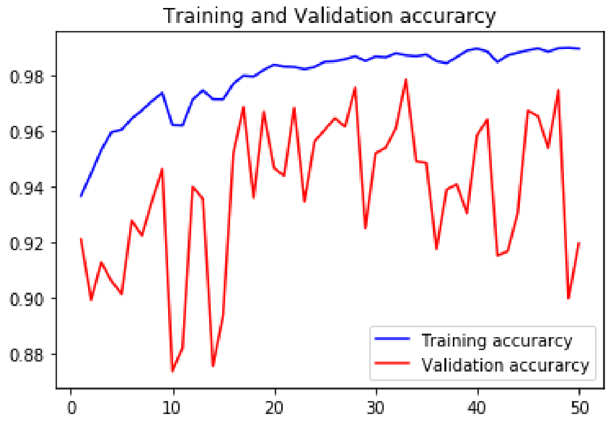

Figure-11: Plot showing training accuracy from 94-99% and validation accuracy from 91-95% for training model

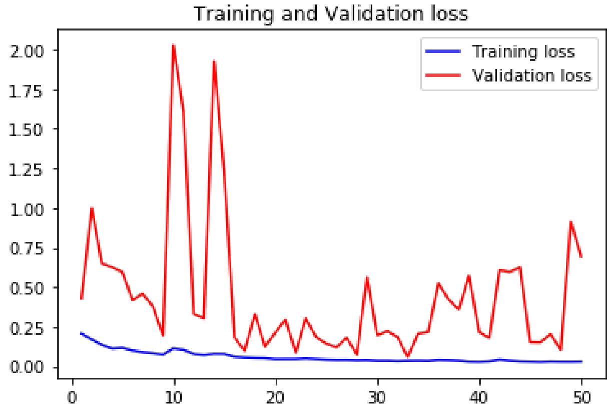

Figure-12: plot shows training loss is 0-20% and validation loss is 15-60% for training model.

For finding the model's performance we apply confusion matrix. In confusion matrix we find firstly true positive, in which we give correct data and model predicts it correctly. Secondly, we find true negative, in which we give correct data but model predicts it wrongly. Thirdly false positive, in which we provide wrong data but model predict it as a correct output. Fourthly we provide incorrect input and model also predict it as wrong then it is called as false negative.

Table-1: confusion matrix for checking performance of model for prediction.

	Prediction (Yes)	Prediction (No)
Actual (Yes)	Pepper bell bacterial blight disease leaf detected correctly. Potato late blight disease leaf detected correctly. Potato early blight disease leaf detected correctly. Tomato healthy plant detected correctly. Tomato yellow curl virus detected correctly. Tomato late blight detected correctly.	Pepper bell healthy leaf detected as pepper bell bacterial blight leaf. Potato healthy leaf wrongly detected as pepper bell bacterial blight. Tomato mosaic virus detected wrong as tomato healthy leaf.
Actual (No)	Lotus plant leaf provides error. Mango plant leaf provides error.	Banyan leaf detected wrongly as pepper bell healthy leaf. Watermelon bacterial blight detected as pepper bell bacterial blight.

3.6. System Architecture

For plant disease detection the plant leaf image goes through many phases as discussed above.

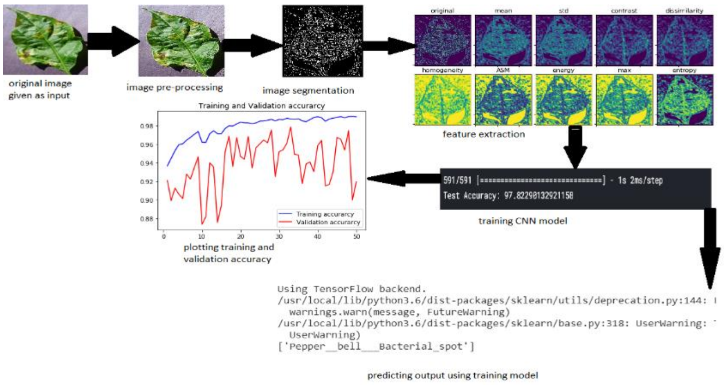

Figure-13: system architecture of the model implemented to detect plant disease.

The system which is implemented in this paper use the plant village dataset and implemented convolution neural network for training the model for dataset. Firstly, we provide original plant village dataset's image as input image which is then pre-processed to remove noise and smoothening of edges using unsharp mask technique. Later on, the image is segmented using canny edge detection then feature of image is been found out which helps in model training. Convolution neural network helps us to classify object, once the model gets trained, we find model accuracy for training and validation. The trained model is used for predicting the test image's disease.

4. Future Scope

There is always a scope for more improvement in technology. In Convolution neural network there are mainly three layers, which are convolution layer, max pooling layer and fully connected. Out of which convolution and max pooling layer are more complicated computationally but the fully connected layer is more memory consuming. In future we need a model which is less memory consuming and more complex computationally. As day by day everyone wants to get work done in less and less memory.

5. Conclusion

In this paper our main motive was to implement a new technology for plant disease detection. Agriculture is not only important because we are dependent on it for our basic needs like food but also good productivity of grains and cash crops is also very important as 17% of our GDP depends on agriculture. Thus, there is also a need there is also a need to keep plant healthy and increase crop productivity which is possible by spreading good quality fertilizer and accurate insecticide and pesticide all over the farm to prevent disease which has been occurred and to protect crop from further damage which will maximize profit of Indian farmer.

We implemented convolution neural network with 97.82% accuracy and average 10% validation loss. For image processing we applied unsharp mask technique. For segmenting image further for extracting feature, we used canny edge detection technique for segmentation and GLCM technique for extracting features. [7][12][15][20]

Table-2: performance of model for prediction.

Dataset taken for Implementation	Plant-Village Dataset
No. of images taken for training	15*200 = 3000 images
Accuracy of model	97.82%
Validation loss	10%
Time taken for prediction	12 seconds

REFERENCES

[1] *J. N. Reddy, K. Vinod and A. S. Remya Ajai, "Analysis of Classification Algorithms for Plant Leaf Disease Detection", IEEE, ICECCT*, Coimbatore, ***(2019)***, *pp. 1-6*.

[2] *Subhadra k, Dr. N. Kavitha, "Plant Leaf Diseases Detection Using Datamining – A survey", IJRAR – International Journal of Research and Analytical Reviews (IJRAR), (2018), pp. 223-237.*

[3] *S. Renuga, "An Identification of Variety of Leaf Diseases using Datamining Techniques", IJCRCST – International Journal of Contemporary Research in Computer Technology, (2017), pp. 100-102.*

[4] *M. Akila, P. Deepan, "Detection and Classification of Plant Leaf Diseases by using Deep Learning Algorithm", International Journal of Engineering Research & Technology (IJERT), (2018), pp. 1-3.*

[5] *G. Prem Rishi Kranth, M. Hema Lalitha, Laharika Basava, Anjali Mathur, "Plant Disease Prediction using Machine Learning Algorithms", International Journal of Computer Applications, (2018), pp. 1-7.*

[6] *Kshyanaprava Panda Panigrahi, Himanshu Das, Abhaya Kumar Sahoo, Suresh Chandra Moharana, "Maize Leaf Disease Detection and Classification using Machine Learning Algorithms", (2020), pp. 659-669.*

[7] *Kishori Patil, Santosh Chobe, "Leaf Disease Detection using Deep Learning Algorithm", (2020), pp. 3172-3175.*

[8] *Sharath D M, Akhilesh, S Arun Kumar, Rohan M G and Prathap C, "Image based Plant Disease Detection in Pomegranate Plant for Bacterial Blight", International Conference on Communication and Signal Processing (ICCSP), (2019), pp. 645-649.*

[9] *Abirami Devaraj, Karunya Rathan, Sarvepalli Jaahnavi and K Indira, "Identification of Plant Disease using Image Processing Technique", International Conference on Communication and Signal Processing (ICCSP), Chennai, (2019), pp. 749-753.*

[10] *Santhosh Kumar.S, B.K.Raghavendra, "Diseases Detection of Various Plant Leaf Using Image Processing Techniques: A Review", (2019), pp. 313-316.*

[11] *L. Sherly Puspha Annabel, T. Annapoorani, P. Deepalaksmi, "Machine Learning for Plant Leaf Disease Detection and Classification – A Review", International Conference on Communication and signal processing (ICCSP), Chennai, (2019), pp. 538-542.*

[12] *Suma V, R Amog Shetty, Rishab F Tated, Sunku Rohan, Triveni S Pujar, "CNN based Leaf Disease Identification and Remedy Recommendation System",* 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), *Coimbatore, India, (2019), pp. 395-399.*

[13] *M Sankar, D N Mudgal, Todkar, varsharani jagdish, Nandi wale Geetanjali Laxman, Mane Mahesh Jalinder, "Green Leaf Disease Detection Using Raspberry pi",* 1st International Conference on Innovations in Information and Communication Technology (ICIICT), *CHENNAI, India, (2019), pp.1-6.*

[14] *Amrita S.Tulshan, Nataasha Raul, "Plant Leaf Disease Detection using Machine Learning",* 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), *Kanpur, India, (2019), pp. 1-6.*

[15] *Sukhvir Kaur, Shreelekha Pandey, Shivani Goel, "Plant Disease Identification and Classification through Leaf Images: A Survey", (2019), pp. 1-24.*

[16] *Nirmalakumari K., Rajaguru H., Rajkumar P., "Plant Leaf Disease Detection Using Modified Segmentation Process and Classification", Computational Vision and Bio-Inspired Computing. ICCVBIC 2019. Advances in Intelligent Systems and Computing, vol 1108. Springer, Cham, (2020), pp. 1-4.*

[17] *S. D. Khirade and A. B. Patil, "Plant Disease Detection Using Image Processing",*

International Conference on Computing Communication Control and Automation, (2015), pp. 768-771.

[18] *G. K. Sandhu and R. Kaur, "Plant Disease Detection Techniques: A Review",* International Conference on Automation, Computational and Technology Management (ICACTM), *London, United Kingdom, (2019), pp. 34-38.*

[19] *P. Panchal, V. C. Raman and S. Mantri, "Plant Diseases Detection and Classification using Machine Learning Models", (2019), pp. 1-6.*

[20] *R. Gandhi, S. Nimbalkar, N. Yelamanchili and S. Ponkshe, "Plant disease detection using CNNs and GANs as an augmentative approach", International Conference on Innovative Research and Development (ICIRD), (2018), pp. 1-5.*

[21] *Pillai S.K., Raghuwanshi M.M., Shrawankar U., "Deep Learning Neural Network for Identification of Bird Species" Peng SL., Dey N., Bundele M. (eds) Computing and Network Sustainability. Lecture Notes in Networks and Systems, vol 75. Springer, Singapore, (2019), pp. 1-5.*

[22] *Pillai S.K., Raghuwanshi M.M., Gaikwad M., "Hyperparameter Tuning and Optimization in Machine Learning for Species Identification System" Dutta M., Krishna C., Kumar R., Kalra M. (eds) Proceedings of International Conference on IoT Inclusive Life (ICIIL 2019), NITTTR Chandigarh, India, Lecture Notes in Networks and Systems, vol 116, Springer, Singapore, (2020), pp. 1-7.*

[23] *Sankaran, Sindhuja & Mishra, Ashish & Ehsani, Reza & Davis, Cristina, "A review of advanced techniques for detecting plant diseases", Computers and Electronics in Agriculture, (2010), pp: 1-13.*