Automated Leaf Disease Detection in Corn Species through Image Analysis

Lakshmi Praneetha.S.K[1], Anusha.K[2], Geetha Viharika.R[3], Divya Sree.M[4], Vidyullatha.P[5]
[1] CSE Student of 4th Year at KLEF, Green Fields, Vaddeswaram, India, lakshmipraneethahhs1@gmail.com
[2] CSE Student of 4th Year at KLEF, Green Fields, Vaddeswaram, India, anusha.kusam11@gmail.com
[3] CSE Student of 4th Year at KLEF, Green Fields, Vaddeswaram, India, geethaviharika@gmail.com
[4] CSE Student of 4th Year at KLEF, Green Fields, Vaddeswaram, India, medarametladivyasree@gmail.com
[5] Associate Professor, Dept of CSE (KLEF) Green Fields, Vaddeswaram, India, pvidyullatha@kluniversity.in

ABSTRACT

India is a land of agriculture where they constitute a major part of the country's economy. The primary sector accounts for 18% of GDP in 2018 and employs more than 50% of countrymen. As per 2014, Crop productivity contributed 31% of GDP. To increase economic efficiency, the crop must be healthy. Crop production problems are very common which severely affects the productivity resulted in the loss to the rural farmers and the agriculture sector. Leaf plays an important role as it gives information about the quantity and quality of agriculture yield in advance depending upon the condition of the leaf. Therefore, early detection and diagnosis of plant diseases are very important. Plant diseases often appear on the leaves, and the characteristics of the affected leaves can be varied and difficult to distinguish. Various techniques are used to identify the diseases among leaves but the developed automated system is an advanced system that helps to find the total infected area and hence calculates the percentage of infection and improving the learning process to distinguish the disease. The proposed system is validated with Kaggle - plant village dataset (corn species) containing 3852 images. The dataset includes both infected as well as healthy leaves. We implemented the image enhancement techniques such as CLAHE algorithm and clustering process to extract features and keras sequential model to classify whether the leaf is diseased or healthy. Our model is 91% accurate to produce the testing results.

Key Words: Clahe, Classify, Detection, Kaggle, Keras, Predict, Sequential.

1. INTRODUCTION

Detecting the leaf whether it's healthy or unhealthy is a major problem for the farmers. If we do not take proper care to the leaf at its starting stage of disease, then the leaf may cause severe issue to its respective products, people will leads to a severe health issues on having these products. To save the people's health and money, we need to find out the leaf whether it's healthy or unhealthy at the early stages and if a leaf is unhealthy, do necessary steps to make it healthy. Automatic leaf disease detection helps farmer to detect the leaf state whether it's healthy or unhealthy, the farmer will get an idea about the plant and these can be rectified and we will obtain good products from the leaf.

There are many approaches available for plant leaf disease detection. In [1] they used K-means clustering for the detection of disease in a leaf. First, in this approach it takes the image input and do necessary image pre-processing techniques and later using the image segmenting algorithms it check the intensity values of the image, if there is any sudden change in the intensity value then it consider as the edge of the image and by using the K-means clustering algorithm it classifies the image and predict the result. In this paper they proposed few image segmentation techniques such as amplitude thresholding, Component labelling, Boundary based approach, Region based approach, Texture segmentation. However, in K-means clustering it is very hard to set the K value.

In [2] they used feed forward back propagation neural network in this paper they used the Prewitt edge detection algorithm and Thining algorithm, First image edge detection is applied on the input image and the thining algorithm reduces to the threshold-based edge values and then by using the backpropagation neural network they predict the health state of the leaf.

On, seeing the above two approaches [3] is the better approach. In this approach, they used the neural networks to get the maximum accurate result. For making the result more accurate they added AlexNet to the CNN model which will make the CNN architecture larger. They are 19 layers presented in VGG-19 [4] while GoogLeNet [5] has 25 Layers in it. However CNN [6] has high architecture which can produce high accurate result by adding these the architecture becomes more complex and causes more problems.

In [7] they used the neuro-fuzzy inference system for the detection of disease in the leaf. For the segmentation of image they used graph cut method to extract the color layout features using of these values they trained the model using the adaptive fuzzy inference system and the neural network classifies the leaf disease based on the trained model.

In [13] they determine the disease type and measure crop damage, thus offering potential economic assistance to the farmers affected. It is possible to quickly identify and quantify

the infection utilizing image analysis methods on the various regions of the cotton plant. This study analyzes the methods used by various authors to predict cotton leaf disease.

In [14] they concentrated on early diagnosis image recognition and used sophisticated vector support machines to recognise and classify leaf diseases. On referring the above approaches we have gone through the keras [8] model and we understood that combination on CNN with keras will give best results, so we builded a keras model which classifies the leaf disease based on the trained model [12]. The figure shown below is the architecture of the CNN

Figure 1: shows the hidden layers of CNN and its inner layer

Building of the model with keras comes in two forms one is sequential and other one is functional api. Most of the models use the sequential form [11] because it easily allows stack sequential layers of the network. We added 2D convolution layer to process the Image, and we kept the kernel size as 3x3 which acts like a moving window and we used the rectified linear unit as activation function followed by the max pool function and we trained our model with the corn crop dataset available in Kaggle plant village dataset and we achieved the better accuracy result.

First, we will give input as an image and pre-processing algorithms are applied on the image for the feature extraction, here we used the clahe algorithm as the pre-processing algorithm because the clahe algorithm increase the contrast of the local region values where we can see the image much more clearly than the original image
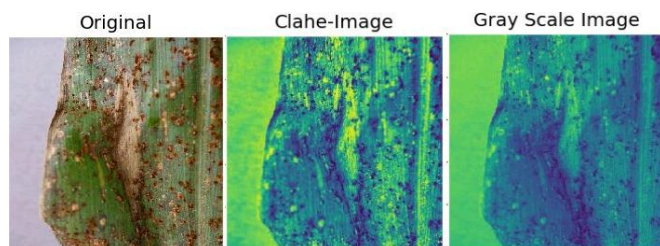

Figure 2: shows the Original, Clahe-image, Gray scale algorithm.

On seeing the above image, you can understand the reason for using the clahe algorithm. On comparing with the gray scale image we have better quality in the clahe image, so we used the clahe algorithm. After apply the clahe algorithm we used the canny-edge detection for finding the edges in the image and to make the image smoother, we used the Gaussian blur later on using the contour for drawing the outline of edges and imghsv package in CV for tracking of the coloured disease contained portion in leaf.

Not only on the corn crop dataset we trained our model using various leaf datasets like the rice crop dataset, tomato dataset and we achieve the better accuracy.

2. RELATED WORK

They are plenty of techniques for predicting the crop diseases here are the few approaches that we studied.

2.1 K-Means Clustering

Here, They[1] used K-means clustering for the detection of disease in a leaf. First, in this approach it takes the image input and do necessary image pre-processing techniques and later using the image segmenting algorithms it check the intensity values of the image, if there is any sudden change in the intensity value then it consider as the edge of the image and by using the K-means clustering algorithm it classifies the image and predict the result. In this paper they proposed few image segmentation techniques such as amplitude thresholding, Component labelling, Boundary based approach, Region based approach, Texture segmentation. However, in K-means clustering it is very hard to set the K value.

2.2 Feed Forward Back Propagation Neural Network

In this paper[2] they used the Prewitt edge detection algorithm and Thining algorithm, First image edge detection is applied on the input image and the thining algorithm reduces to the threshold-based edge values and then by using the backpropagation neural network [10] they predict the health state of the leaf.

2.3 Alex-Net and GoogleNet with CNN Architecture

It is the better approach[3]. In this approach, they used the neural networks to get the maximum accurate result. For making the result more accurate they added AlexNet to the CNN model which will make the CNN architecture larger. They are 19 layers presented in VGG-19 [4] while GoogLeNet [5] has 25 Layers in it. However CNN has high architecture which can produce high accurate result by adding these the architecture becomes more complex and causes more problems.

2.4 Neuro-Fuzzy Infernce System

In this paper they used the neuro-fuzzy inference system[6] for the detection of disease in the leaf. For the segmentation of image they used graph cut method to extract the color layout features using of these values they trained the model using the adaptive fuzzy inference system and the neural network classifies the leaf disease based on the trained model.

3. PROPOSED SOLUTION

In this project, building of the model with keras comes in two forms one is sequential and other one is function-al api. Most of the models use the sequential form because it easily allows stack sequential layers of the network. We added 2D

convolution layer to process the Image, and we kept the kernel size as 3x3 which acts like a moving window and we used the rectified linear unit as activation function followed by the max pool function and we trained our model with the corn crop dataset available in Kaggle plant village dataset and we achieved the better accuracy result.
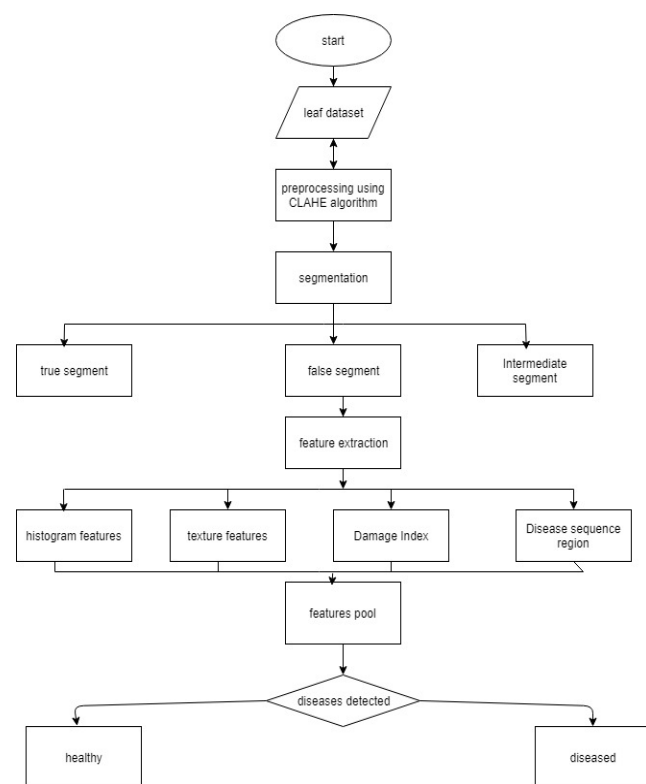
3.1 Work Flow Chart

Figure 3: shows the flow chart of proposed system

3.2 Collecting Data Set

In this section, the image processing and clustering techniques are applied to our model, and then an excision study of Kaggle Plant-Village dataset (Corn Species) is used to test our model to analyze the efficiency of the built model.

3.2.1 Kaggle –Plant Village Dataset (Corn species)

We have seen different datasets widely available in internet but those datasets are not suitable for training and testing our model. Only Corn species is chosen to attain better results among all datasets. The data set contains four kinds of classes named Cercospora leaf spot Gray leaf spot(513 images), Common rust(1,192 images), Northern Leaf Blight(985 images), and corn healthy species(1,162 images). Plantvillage dataset from kaggle has a total of 3852 images.

3.3 Clahe Algorithm

Histogram processing techniques are global processing techniques that pixels are changed based on the grey-level content of the whole image by a transformation function.

Equalization of Histogram is an example of this. A local enhancement algorithm acts within an image on local regions. The mapping applied in the input image on each pixel is determined by some property of that neighbouring pixel. These methods vary from one another depending on the chosen property and how it appears in the mapping. In such approaches, it is possible to adjust the width of the neighbourhood or window size. Most enhancement algorithms make use of the user to pick those enhancement input parameters.

3.3.1 Clahe: (Contrast Limited Adaptive Histogram Equalisation)

CLAHE is an intuitive method for improving contrast. It is dependent on AHE, where a pixel's spatial area measures the histogram. Therefore, the frequency of a pixel in the local intensity histogram is converted into value in the display range relative to the rank of the pixel intensity. There are two vital parameters in CLAHE, suggested by Zuierveld et al: block size (N) and clip limit (CL). Such criteria are used primarily to monitor image quality, but users have defined them heuristically. Originally, CLAHE was designed for medical imaging. CLAHE was also alleged to enhance the contrast in the modification of the underwater and aerial image.

3.4 Mean-Shift Algorithm

Mean shift is one of the clustering algorithms that iteratively assigns the data points to the clusters by moving points to the mode. The mode is seen as the highest data point density (in the area, in the Meanshift context). Like something else, it is also recognised as the algorithm that seeks mode. Mean shift algorithm has many image recognition and computer vision applications.
Based on a set of data points, each data point is assigned to the nearest cluster centroid by the algorithm recursively. The distance to the nearest centroid cluster is defined by in which most of the neighbouring points are located. And each iteration every data point will move closer to wherever the most points are, this will lead to the centre of the cluster. When Each point is allocated to a cluster, then the algorithm ends.

Algorithm

Original condition:
 In the first step, the red and blue data points fully overlap before the beginning of the Meanshift algorithm.
Iteration end 1:
 All red data points are getting closer to clusters. It looks like there are going to be four clusters.

Iteration end 2:
 Using only two iterations, the extreme right and bottom left clusters tend to have achieved convergence. As the two centroids are very similar, the middle and bottom right clusters tend to be merging.

Iteration end 3:
No shift in the centroids of the top right and bottom left. Another two centroids are drawn together as each cluster is affected by the data points. This is a Meanshift signature, there is no fixed number of clusters.

Iteration end 4:
Almost all of the clusters were supposed to converge.

Iteration end 5:
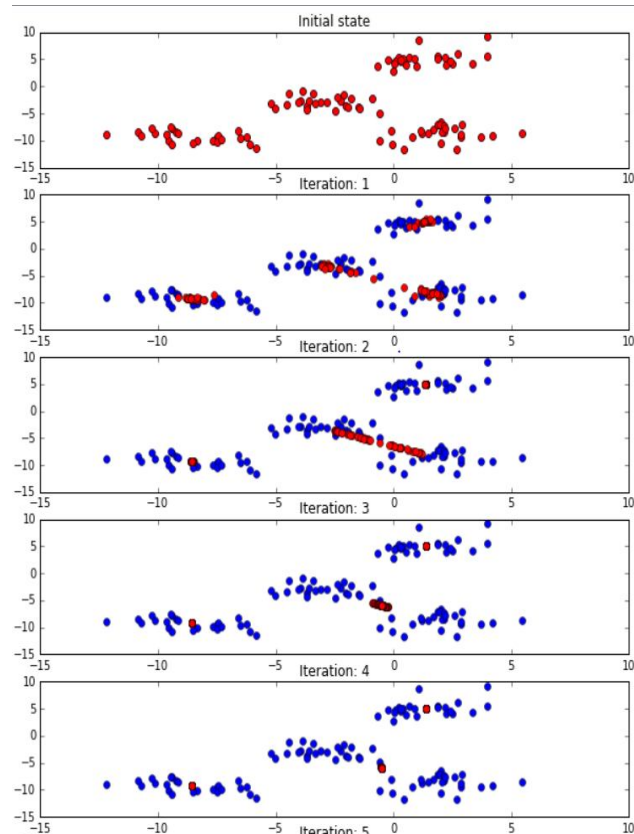There's no movement in all the clusters. The algorithm ends here as no shift for all red data points is observed.

Figure 4 :shows the iterations of Mean shift algorithm
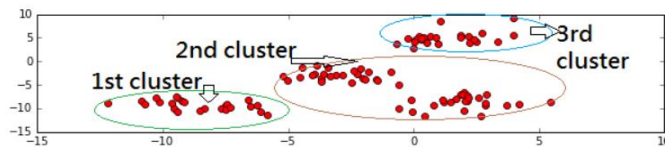
Figure 5: shows the clusters formation using mean shift algorithm

3.5 Gaussian Blur

In Gaussian Blur operation, instead of the box filter, the image is morphed with a Gaussian filter. The Gaussian filter that reduces the components of high frequency. Hence it is called as a low-pass filter.

Using the imgproc class Gaussianblur() method, you can perf orm this operation on an image.

3.6 Canny Edge Detector
The Canny edge detection technique is an operator for edge detection that makes use of a multi-stage algorithm to identify a variety of object edges. This was designed in 1986 by John F. Canny. Canny also developed an edge detection mathematical theory explaining why the technique is working.

Algorithm:
1. Reduction in noise;
2. Calculation of gradient;
3. Suppression of non-maximum;
4. Double threshold;
5. Hysteresis edge tracking.

1) Reduction of Noise
By applying Gaussian blur for smoothing purpose, is one way to get rid of the noise on the image. To do this, a Gaussian kernel (3x3, 5x5, 7x7 etc...) is used to convert images. The size of the kernel relies on the desired effect of blurring. In general, the smallest kernel is the blur, the less noticeable.

2) Calculation of gradient
The step of gradient calculation measures the intensity of edge and direction of the edge by using edge detection operators to measure the gradient of the image.
Edges refer to a change in the intensity of the pixels. The best way to detect it is to use filters that highlight this shift in frequency in horizontal(x) and vertical (y) directions.

3) Suppression of non-maximum
Ideally, there should be thin edges of the final image. Therefore, to thin the margins, we must conduct non-maximum suppression. The idea is simple: the algorithm runs through all of the points on the matrix of gradient strength and identifies the pixels in the edge directions with the maximum value.

4) Double Threshold
The double threshold stage intends to identify 3 types of pixels: strong ones, weak ones, and non-relevant ones: strong ones are pixels with such a higher intensity that we are certain they make a significant contribution to the final edge.
Weak ones are pixels with an intensity value which is not sufficient to be considered strong, but not small and want to be regarded as non-relevant to the detection of the edge.
Others, for the edge, are known to be irrelevant.

5) Hysteresis edge tracking
The hysteresis consists of converting weak pixels into strong ones based on the threshold values, if and only if at least one of the pixels around the one being processed is solid.

3.7 Theoretical Analysis

For the classification of the leaf health state we build a model using keras sequential form with 5 Conv2D layers. To normalize the input layer and hidden layer we used the batch normalization for each layer it adjust the mean and scaling of activation without effecting the neural network so we without

vanishing we can use the high learning rate and for reducing the dimensions of the image we used Max pooling technique for the each layer and at the end we used soft max function to get the probalistic values.

Before send the image to our model, we applied few pre processing techniques on image. As earlier we mentioned that we used clahe algorithm because the output after applying the clahe algorithm is the contrast image where each pixel in the image can be easily known. After applying the clahe algorithm we use the canny edge detection algorithm to find the edges in the image, here edge means the ending parts of the leaf. The edges detected by the canny edge detector are sent to the contour , using the contour the lines are drawn on each edge. After the drawing of the edges the result of the image is as shown as below
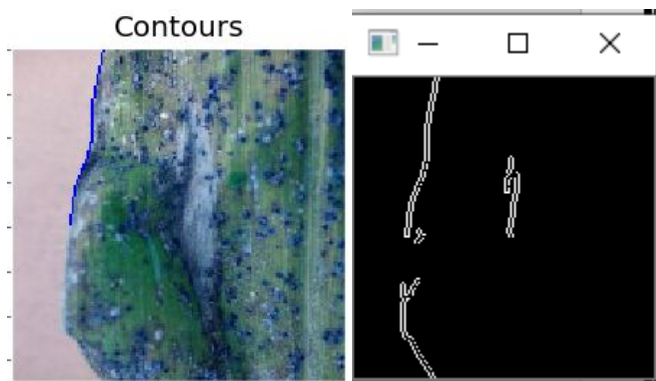
Figure 6: Contours and Edged Image

Later we used imghls in CV package to findout the spots on the leaf and from these results we estimated the disease percentage contained in the leaf.

The prediction of the image will result in some pattern. In this we observed that Healthy leaf contains pattern[0.0.0.1.] other than this pattern can be predicted that unhealthy leaf.

3.8 Trained Model

we perform a detailed study of various image enhancement algorithms and clustering techniques. It's a challenging task to make the system to identify the pattern of infection and to distinguish a leaf to be healthy or infected based upon the texture. The data set contains four kinds of classes named Cercospora leaf spot Gray leaf spot(513 images), Common rust(1,192 images), Northern Leaf Blight(985 images), and corn healthy species(1,162 images). Plantvillage dataset from kaggle has a total of 3852 images. We used Contrast limited adaptive histogram equalisation and grayscale techniques as pre-processing techniques which enhance the image pixels. Gaussian blur for smoothing purpose, Mean shift algorithm uses Euclidean distance for clustering and Contour algorithm detects the coloured regions obtained from edge detection through Canny. This procedure followed by Imghls and prediction through Keras sequential model.

Corn leaf is said to be healthy if the matrix resulted in the form of [0.0.0.1] and diseased if the matrix resulted in the form of [0.1.0.0]. Our trained model give better results while prediction and attained an accuracy percentage of 91%.
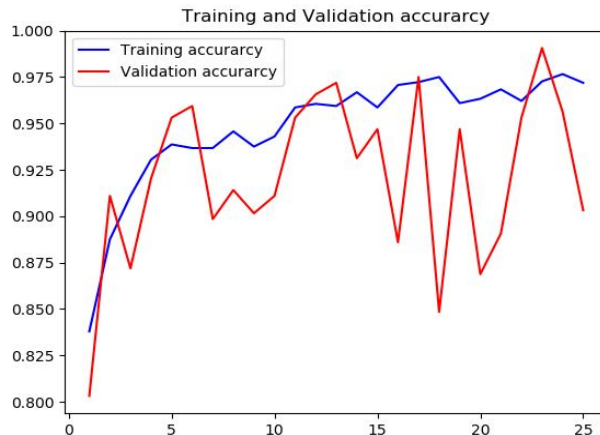

Figure 7: shows the comparison graph of training and validation accuracy.
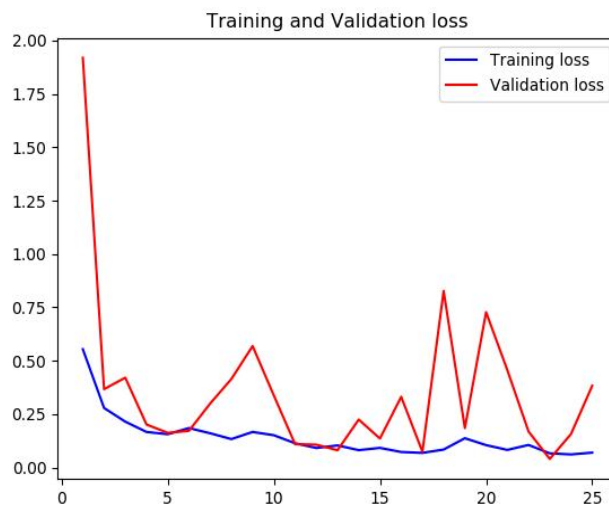

Figure 8: shows the graph of training and validation loss

4. RESULTS

In this section, we tested our model using kaggle leaf dataset. It detected the leaf healthy with high accuracy.
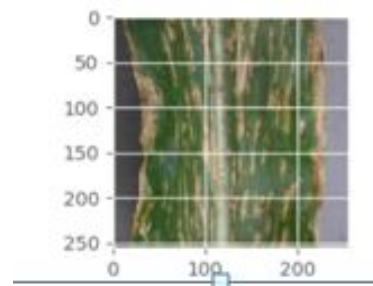
4.1 Diseased Leaf Prediction:

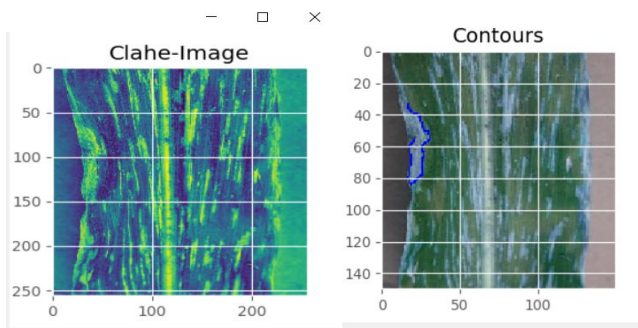**Figure 9 :**shows the original corn leaf.

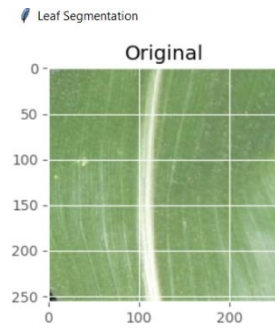**Figure 10:** shows the clahe-image and Infected Region of diseased leaf.

Figure 11: shows imghls, contour, masked images of diseased leaf

Figure 12: Predicted Leaf Result

The above figure shows the total area of leaf, and infected area and the percentage of infected region which helps to assess the leaf to be diseased or not.

The above all figures, resulted in predicting the leaf to be diseased depending on the spots and pattern of infection on the leaf.

4.2 Healthy Leaf Prediction

Figure 13: shows the original corn leaf.

Figure 14: Shows ingles, contour, masked images of healthy leaf
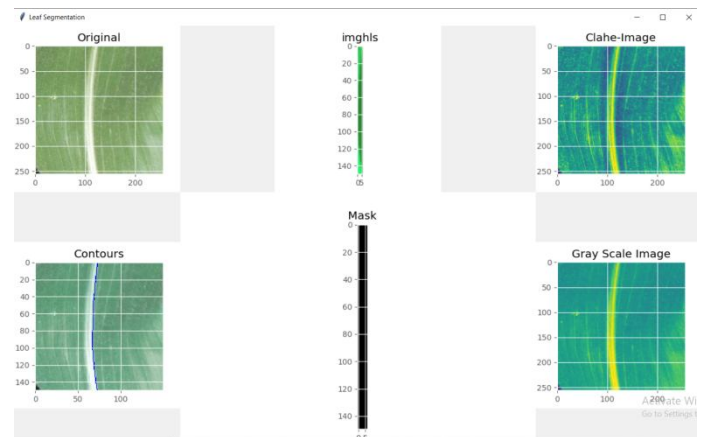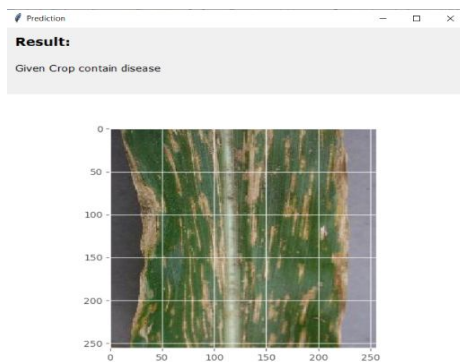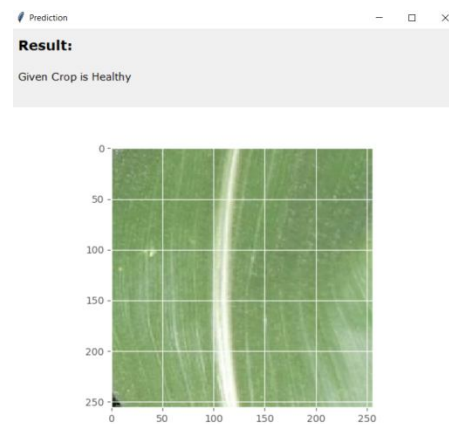

Figure 15: Predicted Leaf Result

5. CONCLUSION

The proposed system assesses the total infected area, the percentage of the infected region and enables the system to distinguish the disease. The system is validated with corn dataset obtained from Kaggle - plant village containing includes both diseased and healthy leaves of 3852 samples. We carried out the model with pre-processing methods such as CLAHE algorithm and clustering techniques and then followed by keras model to classify the leaf either to be diseased or healthy. Our designed model predicts with 91% of accuracy to classify the leaf.

6. ACKNOWLEDGMENT

REFERENCES

1. Jimita Baghel , Prashant Jain. K-Means Segmentation Method for Automatic Leaf Disease Detection. Int. Journal of Engineering Research and Application ISSN : 2248-9622, Vol. 6, Issue 3, (Part -5) March 2016, pp.83-86
2. M. S. Prasad Babu & B.Srinivasa Rao. Leaves Recognition Using Back Propagation Neural Network-Advice For Pest & Disease Control On Crops.
3. Yosuke Toda and Fumio Okura. How Convolutional Neural Networks Diagnose Plant Disease. Published in 2019 in Plant Phonemics (A Science Partner Journal).
   https://doi.org/10.1155/2019/9237136
4. Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. Published as a conference paper at ICLR 2015.
5. Mohammed Ismail. B, B. Eswara Reddy, T. Bhaskara Reddy "Cuckoo Inspired Fast Search Algorithm for Fractal Image Encoding" Elsevier Journal of King Saud University Computer and Information Sciences volume 30 issue 4, 2018, ISSN: 1319-1578 Pages 462–469 DOI 10.1016/j.jksuci.2016.11.00.
6. Christian Szegedy, Wei Liu , Yangqing Jia , Pierre Sermanet , Scott Reed , Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich. Published in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
7. P. R. Rothe, R. V. Kshirsagar. Adaptive neuro-fuzzy inference system for recognition of cotton leaf diseases. Published in: 2014 Innovative Applications of Computational Intelligence on Power, Energy and Controls with their impact on Humanity (CIPECH).
   https://doi.org/10.1109/CIPECH.2014.7019039
8. Recognition of Fake Currency Note using Convolutional Neural Networks , Navya Krishna , Rajarajeswari P et al, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-5 March, 2019.
9. Sakhare, N.N. Sagar Imambi, S. "Performance analysis of regression based machine learning techniques for prediction of stock market movement" International Journal of Recent Technology and Engineering Vol 7.No. 6, pp. 655-662
10. pellakuri Vidyullatha, Rajeswara Rao D, "Knowledge based information mining on unemployed graduates data using statistical approaches", International Journal of Pharmacy and Technology, Volume 8, Issue 4, December 2016, Pages 21961-21966.
11. Kiran Jammalamadaka, Nikhat Parveen, " Holistic Review of Software Testing and Challenges" , International Journal Of Innovative Technology and Exploring Engineering ISSN:2278-3075, Volume.8, Issue 7, Page No pp.323-326, June 2019.
12. Minxing Si, Tyler J. Tarnoczi, Brett M. Wiens , Ke Du. Development of Predictive Emissions Monitoring System Using Open Source Machine Learning Library - Keras: A Case Study on a Cogeneration Unit. Published in: IEEE Access (Volume: 7)
13. Supriya S Patki, Ganesh Sable. A Review: Cotton Leaf Disease Detection Article in International Journal of Advanced Trends in Computer Science and Engineering June 2019 DOI: 10.9790/4200-0603017881
14. Rima Herlina S. Siburian, Rahmi Karolina, Phong Thanh Nguyen, E. Laxmi Lydia, K. Shankar International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8, Issue-6S August 2019 712. DOI:10.35940/ijeat.F1138.0886S19 Leaf Disease Classification using Advanced SVM Algorithm.