iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 Issue: VII Month of publication: July 2022

DOI: https://doi.org/10.22214/ijraset.2022.45579

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Leaf Disease Detection using Polynomial SVM and Euclidean Distance Metric

Ritika Chouksey[1], *Preety D. Swami*[2]
[1, 2]Electronics &Communication Engineering Department, University Institute of Technology, Bhopal, Madhya Pradesh, India

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 taken properly then it directly affects 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 that have been done in this field but there are certain flaws present in the resulting system. Proposed system is based on Polynomial SVM (Support Vector Machine) and an 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. The proposed method provides 97.30% of accuracy which is bit higher than the KNN classifier.

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

I. INTRODUCTION

Agricultural is something that requires to be taken care of today on an emergent basis. 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 identification of plant diseases is done. For doing this, many specialists as well as nonstop checking of plants is required which increases with increase in the farm size.In many places, farmers don't have contacts with the specialists. Also the counseling specialists even charge high and the process is tedious as well. In such circumstances, the proposed strategy ends up being useful in checking enormous fields of harvests. Finding the nature and location of the diseases simply by seeing the pictures of the plant leaves makes it simpler as well as less expensive [1].Distinguishing between plant diseases by visual ways is a relentless assignment and simultaneously less exact and is possible just in restricted regions. While assuming that programmed identification procedure is utilized, it will take less time and will be 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 between the various diseases [2].

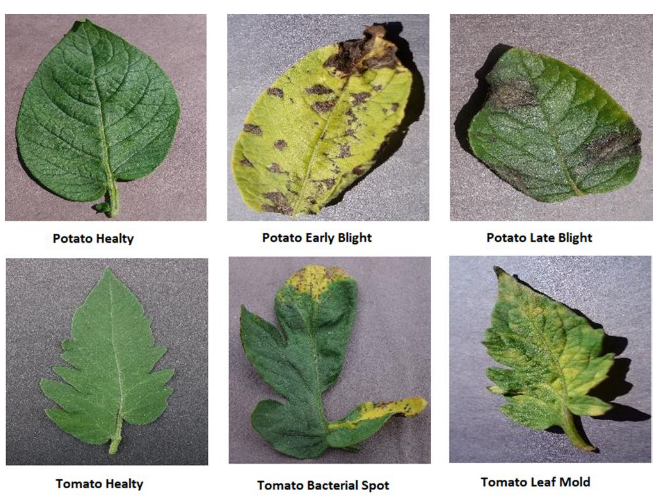


Fig. 1. Leaf Disease Example

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Fig. 1shows the leaf disease image of potato and tomato leaf.Any imaging test does not confirm that any leafpertains 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. Artificial Neural Networks (ANN) are still being used for classification provided the network is trained through features that have been properly selected [3].Automatic diagnosis is very helpful source for diagnosing leaf disease. So, many researches rely on machine learning technique like CNN, RNN, ResNet and many more. But all are very deep neural networks that require lots of samples to train and they take more computational time for training and testing the datasets. A network model should be light in weight and use small intelligent filters in hidden layers through which effective result can be found with high level of accuracy. There is an alternative approach i.e. classifier that can classify the normal and abnormal pictures and can take a decision accordingly. There are so many classifiers available through which classification can be performed for better results like Naive 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 classifydifferent patterns, textures and variations [4].

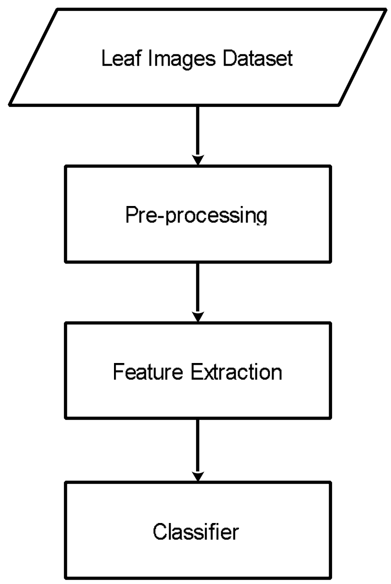


Fig. 2. Block Diagram of Operation Model

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

II. RELATED WORKS

Many researchers tried to extract the lesion from leaf images and obtained good accuracy with certain false alarm rate. Jaskaran Singh et al. [5] proposed a research which is based on Region-Based Segmentation and KNN Classifier.In their 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 classification. Eftekhar Hossain et al. [6] proposed a research which is based on KNN classifier that detects the disease on the basis of color and textures. 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 done by utilizing the k-nearest neighbor classifier and the 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. [7] proposed a research which is based on ensemble classifier. Their method uses ensemble classifier that combines two classifiers and obtains the result accordingly usingKNN and Random Forest. But SVM is much modern and better classifier as compared to other classifiers. SVM is able to deal with linear as well as non linear data with high prediction rate. SVM is considered as the best classifier to diagnose diseases or faults related to image processing [8]. GLCM is also used to extract the textual feature of the image and later classifiers detect the disease accordingly.
Ch. Usha Kumari et al. [9] proposed a research which is based on K-means clustering and ANN. In this paper the leaf disease location is foundby utilizing a neural network classifier. The distinction is done utilizing k-means clustering.

Different elements like Mean,Standard Deviation,Variance, Energy,Correlation,Contrast, and Homogeneityare computed 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 for the classifier to perceive and characterize 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 as target spot and two examples are misclassified as bacterial leaf spot.Image processing based approaches are quite popular in this area [10-13]. Abirami Devaraj et al. [14] also proposed an 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. Image denoising can also be treated as an important preprocessing step in image processing. Training a neural network or a classifier with features selected from denoised images produces far better results than training with the features of a noisy image. Many image denoising methods are available and from themthe method that is less computationally expensive can be chosen [15].

Table I Models' Comparison

Method	Finding
VGG-16	VGG16 has so many weight and bias ed parameters due to that the model b ecomes very heavy in size
DenseNet	Excessive dense connections may co mplicated the network that affected computation efficiency
AlexNet	AlexNet is not very deep model due t o that it struggles to scan for all the fe aturesthat resulting poor in performin g
CNN	Conventional CNN model is poor in t raining and building heavy network that directly affects the execution tim e
Ensemble Classifier	Ensemble classifier is bit harder to int erpret and it costs high to train, evalu ate and deploy the model.

Table I shows the drawbacks of certain models through which leaf disease can however be diagnosed with satisfying precision. A comparison between VGG-16, Densenet, AlexNet, CNN and Ensemble Classifier is presented that have been used in various researches for implemention of automatic leaf disease detection. This table tells where each model suffers and what kind of perspective should be retained in mind to develop an ideal system that can diagnose leaf diseases. A better approach is required that can train the system with less filters or light weight filters with proper training through which results with high precision can be obtained with less false alarm rate.

III. IMPLEMENTATION DETAILS

Proposed system is based on Polynomial Support Vector Machine and Euclidean Distance Metric. This method is able to diagnose leaf disease automatically with classifying its category. In leaf image, there is presence ofnoise 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 regions. SVM is a method through which similar kind of cells can form a group or cluster and classify them as per the patterns. The proposed method uses polynomial SVM which is a non linear SVM that can classify non linear data. Leaf image is bit complicated in structure, so it is better to use non linear classifier to obtain high precision in classification.

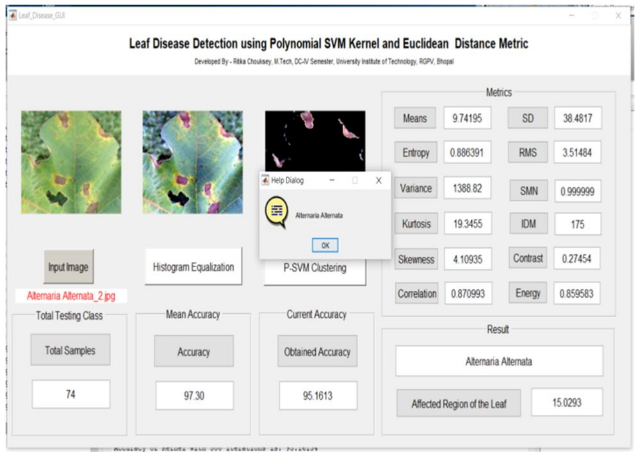


Fig. 3. Graphical User Interfacefor the Proposed Method

Fig.3 shows the graphical user interface of the proposed system where certain steps involved such as histogram equalization and clustering can be seen.

A. *Histogram Equalization*

There are various steps involved in the proposed method but the 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 is improved.

For an image f, with n pixels and where the variations in pixel intensities is from 0 to $L-1$ (L is the number of intensity values), the normalized histogram p_n of the image f for each possible intensity is given by

$$p_n = \frac{number\ of\ pixelshavingintensity\ n}{Total\ number\ of\ pixels}$$
$$n = 0,1 \ldots . L-1$$

The histogram equalized image γ will be defined as

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

It is equivalent to transforming the pixel intensities$k$ of the image f by the function

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

Fig. 4. Original Image

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Fig. 5. Histogram Equalization of the Affected Image

Fig.4 shows the original image of leaf and Fig.5 shows the histogram equalization of the affected image where it can be seen that the visibility of the image has been enhanced that will help the model to obtain good accuracy.

B. Support Vector Machine

Support Vector Machine is a tool for 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 precision. The proposed method uses nonlinear SVM to deal with the nonlinear data. Most of the images with leaf diseases belong to the non-linear classes because of their complex structure. Fig. 6 shows the separation of data with a hyperplane.

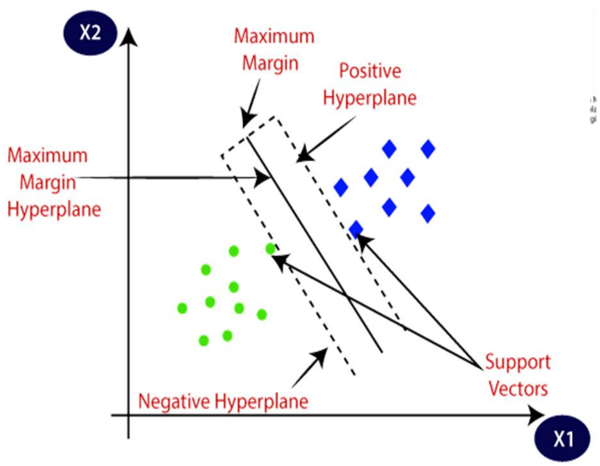


Fig. 6. SVM for Two Sample Classes

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 set of 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.

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

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

The Euclidean distance between the input image and the segmented image is to be calculated. Afterwards, the Euclidean distance between the input leaf and the next segmented image and so on is calculated till all the distances have been found. For the computation of the Euclidean distance, following formula can be used

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

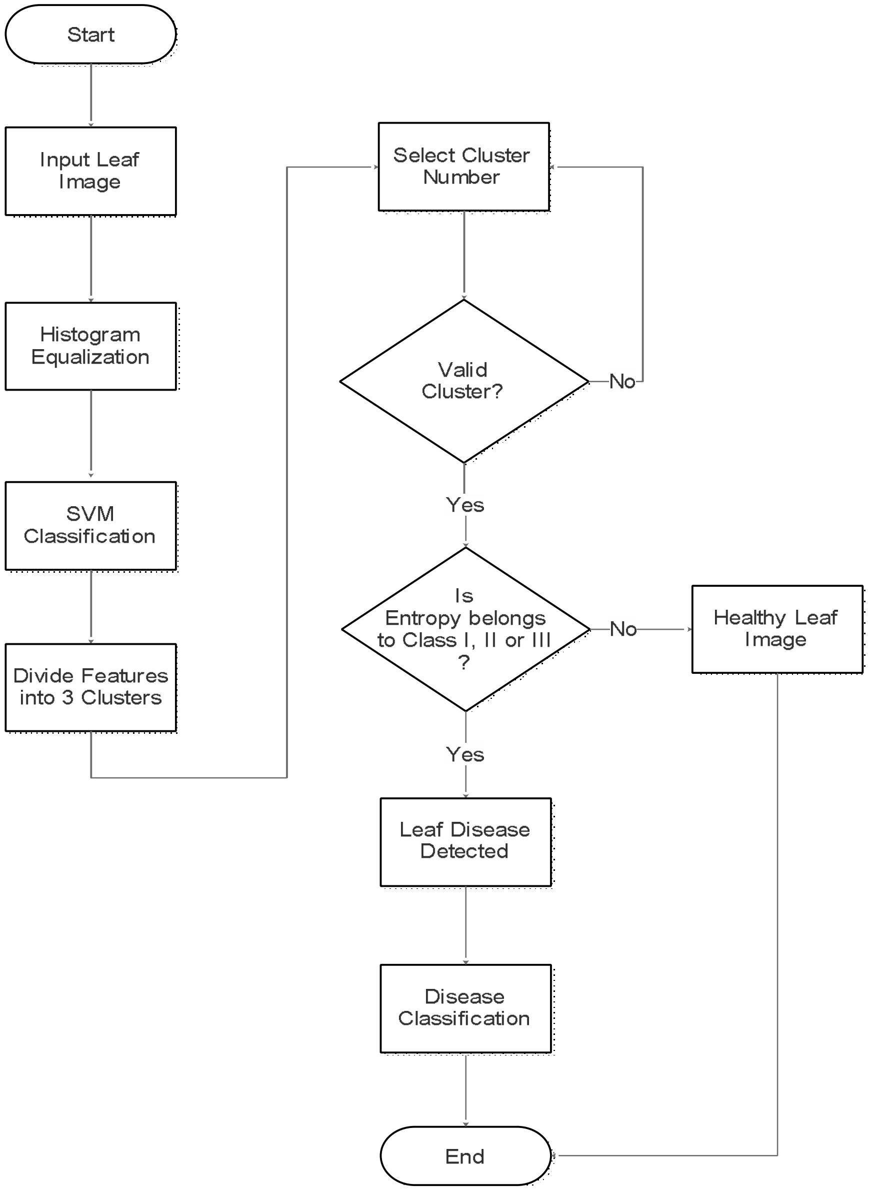


Fig. 7. Flowchart of the Proposed Method

Fig.7 shows the flowchart of proposed system where system firstly loads 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 is later compared 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. If entropy satisfies case 4, then it would be considered as healthy leaf image. So, system also classifies the affected region as per the density of the lesion.Table II tabulates the various steps involved in the proposed algorithm.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Table II Proposed Algorithm

Polynomial SVMand Euclidean Algorithm

Initialization

Input: Set of Image I=(i_1, i_2,i_3,………i_n)

Output: Entropy

Step 1: Input image

Step 2: Apply histogram equalization

$$p_n = \frac{number\ of\ pixels\ with\ intensity\ n}{Total\ number\ of\ pixels} \quad n = 0,1\ldots.L-1$$

Where p_n is the normalized histogram of the image.

Step 3: Collect data points as vectors w_i.

$$y = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 \ldots\ldots\ldots$$

$$= w_0 + \sum_{i=1}^{m} w_ix_i$$

$$w_i = w_0, w_1, w_2 \ldots\ldots\ldots\ldots. w_m$$

Where w_i is the vector, b is the bias and x is the variable

Step 4: Calculate the margin

$$w * x + b = 1$$

$$w * x + b = 0$$

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

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

Step 5: Compute loss function

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

Step 6: 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.

Step 7: if E is True for Class I, II & III then

Leaf disease detected;

Classify disease;

else

Healthy Leaf Image Detected;

end else

end if

Step 8: End

IV. EXPERIMENTAL RESULT

Experimental resultsare based on four metrices;that are True Positive (TP), False Positive (FP), True Negative (TN) and False Negative(FN).

Truepositive means if an image belongs to either class 1, 2 or 3and 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 3but system diagnosed it as normal.

There are total 74 testing imageswhere 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 Kagglebenchmark.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Table III Experimental Results

Terms	Proposed Method
Total Testing Classes	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	Proposed Method
Mean	31.77
Standard Deviation	59.48
Entropy	2.73
RMS	7.24
Variance	99.99
Smoothness	1
Kurtosis	9.29
Inverse Difference Movement (IDM)	99.99
Skewness	2.36
Contrast	0.97
Correlation	0.83
Energy	0.57

Table IV shows the average of the featuresfrom the proposed results based on fourclasses such as class1, class2, class3 and class4.

Table V Result Comparison

Method	KNN Classifier [6]	P-SVM Classifier (Proposed)
Precision in %	94.00	96.61
Recall in %	91.56	100
*F*1-Score in %	92.39	98.28
Accuracy in %	96.76	97.30

Table Vand Fig. 8 represent the proposed result and its comparison with the results of KNN classifier in [6].

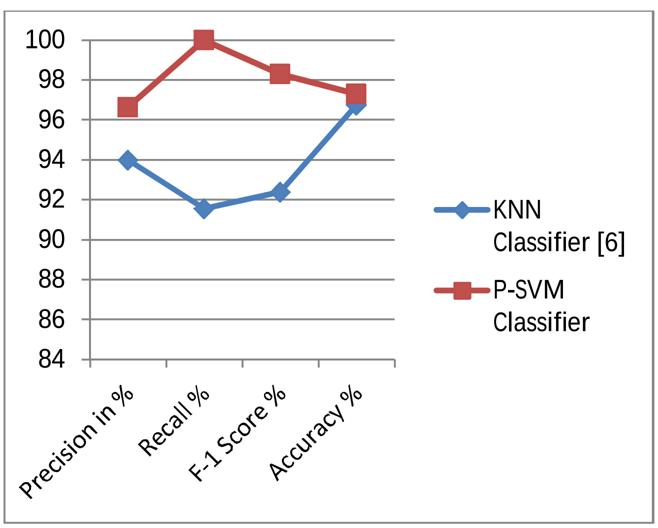


Fig. 8. Graphical Comparison of the Results

V. CONCLUSION & FUTURE SCOPE

Proposed leaf disease detection method is based on Polynomial Support Vector Machine that classifies the abnormal and normal leaves and on the basis of that decision can be make effectively with high accuracy. System has been tested with *kaggle* benchmark dataset and has proved to achieve better results as compared to the KNN classifier. The comparison of the results of the proposed method with the KNN classifier has been done on the basis of four parameters: precision, recall, F-1 score and accuracy. It has been observed that percentage increase in these four parameters as compared to the KNN classifier are 2.61, 8.44, 5.89 and 0.54 respectively. Hence, it could be concluded that the proposed classification technique is bit powerful in classifing the leaves in normal and abnormal categories and in takingood decisions about the image or disease belonging to the respective category of the disease.

In future, the system can be tested with different benchmarks that contain several images or data. The accuracy can be enhanced furhther by using certain modern image 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] Rakesh K Jha, Preety D Swami,“Intelligent fault diagnosis of rolling bearing and gear system under fluctuating load conditions using image processing technique”, Journal of Mechanical Science and Technology 34 (10), pp. 4107-4115, doi: 10.1007/s12206-020-0903-z.

[4] 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)

[5] Singh, Jaskaran & Kaur, Harpreet. (2019). Plant Disease Detection Based on Region-Based Segmentation and KNN Classifier. 10.1007/978-3-030-00665-5_154.

[6] 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.

[7] 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.

[8] Rakesh Kumar Jha, Preety D Swami “Fault diagnosis and severity analysis of rolling bearings using vibration image texture enhancement and multiclass support vector machines”, Applied Acoustics, vol. 182, pp. 1-12, November 2021, doi: 10.1016/j.apacoust.2021.108243

[9] 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.

[10] Dhaygude Sanjay B, Kumbhar Nitin P (2013), "Detection of leaf plant diseases through image processing", Int J Adv Res Electron Electro Instrum, Volume 2 (1).

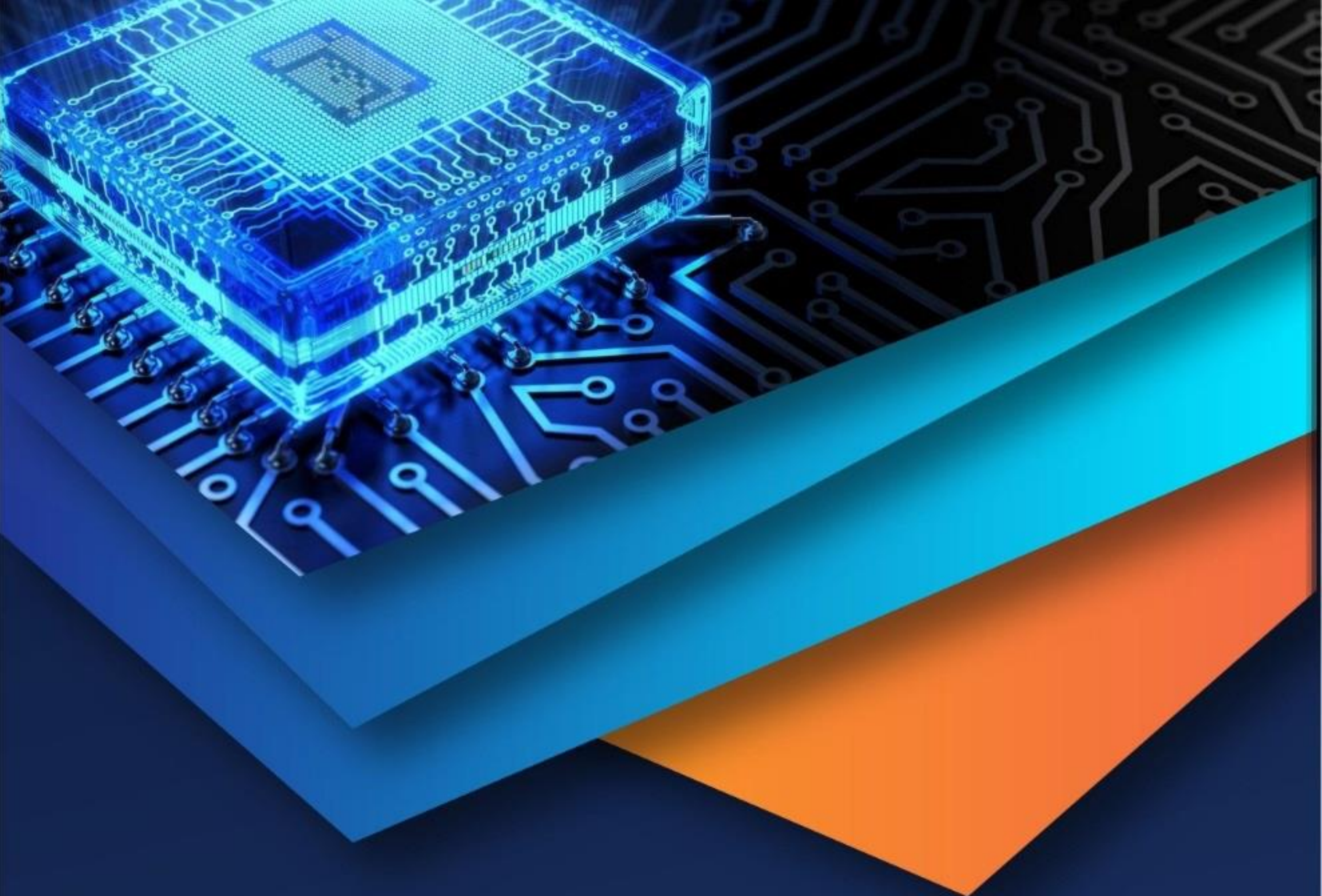
[11] Bashir Sabah, Sharma Navadepe (2012). Remote detection of plant diseases through image processing. IOSR J Electron Commun Eng; 2 (6): 31-4. ISSN: 2278-2834.

[12] Chaudhary Piyush et al. (2012). Color transformation based approach for detecting diseases in the leaves of IntComput SciTel Telecommun; 3 (6).

[13] Kulkarni Anand H., Ashwin Patil RK (2012). The use of image processing technology to detect plant diseases. Input J Mod Eng Res; 2 (5): 3661-4.

[14] 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.

[15] Brajesh Kumar Sahu, Preety D Swami , “Image Denoising using Principal Component Analysis in Wavelet Domain and Total Variation Regularization in Spatial Domain,” International Journal of Computer Applications, vol. 71, no. 12, pp. 40-46,May 2013, doi: 10.5120/12414-9183.



doi crossref

10.22214/IJRASET

INDEX COPERNICUS

45.98

ISRA JIF

IMPACT FACTOR: 7.129

SJIFactor Scientific Journal Impact Factor TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY