Sasya Vaidya

Software to Detect and Diagnose the disease of the Crop/Plant using AI, ML and Image Processing.

[1]Suvarna Aranjo, [2]Tejas Relakar, [3]Pranav Pitla

[1] Professor, [2]Student, [3]Student,

Department of Information Technology
Xavier Institute of Engineering, Mumbai, India
[2]tejasrelekar@gmail.com

Abstract*— *The development of such an intelligent system is justified by its economical relevance and by hard efforts necessary to perform a correct diagnosis.

Keywords—Machine Learning, Plant, Diseases, K-Nearest Neighbours, K-Means.

I. INTRODUCTION

Leaves are considered the initial stage of onset of diseases in plants. Diagnosing the different aspects of leaves allows identifying and differentiating symptoms that indicate the presence of disease-causing microbes. Spots on leaves are considered significant indicators of microbial growth. Identifying various kinds of spots on leaves is essential to determine the type of plant disease. This identification can be carried out through recognition of varying spots on images of leaves. Image processing is utilized for the recognition of disease causing spots on leaf surfaces. Image processing paired with Machine Learning algorithms is a crucial step in the identification and recognition process.

II. LITERATURE REVIEW

Selecting a Template (Heading 2)

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the A4 paper size. If you are using US letter-sized paper, please close this file and download the Microsoft Word, Letter file.

Maintaining the Integrity of the Specifications

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

III. CONCEPT

IV. K-Nearest Neighbours algorithm is used for the identification of spots on leaf images. Initially images are captured using a high-resolution camera and passed through an edge detector. Once the leaf edges in each image are detected, the detected edges are further passed to mask creator for creation of binary image mask. This mask is then applied on the original leaf images and background is removed for removal of unwanted noise which might hamper further the processing.

V. The image is cropped digitally and the starting and ending position of the leaf edge is calculated. This position value indicates the horizontal and vertical positioning of the leaf within entire captured image. This resized image is then used to produce a histogram. The image comprises pixels ranging from 0 to 255 with 0 valued pixels indicating dark regions and 255 valued pixels indicating light regions. All values lie within this range and create the histogram.

VI. The histograms of leaf images are all fed to a Disease Area Counter to determine and differentiate the healthy and unhealthy parts of the leaf. Each histogram goes through feature extraction where parameters such as Correlation, Dissimilarity, Homogeneity etc are calculated. These parameters are used to determine the accuracy of leaf disease-prediction using KNN.

VII. The collected features are arranged and the KNN algorithm is applied to it to predict the output. Based on the generated output chemicals are suggested and displayed to the user for treating the diseased plants.

VIII. Fig.1. below depicts the block diagram of the entire process.

Fig.1. Block Diagram

A. Algorithm

The user captures images from live video and these images are input to the algorithm for processing. Initially background removal occurs to ensure that unwanted noise in images is eliminated. Edge detection occurs on the image to create a mask according to the outer edges. This mask is then applied to the original image.

NumPy library in python is used to crop the image and remove areas which aren't required in the processing. Further the image is again resized to 256x256 pixels for creation of histogram. Histogram is utilized for the analysis of image to identify and recognize spots and predict the output.

The green colored leaf pixels are masked using color detection to show only the areas with spots. The detected spots are counted and image features are extracted using GCLM method. The collected data is arranged and provided to the K-Nearest Neighbours algorithm for prediction of disease causing spots in the leaf image. Certain chemical treatment suggestions are also provided to the user on identification of disease causing spots and an overall report of data is displayed to the user.

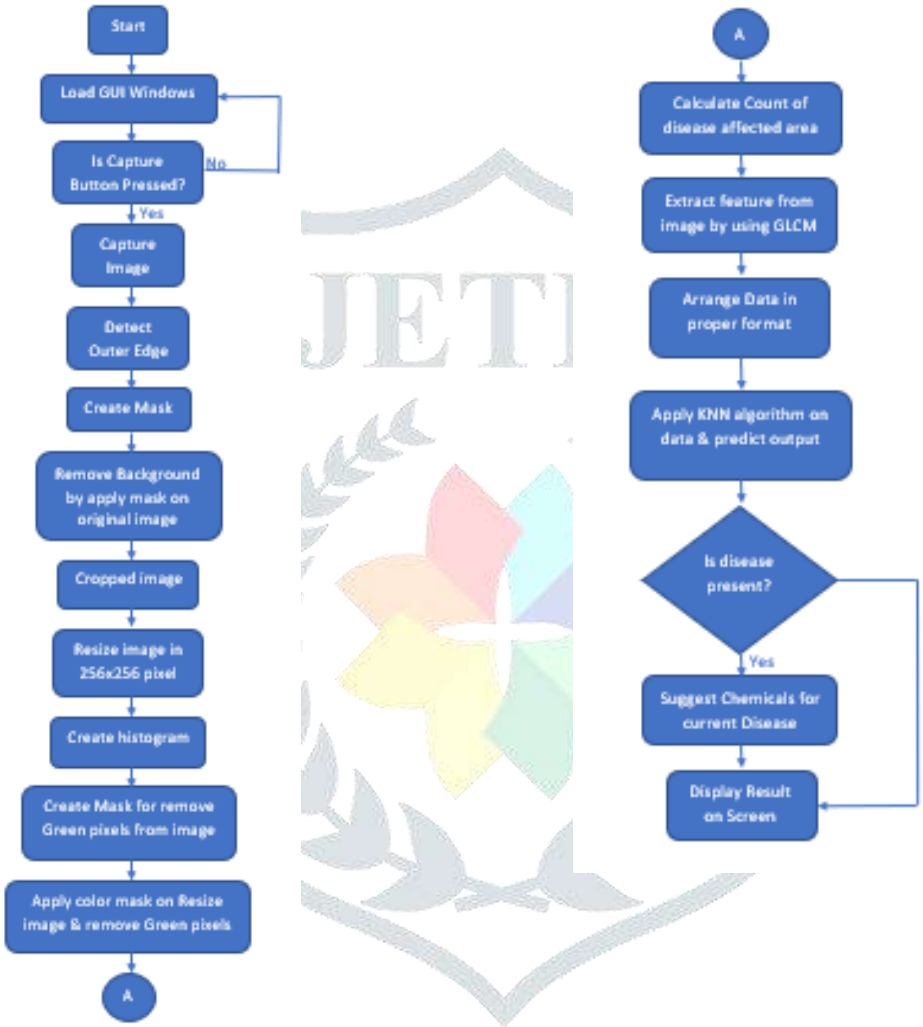

Fig.2. Flow Chart

B. Equations

KNN runs this formula to compute the distance between each data point and the test data. It then finds the probability of these points being similar to the test data and classifies it based on which points share the highest probabilities.

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

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

IV.RESULTS

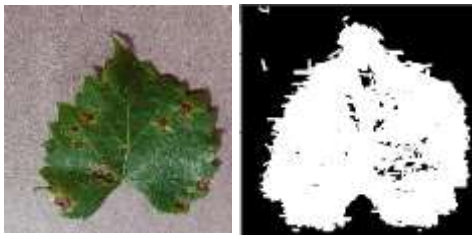

Fig. 3. (a) Read RGB Image from Camera (b) Detect edges using Edge Detector

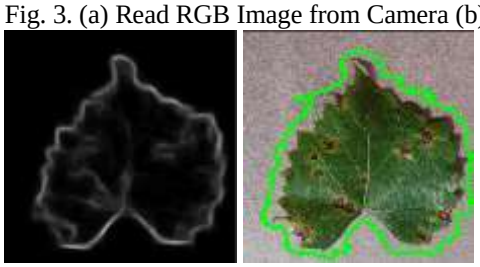

Fig. 4. (a) Addition of Salt and Pepper Noise (b) Creation of rough border

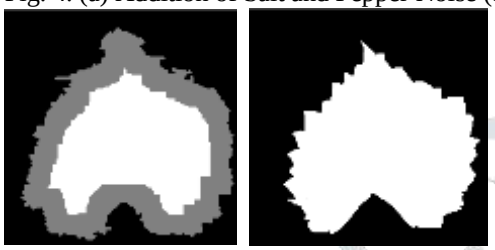

Fig.5. (a) Creation of a rough mask (b) Creation of proper mask from the rough mask as a reference

Fig. 6. (a) Apply masked image on original image (b) Crop masked image and remove unwanted noise

Fig.7. (a) Resize cropped image in 256x256 pixel image (b) Apply Histogram on resized image

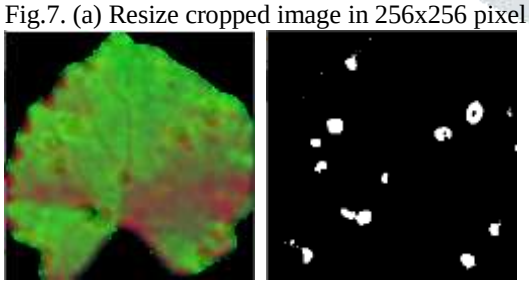

Fig. 8. (a) Convert blur image to HSV color space (b) Create mask for green pixels

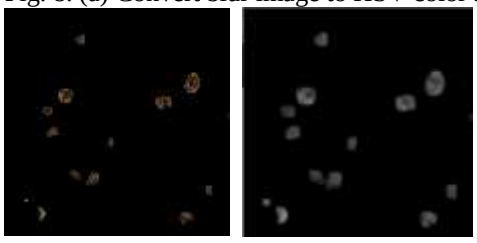

Fig.9. (a) Apply mask on original image for removal of green pixels from image (b) Apply Dilation on image to reduce pixel information in image

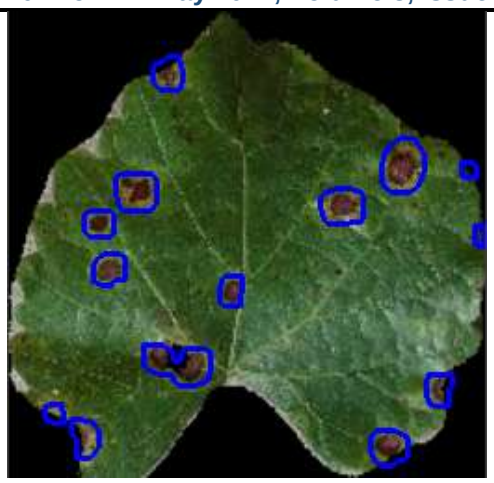

Fig. 10. Find all contours from Dilation image and draw contours on original image.

A. K-NN Accuracy

The K-Nearest Neighbors algorithm separates classes by creating a boundary between the different classes. The best choice of 'k' depends upon the available data. For our model, 1-N neighbour value provided 90% accuracy for test data set of leaf images.
We observed that the accuracy decreases with increase in neighbor count, indicating the accuracy being inversely proportional to neighbor count.
Fig.11 shows the accuracy graph for K-NN algorithm

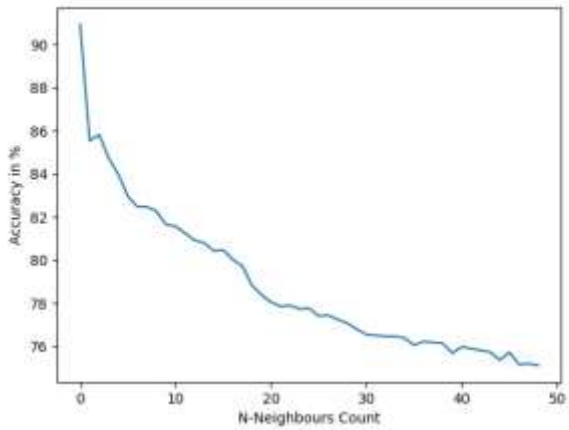

Fig. 11. K-NN Accuracy Graph

B. K-NN Error Rate

The K-Nearest Neighbors algorithm error-rate indicates the reduction in accuracy with respect to errors. For our model, 1-N neighbour value the error rate is less than 0.7 approximating zero.
We observed that the accuracy decreases with increase in error-rate, indicating the accuracy is inversely proportional to error-rate as well.

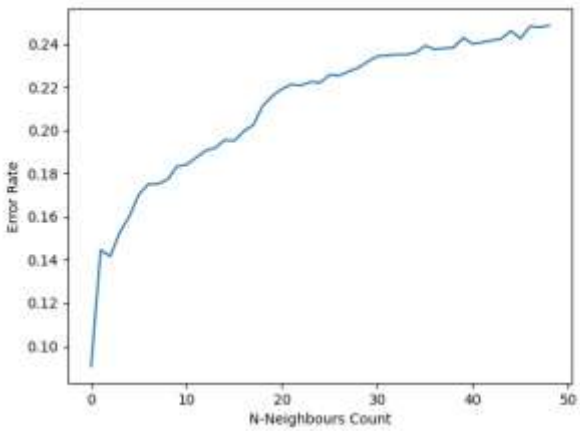

Fig.12. K-NN Graph

C. Confusion Matrix

Confusion matrix shows the True Positive versus False Positive values.

We utilized six classes and thus a 6x6 matrix is created, if classes are increased the size of matrix will eventually increase. This matrix shows actual values in X axis and predicted values in Y axis. As shown in matrix for class 1, 257 True values predict and other False value predict by the algorithm. In this confusion matrix diagonal values shows high value which indicate that KNN algorithm is perfectly suitable for this kind of data.

Fig.13 shows the Confusion matrix.

```
Confusion Matrix:
[[257  18   5  15   6  12]
 [ 15 315   5  28   7  12]
 [  0   1 208   0   3   6]
 [  5  22   1 354   0   4]
 [  2   0   9   1 179   2]
 [ 13  18   5   6   1 914]]
```

D. Classification Report

```
Classification Report:
                    precision    recall  f1-score   support

     Bacterial_spot       0.88      0.82      0.85       313
          Black_rot       0.84      0.82      0.83       382
      Early_blight        0.89      0.95      0.92       216
Esca_(Black_Measles)      0.88      0.92      0.90       386
        Leaf_scorch       0.91      0.93      0.92       193
            healthy       0.96      0.96      0.96       957

           accuracy                           0.91      2449
          macro avg       0.89      0.90      0.90      2449
       weighted avg       0.91      0.91      0.91      2449
```

V. CONCLUSION

The plant disease detection required three main steps namely; Feature Extraction, Segmentation and Classification. In the existing technique feature extraction is used to extract the textural features. The k-mean clustering is applied to remove background of input image. The SVM classifier is replaced with the KNN classifier in the proposed work to classify data into multiple classes. The performance of KNN is tested in terms of accuracy with is increase up to 10% as compared to SVM algorithm. K-Means clustering removes some kind of data from the image which will affected on the model accuracy. Image feature extraction increase the accuracy by 5%. Image histogram data increase speed of operation. Image uneven distribution of dataset effect on model accuracy.

VI. FUTURE SCOPE

Improve accuracy by using different classification algorithms and other image extraction techniques. Increase number of classes by using online data set. Improve hardware of industrial usage. Increase number of features and number of inputs. Implement algorithm on web application.

VII. REFERENCES

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, “Title of paper if known,” unpublished.

[5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.