

# Leaf Disease Detection using Polynomial SVM and Euclidean Distance Metric

Ritika Chouksey

Electronics & Communication Engineering  
University Institute of Technology  
Bhopal, Madhya Pradesh, India  
ritii.2908@gmail.com

Preety D. Swami

Electronics & Communication Engineering  
University Institute of Technology  
Bhopal, Madhya Pradesh, India  
preetydswami@gmail.com

**Abstract—** Agriculture productivity is highly important in the world to survive. There are lots of involvements of artificial intelligence in agriculture to help the productivity. Automatic leaf disease detection is one of them. It is hard to diagnose the leaf disease by normal vision because it looks quite natural. If care is not handled properly then it directly affect the quality of the production. So, it is important to detect the disease at early stage through which production can be improved and proper care can be taken place. There are so many researches have been done in this field but there are certain flaws present in resulting the system. Proposed system is based on Polynomial SVM (Support Vector Machine) and Euclidean Distance Metric. Polynomial SVM is a classifier that can handle the non-linear data in a very effective manner. Euclidean Distance Metric calculates distance between two different clusters or points; through which decision can be made easily. Dataset has been taken from kaggle for four different categories; such as *Alternaria Alternata*, *Bacterial Blight*, *Cercospora Leaf Spot* and *Healthy Leaves*. System pertained 97.30 % of accuracy which is bit higher than previous one.

**Keywords—** Leaf Disease, Polynomial SVM, Euclidean Distance Metric, *Alternaria Alternata*, *Bacterial Blight*, *Cercospora Leaf Spot*.

## I. INTRODUCTION

The agricultural land mass is something beyond being a taking care of obtaining today. Indian economy is exceptionally reliant of agricultural efficiency. Hence in field of agriculture, discovery of disease in plants assumes a significant part. To distinguish a plant disease in exceptionally introductory stage, utilization of programmed disease recognition method is advantageous. The current strategy for plant disease detection is essentially unaided eye perception by specialists through which ID and identification of plant diseases is finished. For doing as such, an enormous group of specialists as well as nonstop checking of plant is required, which costs exceptionally high when we do with huge ranches. Simultaneously, in certain nations, ranchers

don't have legitimate offices or even thought that they can contact to specialists. Because of which counseling specialists even expense high as well as tedious as well. In such circumstances, the proposed strategy ends up being useful in checking enormous fields of harvests. Programmed location of the diseases simply by seeing the side effects on the plant leaves makes it simpler as well as less expensive [1]. Plant disease distinguishing proof by visual way is more relentless assignment and simultaneously, less exact and should be possible just in restricted regions. While assuming that programmed identification procedure is utilized, it will take less endeavors, less time and become more precise. In plants, a few general diseases seen are brown and yellow spots, early and late singe, and others are parasitic, viral and bacterial diseases. Picture handling is utilized for estimating impacted area of disease and to decide the distinction in the shade of the impacted region [2].

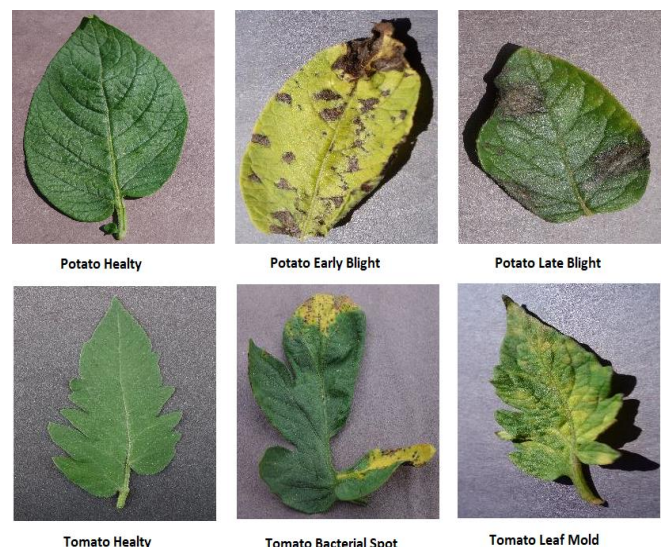


Fig. 1. Leaf Disease Example

Fig 1 shows the leaf disease image of potato and tomato leaf. Any imaging test does not confirm that any leaf pertains any disease but it gives lot of information through which any disease can be predicted with certain level of accuracy. Every disease has very less symptoms at the early stage and it is required to diagnose that disease in early stage to treat well. Automatic diagnosis is very helpful source for diagnosing leaf disease. So, many researches are based on machine learning technique like CNN, RNN, ResNet and many more. But all are very deep neural network that requires lots of samples to train that take more computational time for training and testing the datasets. A network model should be light in weight and uses small intelligent filters in hidden layers through which effective result can be pertained with high level of accuracy. There is an alternative approach i.e. classifier that can classifies the normal and abnormal cells and can take a decision accordingly. There are so many classifiers available through which classification can be performed for better result like Navie Bayes, K-means clustering, SVM and many more. But SVM is considered as the best classifier among them because it has better prediction level to classifies different pattern of cells [3].

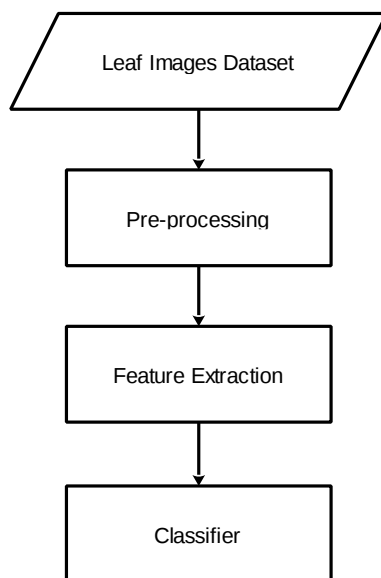


Fig. 2. Block Diagram of Operation Model

Fig 2 shows the basic process model for leaf disease detection where the leaf images dataset is to be downloaded first and pre-process the images. Later, the system extracted the features and classification can be done.

## II. RELATED WORKS

Fig. 3. There are many researchers tried to extract the lesion from leaf images and obtained good accuracy with certain false alarm rate. Jaskaran Singh et al. [4] proposed an research which is based on Region-Based Segmentation and KNN Classifier. In this paper, it is inferred that plant disease identification is the way to deal with identify diseases from the plants. In this work, the GLCM calculation is applied for the textural highlight examination, k-means clustering is applied for the region based division, and KNN classifier is applied for the disease expectation. The reproduction of the

proposed modular is done in MATLAB and results are displayed in type of figures and tables. Eftekhari Hossain et al. [5] proposed a research which is based on KNN classifier that detects the disease on the basis of color and textures. This work proposed a strategy which utilizes KNN to deal with recognize and characterize different diseases that are available in plant leaves. Diseases, for example, alternaria alternata, anthracnose, bacterial curse, leaf spot, and blister of plant leaves are considered for the trial. The division of the disease segment is finished by utilizing the k-nearest neighbor classifier and GLCM surface elements are utilized for the grouping. The KNN classifier based division result gives certain ideal precision in plant disease identification and the quantitative performance of the proposed calculation is acquired by estimating the DSC, MSE and SSIM boundaries. Aamir Yousuf et al. [6] proposed a research which is based on ensemble classifier. System uses ensemble classifier that combines two classifiers and obtains the result accordingly such as KNN and Random Forest. But SVM is much modern and better classifier as compare to other classifiers. SVM is able to deal with linear as well as non linear data with high prediction rate. SVM is consider as the best classifier to diagnose any disease related to the image processing. GLCM is also used to extract the textual feature of the image and later classifiers detect the disease accordingly.

Ch. Usha Kumari et al. [7] proposed a research which is based on K-means clustering and ANN. In this paper the leaf disease location is finished utilizing neural network classifier. The division is finished utilizing k-means clustering. Different elements like Contrast, Correlation, Energy, Homogeneity, Mean, Standard Deviation and Variance are separated for cotton and tomato diseases. The diseased leaves considered for reproduction are bacterial leaf spot, target spot septoria leaf spot and leaf form disease. Highlights are processed from disease impacted bunches 1 and 3. The highlights are taken care of to the classifier for perceiving and characterizing the diseases. Out of twenty cotton tests nine examples are arranged accurately as bacterial leaf spot and one example is misclassified as target spot. Eight examples are delegated target spot and two examples are misclassified as bacterial leaf spot. Abirami Devaraj et al. [8] proposed a research which is based on image processing based approach. It includes stacking an image, image preprocessing, image division, highlight extraction and order. Improvement of programmed detection framework utilizing cutting edge innovation like image process work with to help the ranchers inside the recognizable proof of diseases at an early or beginning stage and supply supportive information for its administration. They could like to expand their work extra on a ton of disease detection.

Table I Models Comparison

| Method | Finding                                                                                         |
|--------|-------------------------------------------------------------------------------------------------|
| VGG-16 | VGG16 has so many weight and biased parameters due to that the model becomes very heavy in size |

|                     |                                                                                                                        |
|---------------------|------------------------------------------------------------------------------------------------------------------------|
| DenseNet            | Excessive dense connections may complicated the network that affected computation efficiency                           |
| AlexNet             | AlexNet is not very deep model due to that it struggles to scan for all the features that resulting poor in performing |
| CNN                 | Conventional CNN model is poor in training and building heavy network that directly affects the execution time         |
| Ensemble Classifier | Ensemble classifier is bit harder to interpret and it costs high to train, evaluate and deploy the model.              |

Table I shows the drawbacks of certain models through which leaf disease can be diagnosed with satisfying precisions. It shows the comparison among various methods which have been used to implement the automatic leaf disease detection. Table I shows the problem which has been identified at each model of the system and where it suffers and what kind of perspective should be retain in mind to develop an ideal system that can diagnose leaf disease. System requires a better approach that can train the system with less filters or light weight filter with proper training through which percision can be calculated high with less false alarm rate.

### III. IMPLEMENTATION DETAILS

Proposed system is based on Polynomial Support Vector Machine and Euclidean Distance Metric. Proposed system is able to diagnose leaf disease automatically with classifying its category. In leaf image there are various noise present and to classify the lesion area; it is required to mask or erode the background information. Image pre-processing helps to enhance the image and classify the distinct region. SVM is a method through which similar kind of cells can form a group or cluster and classify them as per the patterns. Here system uses polynomial SVM which is a non linear SVM that can classifies non linear data. Leaf image is bit complicated in structure, so it is better to use non linear classifier to pertaining correct precision.

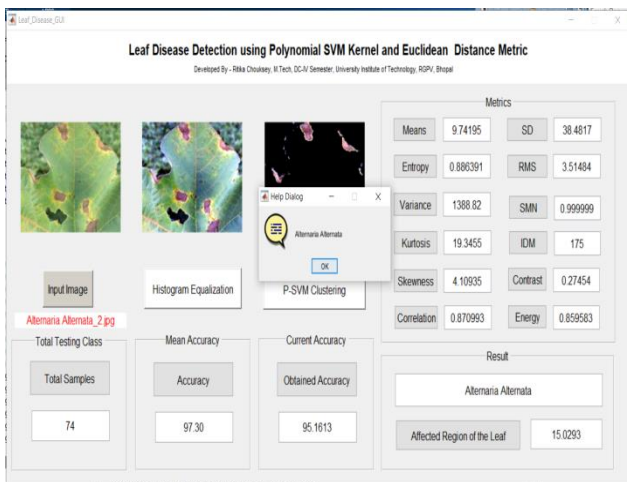


Fig. 4. Graphical User Interface

Fig 4 shows the graphical user interface of the proposed system where certain steps involved such as histogram equalization and clustering.

#### A. Histogram Equalization

There are various steps involved in proposed system but first step is to obtain the original image from dataset and apply histogram equalization. In histogram equalization; the contrast and brightness of an image get enhanced and the visibility of lesion get improved. It helps to classify the background that may contains noise that can be erode while edge detection.

$$P_n = \frac{\text{number of Pixel Intensity } n}{\text{Total number of pixels}} \quad n = 0, 1 \dots L - 1$$

Where  $P_n$  is the affected pixel value after histogram equalization. The histogram equalized image  $\gamma$  will be defined as

$$\gamma_{i,j} = \text{floor} \left( (L - 1) \sum_{n=0}^{f_{i,j}} P_n \right),$$

Where floor() can be considered as the nearest integer. It is equivalent to pixel intensity  $k$ ;

$$k = \text{floor} \left( (L - 1) \sum_{n=0}^{f_{i,j}} P_n \right)$$

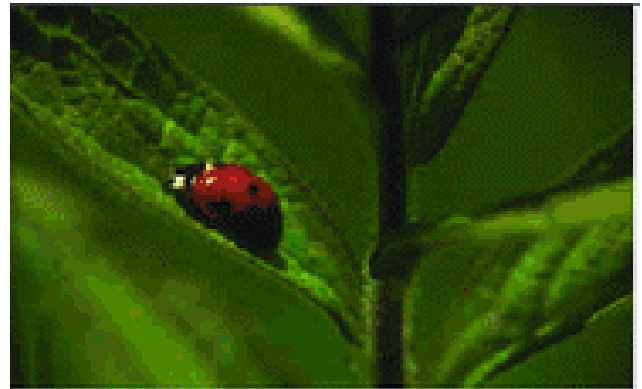


Fig. 5. Original Image

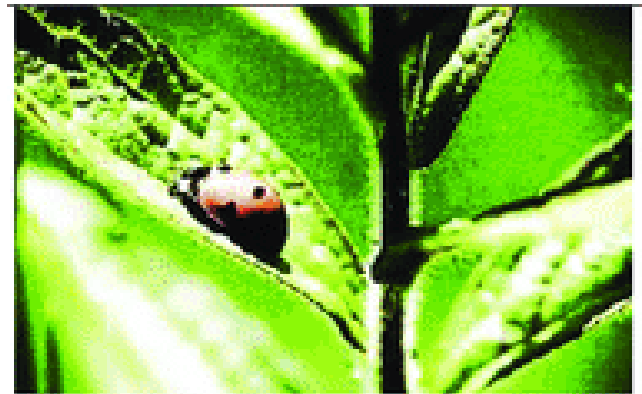


Fig. 6. Histogram Equalization affected Image

Fig 5 shows the original image of leaf and Fig 6 shows the histogram equalization affected image where it can be seen that; the visibility of the image has been enhanced that helps system to pertain good accuracy.

### B. Support Vector Machine

Support Vector Machine is method of classifying data on the basis of their patterns or appearance. SVM is considered as the most robust prediction technique that can classify data with more preciseness. Here system uses non linear SVM to deal with the non linear data. Most of the medical data belongs to the non-linear classes because of complex structure of blood vessels. Fig. 6 shows the separation of data with hyperplane.

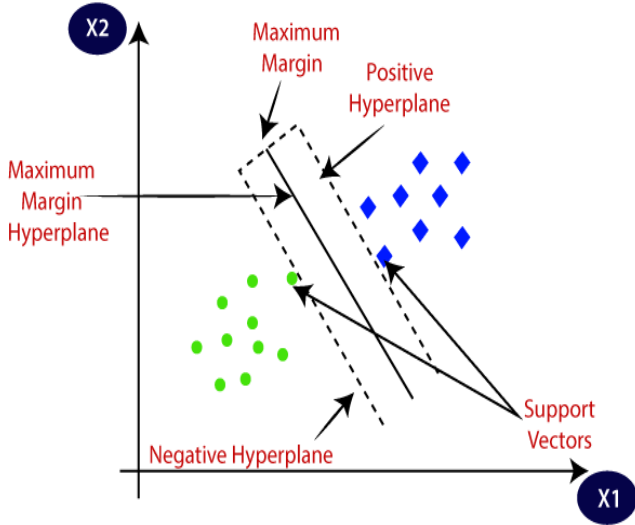


Fig. 7. SVM for Two Sample Classes [9]

Every hyperplane can be written as;

$$w * x + b = 1$$

$$w * x + b = 0$$

$$w * x + b = -1$$

Where  $w$  is the normal vector,  $b$  is the bias,  $x$  is the data points. If data point is on the hyperplane then it would be;  $w * x + b = 0$  otherwise it would be either negative or positive. It is required to know that which data points are closer or nearer to the hyperplane.

$$h(x_i) = \begin{cases} +1 & \text{if } w \cdot x + b \geq 0 \\ -1 & \text{if } w \cdot x + b < 0 \end{cases}$$

It is required to maintain the balance of the classification between maximization and loss. It can be stated as;

$$\min_w \|w\|^2 + \sum_{i=1}^n (1 - y_i \langle x_i, w \rangle)$$

### C. Euclidean Distance Metric

It then computes the Euclidean distance from the ratios of the input image and the segmented image. Afterwards, it calculates the Euclidean distance between the input leaf and the next segmented images and so on till all the distances

have been found. The Euclidean distance,  $D$ , is calculated using the formula below:

$$D = ((X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots + (X_N - Y_N)^2)^{0.5}$$

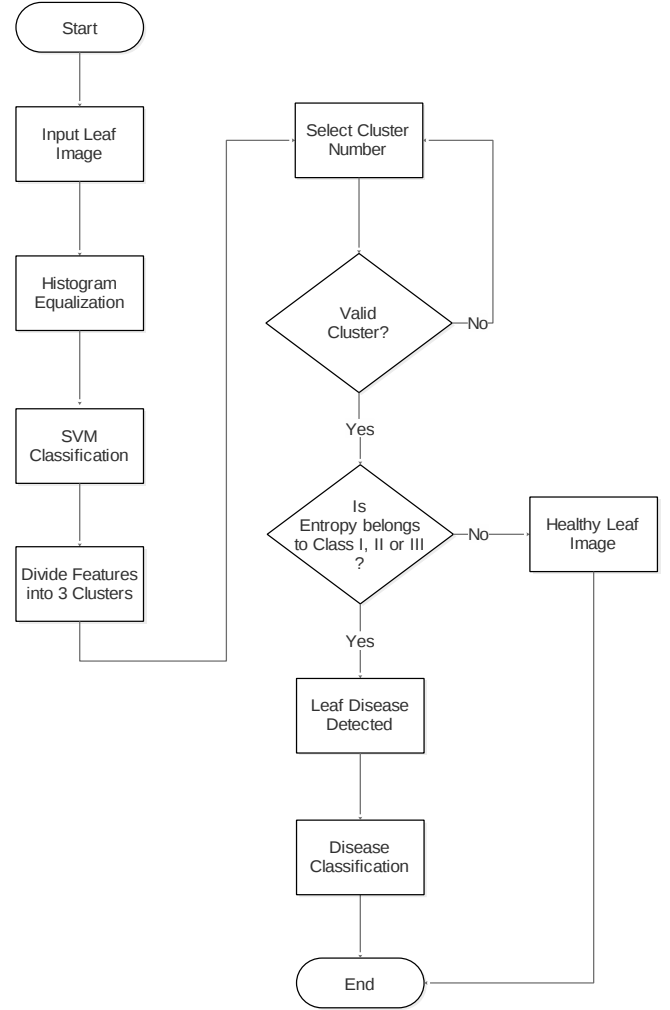


Fig. 8. Flowchart of Proposed System

Fig 8 shows the flowchart of proposed system where system firstly loaded the dataset image as an input data. Then pre-processing module has been initiated for enhancing the visibility of the images. Histogram is one of them, it is responsible to balance the brightness and contrast of the system, once the visibility increase then features gets extracted and after feature selection SVM classification can be initiated to classify the data points. User is required to select the cluster which has been created by the system. Then system calculates the entropy of the extracted lesion. It decides the density of the lesion that later compare with the threshold value. There are four switch cases in the system and comparison can be made accordingly. If entropy satisfies the case 1, 2, 3 then it will be considered as respective disease i.e. Alternaria Alternata, Bacterial Blight or Cercospora Leaf Spot. But if entropy satisfies case 4 then it would be consider as healthy leaf image. So, system also classifies the affected region as per the density of the lesion. Let it be more tangible with the proposed algorithm and the steps involved in the algorithm.

Table II Proposed Algorithm

| Polynomial SVM and Euclidean Algorithm                                                                                          |
|---------------------------------------------------------------------------------------------------------------------------------|
| Initialization                                                                                                                  |
| Input: Set of Image I=(i <sub>1</sub> , i <sub>2</sub> , i <sub>3</sub> ,..... i <sub>n</sub> )                                 |
| Output: Entropy                                                                                                                 |
| <b>Step 1:</b> Input image                                                                                                      |
| <b>Step 2:</b> Apply histogram equalization                                                                                     |
| $P_n = \frac{\text{number of Pixel Intensity } n}{\text{Total number of pixels}} \quad n = 0, 1 \dots L - 1$                    |
| Where $P_n$ is the affected pixel value after histogram equalization.                                                           |
| <b>Step 3:</b> Collect data points as vectors $w_i$ .                                                                           |
| $y = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 \dots \dots$                                                                       |
| $= w_0 + \sum_{i=1}^m w_i x_i$                                                                                                  |
| $w_i = w_0, w_1, w_2 \dots \dots \dots w_m$                                                                                     |
| Where $w_i$ is the vector, $b$ is the bias and $x$ is the variable                                                              |
| <b>Step 4:</b> Calculate the margin                                                                                             |
| $w * x + b = 1$                                                                                                                 |
| $w * x + b = 0$                                                                                                                 |
| $w * x + b = -1$                                                                                                                |
| $h(x_i) = \begin{cases} +1 & \text{if } w.x + b \geq 0 \\ -1 & \text{if } w.x + b < 0 \end{cases}$                              |
| <b>Step 5:</b> Compute loss function                                                                                            |
| $\min_w \ w\ ^2 + \sum_{i=1}^n (1 - y_i \langle x_i, w \rangle)$                                                                |
| <b>Step 6:</b> Calculate Entropy of the cluster                                                                                 |
| $E = - \sum_{i=0}^{n-1} p_i \log_b p_i$                                                                                         |
| Where $n$ is the number of gray-levels, $p$ is the probability of pixel having gray-levels $i$ and $b$ is the base of function. |
| <b>Step 7:</b> if $E$ is True for Class I, II & III then                                                                        |
| Leaf disease detected;                                                                                                          |
| Classify disease;                                                                                                               |
| else                                                                                                                            |
| Healthy Leaf Image Detected;                                                                                                    |
| end else                                                                                                                        |
| end if                                                                                                                          |
| <b>Step 8:</b> End                                                                                                              |

#### IV. EXPERIMENTAL RESULT

Experimental results are based on four metrics; that are True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). True positive means if an image belongs to either class 1, 2 or 3 and system diagnosed it positively, True Negative means if an image does not belong to either Class 1, 2 or 3 and system diagnosed it as healthy. False Positive means if an image class 4 and system diagnosed it as Class 1, 2 or 3, False Negative means if an image belongs to the Class 1, 2 or 3 but system diagnosed it as normal. There are total 74 testing images where 34 images belong to class 1 (Alternaria Alternata), 18 images from class 2 (Bacterial Blight), 7 images belong from class 3 (Cercospora Leaf Spot) and 15 images from normal class in Kaggle benchmark.

Table III Experimental Results

| Terms                         | Proposed |
|-------------------------------|----------|
| Total Testing Class           | 74       |
| True Positive                 | 57       |
| True Negative                 | 15       |
| False Positive                | 2        |
| False Negative                | 0        |
| Specificity in %              | 88.24    |
| Precision in %                | 96.61    |
| Accuracy in %                 | 97.30    |
| F1 Score in %                 | 98.28    |
| Sensitivity in %              | 100      |
| Negative Prediction Rate in % | 100      |
| False Positive Rate in %      | 11.16    |
| False Negative Rate in %      | 0        |
| Recall                        | 100      |

$$\text{Sensitivity} = \frac{TP}{TP + FN} * 100 \%$$

$$\text{Specificity} = \frac{TN}{FP + TN} * 100 \%$$

$$\text{Precision} = \frac{TP}{TP + FP} * 100 \%$$

$$\text{Negative Prediction Rate} = \frac{TN}{FN + TN} * 100 \%$$

$$\text{False Positive Rate} = \frac{FP}{FP + TN} * 100 \%$$

$$\text{False Negative Rate} = \frac{FN}{FN + TP} * 100 \%$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} * 100 \%$$

$$F1 = \frac{2TP}{2TP + FP + FN} * 100 \%$$

$$\text{Recall} = \frac{TP}{FN + TP} * 100 \%$$

Table IV Proposed Classification Result

| Terms              | P-SVM + EDM (Proposed) |
|--------------------|------------------------|
| Mean               | 31.77                  |
| Standard Deviation | 59.48                  |
| Entropy            | 2.73                   |
| Root Mean Square   | 7.24                   |
| Variance           | 99.99                  |

|                             |       |
|-----------------------------|-------|
| Smoothness                  | 1     |
| Kurtosis                    | 9.29  |
| Inverse Difference Movement | 99.99 |
| Skewness                    | 2.36  |
| Contrast                    | 0.97  |
| Correlation                 | 0.83  |
| Energy                      | 0.57  |

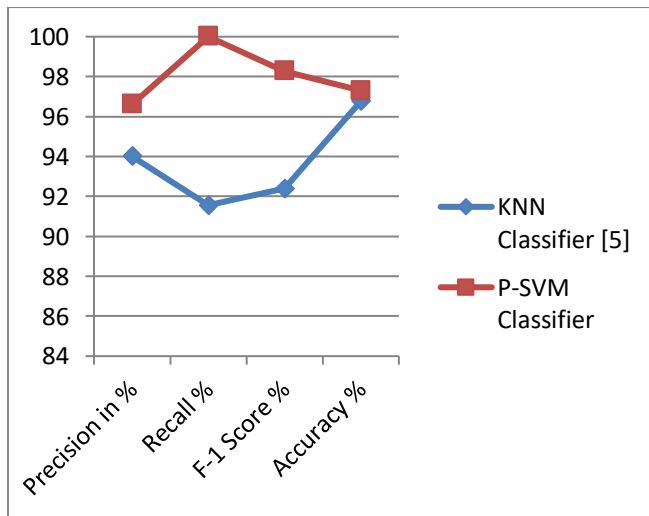
Table IV shows the proposed feature result based on four classes such as class1, class2, class3 and class4.

Table V Result Comparison

| Methods        | KNN Classifier [5] | P-SVM Classifier (Proposed) |
|----------------|--------------------|-----------------------------|
| Precision in % | 94.00              | <b>96.61</b>                |
| Recall in %    | 91.56              | <b>100</b>                  |
| F1-Score in %  | 92.39              | <b>98.28</b>                |
| Accuracy in %  | 96.76              | <b>97.30</b>                |

Table V represents the proposed result and compare with the previous system.

Graph I Result Comparison



## V. CONCLUSION & FUTURE SCOPE

Proposed system is based on Polynomial Support Vector Machine that classifies the abnormal and normal cells and on the basis of that decision can be made effectively with high accuracy. System is bit efficient in all parameters and achieved higher result as compared to the KNN classifier. Proposed classification technique is bit powerful to classify the normal and abnormal cells and take decision whether the image or disease belongs to respective category. The classification result is better than the previous model. System has been tested with kaggle benchmark and achieved better result. In future system can be tested with different benchmarks that contain several images or data. The accuracy can be enhanced by using certain modern pre-processing approaches.

## REFERENCES

- [1] Vijai Singh, A.K. Misra, Detection of plant leaf diseases using image segmentation and soft computing techniques, Information Processing in Agriculture, Volume 4, Issue 1, 2017, Pages 41-49.
- [2] Savita N. Ghaiwat, Parul Arora, Detection and classification of plant leaf diseases using image processing techniques: a review Int J Recent Adv Eng Technol, 2 (3) (2014), pp. 2347-2812.
- [3] Sanjay B. Dhaygude, Nitin P. Kumbhar, Agricultural plant leaf disease detection using image processing, Int J Adv Res Electr Electron Instrum Eng, 2 (1) (2013)
- [4] Singh, Jaskaran & Kaur, Harpreet. (2019). Plant Disease Detection Based on Region-Based Segmentation and KNN Classifier. 10.1007/978-3-030-00665-5\_154.
- [5] E. Hossain, M. F. Hossain and M. A. Rahaman, "A Color and Texture Based Approach for the Detection and Classification of Plant Leaf Disease Using KNN Classifier," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 2019, pp. 1-6, doi: 10.1109/ECACE.2019.8679247.
- [6] Yousuf, Aamir & Khan, Ufaq. (2021). Ensemble Classifier for Plant Disease Detection. International Journal of Computer Science and Mobile Computing. 10. 14-22. 10.47760/ijcsmc.2021.v10i01.003.
- [7] C. U. Kumari, S. Jeevan Prasad and G. Mounika, "Leaf Disease Detection: Feature Extraction with K-means clustering and Classification with ANN," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), 2019, pp. 1095-1098, doi: 10.1109/ICCMC.2019.8819750.
- [8] A. Devaraj, K. Rathan, S. Jaahnavi and K. Indira, "Identification of Plant Disease using Image Processing Technique," 2019 International Conference on Communication and Signal Processing (ICCSP), 2019, pp. 0749-0753, doi: 10.1109/ICCSP.2019.8698056.