Classification of Leaf Diseases in Paddy Plant Based on Combined Approach of Texture and Colour Feature Extraction and Optimized Feature Selection

Nithiya S[1], AnnapuraniK[2]
{ nithiyas@srmist.edu.in[1], annapook@srmist.edu.in[2] }

Research Scholar, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur[1], Associate Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur[2]

Abstract. This paper proposes a methodology for the classification of leaf diseases by using the different characterization of shape, and colour properties. Paddy plant diseases are discussed in this research work. Bacterial light, Brown Spot, Leaf smut diseases is identified in paddy crops. A plant leaf is pre-processed at first utilizing the improved BAHE (brightness adapted histogram equalization). Leaf image is segmented using K mean clustering algorithm. The feature extraction is improved by combination of texture feature and colour features. The extracted feature is given as an input for feature selection using optimized firefly algorithm's after that it is used for the classification of diseases in paddy plant. The optimization result is compared with PSO and Genetic algorithm..

Keywords: BAHE, K mean clustering, PSO, Firefly algorithms, GA.

1   Introduction

The main goal of exploration in farming is to expand the profitability of food and improved in nature of products. The creation of agribusiness and its quality relies upon soil, seed, and agro synthetics. The farming items (vegetables and organic products) are the fundamental for the endurance of individuals. The horticultural efficiency diminishes when the harvests are influenced by different infections. The illnesses of plants are only debilitations that changes the typical plants in their crucial capacities viz., photosynthesis, fertilization, preparation and germination. Because of the impact of antagonistic ecological conditions, infections in plants are brought about by different parasites, microbes and infections. In the event that the infection isn't distinguished at beginning phase, at that point yield diminishes. In this paper, distinguishing proof of leaf infection utilizing numerous descriptors is introduced. At first the pictures are caught utilizing actual gadgets and information base is made containing solid and unfortunate leaf

This research work is used for effective feature extraction and feature selection technique. For this research work 300 leaf image is collected from Villupuram district at krishnapuram village. The collected leaf was undergone for classification process. The direct classification of image may get the chances of misclassification due to the noise and morphological changes in the image quality and capturing device. Appropriate brightness and contrast in the captured image quality is improved at the pre-processing stage.

2.Related Work

Texture analysis is used for determine the homogeneity characteristics of an image which act as the centra unit. texture appear on the leaf disease generally located on specific routine, way and scale. Surface examination dependent on the neighbourhood spatial variety of force or shading brilliance serves a significant part in numerous uses of distant detecting pictures [1,2]. Surface investigation is broadly utilized in picture division, order, what's more, design acknowledgment. Also, surface component extraction is a significant substance of surface investigation, which is a compelling strategy for taking care of the issues of unearthly variance and spatial scatter in a similar classification [3]. Essential to quantify the surface sensibly and viably, on the grounds that the removed surface highlights straightforwardly influence the nature of resulting handling. As of now, the strategies for surface data extraction are as per the following: measurable strategy, wavelet change, fractal technique, Markov arbitrary field (MRF, etc. Measurable technique is basic, simple to execute, and has solid versatility and power. Among measurable techniques, the Dark level Co-event Lattice (DLCL) is widely applied in surface portrayal [4], and the outcomes from the co-event technique are in a way that is more suitable than the surface segregation techniques [5]. The high computational intricacy and absence of worldwide data need for effective segmentation,and also understanding of the relationship between neighbour pixel is very hardwhen we focus on surface measurement.While dealt with surface data of distant detecting pictures dependent on multi-scale Wavelet change and fractal strategy are recently created. The wavelet changes are used for detecting the appropriate information. and has a decent impact on the order of surface pictures with routineness and solid directionality. Notwithstanding, for the complex characteristic pictures, it is regularly inadequate because of commotion impedance. Fractal strategy measures naturally different scale property and scale invariants that can be skilled for harshness highlights in the textural portrayal of a surface picture. Utilizing the MRF model to portray the surface highlights of distant detecting pictures can consider the nearby haphazardness and the general routineness, and notes the multi-goal of the surface, which mirrors the law of topography by and large. Nonetheless, it is fundamentally through model coefficients to distinguish the surface highlights; in this way addressing the model coefficients is troublesome, and the model boundary change isn't advantageous. The GLCM is an exemplary strategy for surface component extraction, which is successful in picture acknowledgment [7], picture division [8], picture recovery [9], picture arrangement [10], and surface investigation techniques [11,12]. The use of GLCM to extricate the surface component happens by means of the joint condition likelihood dispersion of the picture dark level to speak to surface and figures the nearby connection of pixels to acquire the surface element esteem. The GLCM is broadly utilized in numerous fields and has been consistently improved [13]. By ascertaining various bearings and window sizes of the GLCM, Pacific et al. [14] removed multi-scale surface highlights from exceptionally high-goal panchromatic symbolism. Mukherjee et al. [15] pre-processed and consolidated the surface highlights by extricating from the GLCM and utilized a BP-MLP (backpropagation-multilayer perceptron's) neural organization to order two sorts of therapeutic plants. Li et al. [16] played out a foremost segment examination of the picture and utilized the GLCM to extricate surface highlights from the initial two head segments. The surface highlights as another band joined with the first picture band all together to shape another picture, which was utilized for managed characterization. Utilizing the surface highlights separated from the GLCM, Huang et al. [17] proposed the dynamic windows calculation to group far off detecting symbolism as per the blend of dark scale and surface highlights. Rao et al. [18] extricated two-request factual

boundaries from the GLCM of the fluid precious stone surface, including contrast, energy, consistency and relationship, and recognized the stage change temperature of the gem. Teng et al. [19] chose five run of the mill surface class tests from Quick Bird information and utilized GLCM to quantitatively ascertain six factual surface highlights that were gotten by figuring the normal values in four ways and one pixel of pair-wise distance. The paper talked about which boundaries were reasonable for the particular surface arrangement. The above examination applies GLCM to all angles, furthermore, the surface highlights of pictures are separated for ensuing preparing, however less exploration takes into account the spatial directionality of surface appropriation. Periodicity, directionality and haphazardness are the three most significant factors in describing surfaces [20]

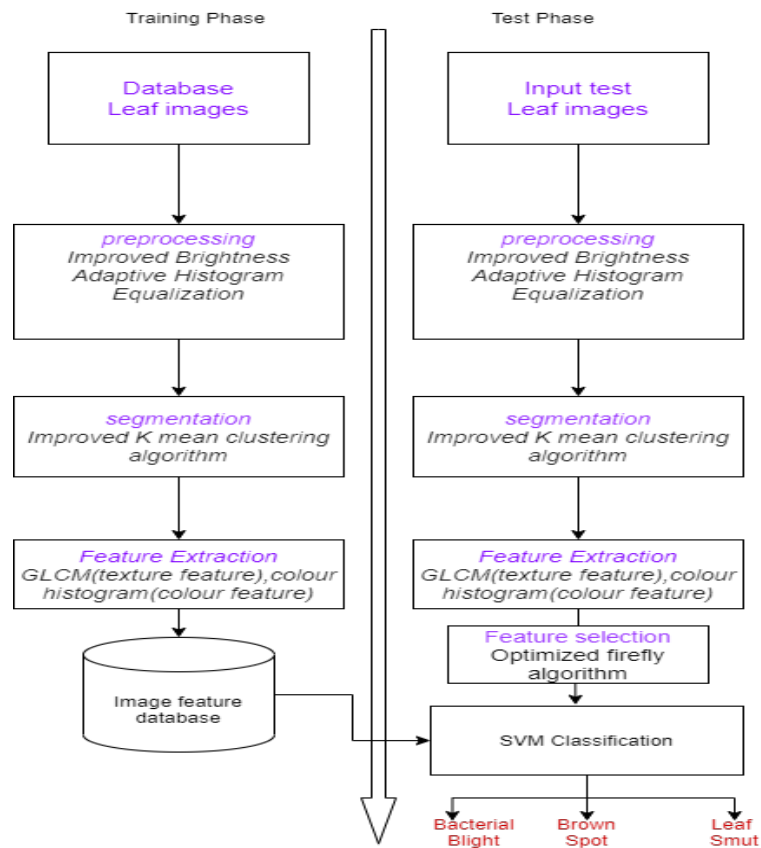## 3 Architecture Diagram Of Proposed Leaf Disease Identification In Paddy Crop System

Fig. 1. Proposed Architecture Diagram

Image was collected from Villupuram district at krishnapuram village totally 300 leaf images are collected. paddy leaf image is collected from the real time environment. Diseases image is given to the input for pre-processing step. Image quality is the important factor for leaf disease classification. In this stage image quality is improved through ImprovedBAHE

algorithm which is used for increase the proper brightness in the images. In next stage K mean clustering algorithm is used for leaf diseases segmentation. Segmented diseased leaf image is given for colour and texture feature extraction Technique. Extracted feature is Optimized using firefly feature selection algorithm. Leaf diseases are classified using Binary SVM.

3.1 PLANT DISEASES

Most plant illnesses – around 85 percent – are brought about by parasitic or contagious like creatures. Notwithstanding, other genuine illnesses of food and feed crops are brought about by viral and bacterial life forms. Certain nematodes likewise cause plant infection. Some plant sicknesses are delegated "abiotic," or illnesses that are non-irresistible and incorporate harm from air contamination, dietary insufficiencies or poison levels, and develop under not exactly ideal conditions. For the present, we'll see infections brought about by the three fundamental pathogenic microorganisms: parasite, microscopic organisms and infection. On the off chance that plant illness is suspected, cautious consideration regarding plant appearance can provide a decent insight with respect to the kind of microorganism included.

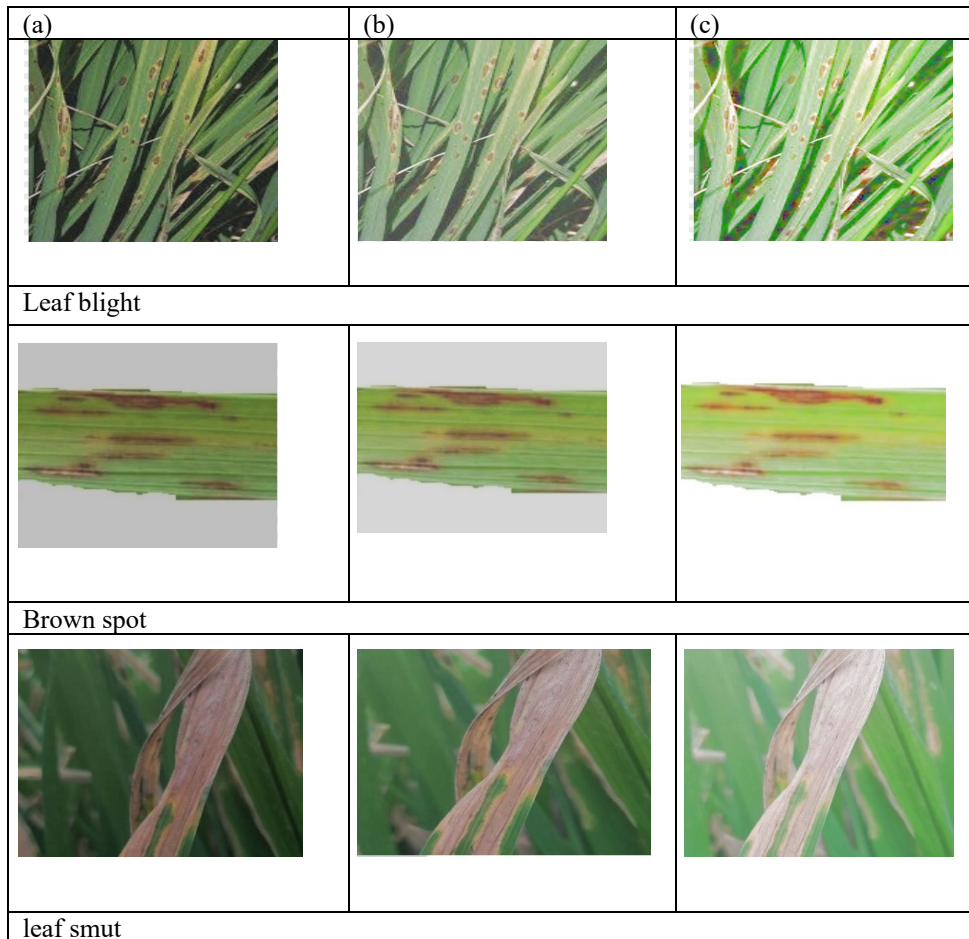| (a)	(b)	(c)
Leaf blight		
Brown spot		
leaf smut		

Fig. 2. Results for BAHE (a) Original image (b) HE (c) BAHE

In figure 2 shows the output of IBAHE algorithm. This algorithm was compared with an existing algorithm of HE (histogram equalization).in this figure 2(a) shows the health image figure 2(b) shows the output of Histogram Equalization Algorithm and figure 2(c) show the improved output of BAHE. Image quality get increased in BAHE algorithm when Compare with the real image and HE image.

3.2 Improved Brightness Adaptive Histogram Equalization Algorithm[Bahe]

This algorithm is used for preserve the brightness in the image. Using the equation 1 histogram of an image is calculated and it is constructed as the two-sub part by using median value of a histogram.

$$X_= k \text{ where } _Z(k) \geq \frac{N}{2} \tag{1}$$

Median based partition having various level of intensity values known by low range to mid-range, mid-range to high ranges in equation 2

$$\text{partition}^1: \omega_0 = \sum_{L=0}^{Xmd-1} \text{pdf}(L), \tag{2}$$
$$\text{partition}^2: \omega_1 = \sum_{L=Xmd}^{Lmax} \text{pdf}(L).$$

Normalization of histogram is the final step. Highest intensity value in each partition is applied for histogram normalization the histogram is applied to partitioned part using equation 3

$$T(L) = L - l\text{mi } L\text{mx} - L\text{mi } (L\text{up} - L\text{lw}) + L\text{lw}, \tag{3}$$

3.3 Improved K Mean Clustering Algorithm

One of the major disadvantages in k mean clustering is centre initialization. this was improved using proposed k mean clustering algorithm this consist of three parts (1) Image background removal (2) Introductory to cluster centre

3.3.1 Image Background Removal

Image is captured in Realtime environment it consists of background potion also. one of the ways to speed up the k mean clustering algorithm is background removal process. The background of white and black colour is removed using equation 4.

$$\text{Diff1} = \text{Ired-Igreen}; \tag{4}$$
$$\text{Diff2} = \text{Ired-Iblue}$$

Centre of the cluster is calculated by below equation

$$d := \sum_{\forall sj \in c}^{n} \left(\bigcup_{y_i}^{m} \frac{R_n G_n B_n}{m} \right) \in \omega' \tag{5}$$

3.3.2 Introductory to Clustering Center

The choice of starting cluster habitats is significant for precise division of infection pictures. On the off chance that a fixed grouping community is utilized, versatility of the calculation will be an issue. For arbitrarily chose beginning cluster communities, Kmeans doesn't ensure novel grouping. Subsequently, we proposed a self-learning strategy for starting cluster places in light of a division crop infection picture. The means of oneself learning strategy for beginning cluster focuses are as per the following:

1. Input: Color image
2. Output: Optimal segmented color image CIsegment
3. Initialize the image (x)
4. Remove the white and back background
5. Diff1=Ired-Igreen;
6. Diff2=Ired-Iblue;
7. Identify the peaks in the RGB histogram

8. Maximum Peak in green pixel= health part
9. Other peaks are treated as lesion
10. for i = 1 : l (pks and pts) j = 1:d(pks and pts)
11. Apply the threshold for obtain the maximum green peak region for healthy part cluster
12. Calculate the centroid for healthy part of the cluster using RGB pair equation (10)
13. Apply the threshold for obtain the less green part region for lesion part
14. Calculate the centroid for lesion part of the cluster using RGB pair equation (10)
15. Calculate the equlitian distance between centroid and cluster points
16. Mean value=(total measure / length of clustering positions)//
17. Choose the new centroid point
18. Again, calculate the distance between the centroid and RGB pairs
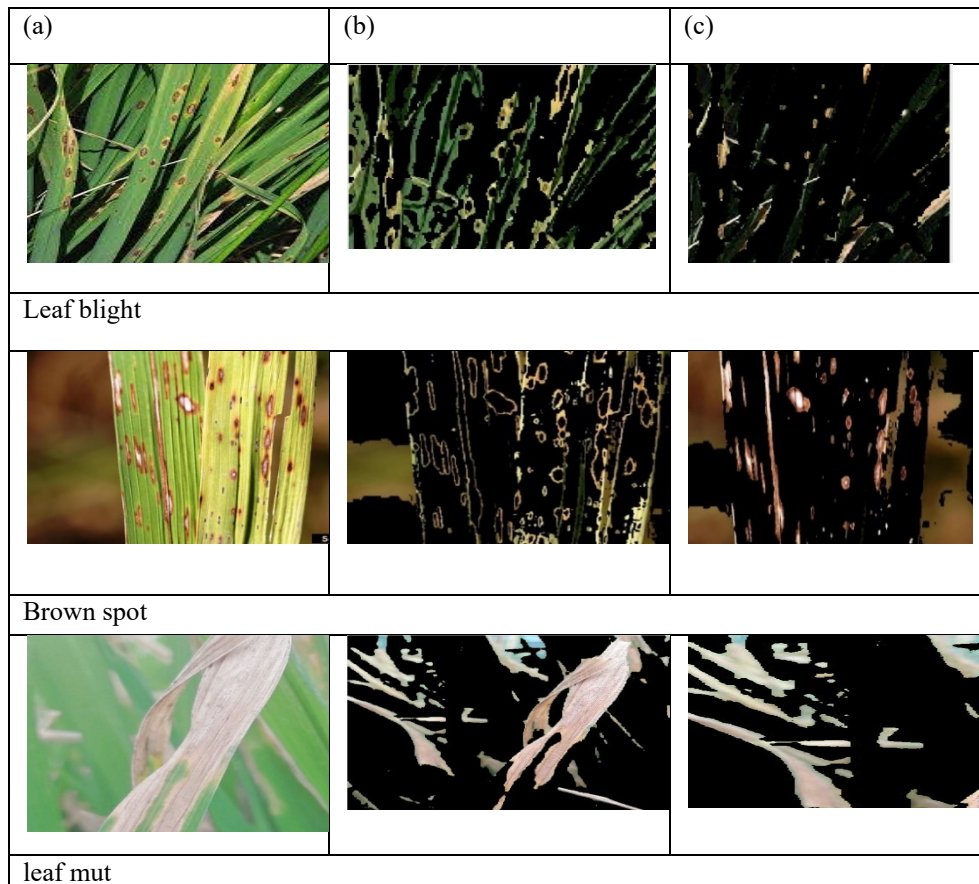19. Until the centroid position does not change

(a)	(b)	(c)
Leaf blight		
Brown spot		
leaf mut		

Fig 3: Results for K mean clustering (a) Original image (b) cluster1 (b) cluster 2

In figure 3 shows the output of K mean clustering algorithm. In this figure 3(a) shows the unhealth image figure after the enhancement of BAHE algorithm 3(b) shows the output for

cluster 1 and 3(c) show the output of cluster 2 . Various clustered output is taken into the iput for the svm

3.4 Feature Extraction Method

Combined features are extracted in this work. Colour and texture features are extracted. 10 texture feature and 27 colour features are extracted in this work. Texture features is extracted from GLCM and colour features is extracted from various colour channel mean and std deviation information.

3.4.1 Grey Level Co-Occurrence Matrix

GLCM is used for the texture feature extraction. Texture features denotes the characteristics of an area information. GLCM is used to perceive the information of various probability of intensity occurrences. GLCM is based on the second order statistic. In table1 show the feature extracted information through the GLCM matrix. Totally 10 features are extracted using the formula mentioned in table1

Table 1: GLCM features

Entropy	$$Entropy = \sum_{i,j=0}^{n-1} P(i,j)\log(p(i,j))$$
Correlation	$$Correlation = \sum_{i,j=0}^{N-1} \frac{(1-\mu_i)(j-\mu_j)P_{i,j}}{\sigma_i \sigma_j}$$
Mean	$$\mu_i = \sum_{i,j=0}^{n-1} iP(i,j)$$ $$\mu_j = \sum_{i,j=0}^{n-1} jP(i,j)$$
Variance (Sums of Squares)	$$\sigma_i = \sum_{i,j=0}^{n-1} (i-\mu_i)^2 P(i,j)$$ $$\sigma_j = \sum_{i,j=0}^{n-1} (j-\mu_i)^2 P(i,j)$$
Sum Average	$$AVER = \sum_{i,j=0}^{2n-2} iP_{x+y}(i)$$
Sum Entropy	$$SENT = -\sum_{i,j=0}^{2n-2} P_{x+y}(i)\log\left(P_{x+y}(i)\right)$$
Sum of Variance	$$SVAR = -\sum_{i,j=0}^{2n-2} (i-SENT)^2 P_{x+y}(i)\log\left(P_{x+y}(i)\right)$$
Difference Entropy	$$SVAR = -\sum_{i,j=0}^{2n-2} (i-SENT)^2 P_{x+y}(i)\log\left(P_{x+y}(i)\right)$$
Information Measure of Correlation 1	$$IMC1 = \frac{\text{Entropy-HXY1}}{max\{HX,HY\}}$$
Information Measure of Correlation 2	$$IMC2 = \sqrt{1 - \exp\left[-2.0(HXY2 - HXY)\right]}$$

3.4.2 Colour Descriptors

For this colour feature extraction various format of image is consider like HSV,LAB,RGB From each colour channel so totally 27 features are extracted.

Moment 1:mean value is calculated using the equation (6).MN is used to represent the total number of pixels

$$\mu = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{i,j}$$
(6)

Moment 2: Standard Deviation of each colour channel is calculated using equation 7. It is used to represent the root difference of a distribution

$$\sigma = \sqrt{\left[\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} \left(P_{ij} - \mu\right)^2\right]}$$
(7)

Moment 3 : Skewness and kurtosis is measured by equation (8)(9).This equations are used to calculate the Asymmetric and peakedness

$$\theta = \sqrt[3]{\left[\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} \left(P_{ij} - \mu\right)^3\right]}$$
(8)

$$\gamma = \sqrt[4]{\left[\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} \left(P_{ij} - \mu\right)^4\right]}$$
(9)

3.5 Optimized Feature Selection Using Firefly Algorithm

Firefly algorithm is based on the attraction's properties of each firefly. Lesser bright firefly will be attracted by highest brightness firefly. If the distance between 2 fireflies increases then the attractiveness also reduced.

1. Objective function f(x), x = (x1, ..., xd)
2. Generate initial population of fireflies xi (i = 1, 2, ..., n)
3. Initialize iteration (max Generation)
4. Initialize the dimension(D)
5. Randomly initialize the position of the fireflies
6. For i=1:n all n fireflies
7. For j=1 to I all n fireflies
8. while (t <max Generation),
9. Calculate the fitness function
10. If(j>i) move the fireflies i towards j
11. Evaluate the new solution and update the light intensity
12. End for i
13. End for j
14. Rank the fireflies and find the optimal selection

optimized feature was compared with other features selection algorithm like PSO and GA. The performance of firefly algorithm is high than the others

Table2: Comparison table of feature selection algorithm

Algorithm	Iteration	Selected features
PSO	1000	27
GA	1500	25
Firefly algorithm	500	10

Table 2 shows the feature selection comparison between various algorithm. Firefly algorithm obtained the better result than the others which choose the optimal features.

3.6 SVMClassification

SVM is used as the classification algorithm for plant disease identification. Optimally selected feature is given for the classification process.In this research work denotes the binary SVM it is used to map the datapoint in two classes.
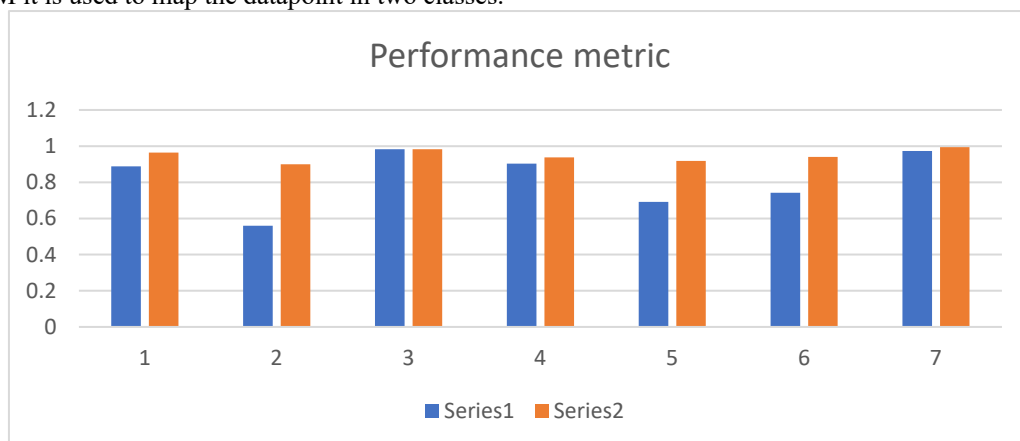

Fig.4. SVM performance

Given input features is separated by hyperplane's which is compared with KNN andSVM.Theperformance of an SVM is better than KNN by analysing various performance metrics which is shown in table 3

Table 3. SVM performance metrics

		KNN	SVM
Sensitivity	$SN = \dfrac{TP}{TP + FN}$		
Specificity	$SP = \dfrac{TN}{TN + FP}$	0.8879	0.9641
Accuracy	$Acc = \dfrac{\sum TP + \sum TN}{\sum TOTALELEMENT}$	0.5600	0.9000
Precision	$Precision = \dfrac{TP}{TP + FP}$	0.9827	0.9827
F_measure	$Acc = \dfrac{2(percision * sensitivity)}{(percision + sensitivity)}$	0.9032	0.9375
Goodness mean	$GM = \sqrt{\dfrac{TPR}{TNR}}$	0.6914	0.9184
Area under curve	$AUC = \int f(a)da$	0.7418	0.9404
		0.9729	0.9946

CONCLUSION

This research work is successfully executed by MATLAB 2020R version. In this research work conclude the many advancement technique for plant diseases identification and classification purpose. Initially at pre-processed stage improvement in histogram equalization algorithm is adopted and it is compared with existing algorithm and K mean clustering algorithm used to segmentation purpose. Many Mathematical derivation is used for texture and feature extraction. by combining the texture and colour feature 37 features are extracted. These 37 features are given to the input for PSO, GA and firefly for feature selection process. Thefirefly performance is high when compare with GA and PSO

References

[1] Kitada, K.; Fukuyama, K. Land-use and land-cover mapping using a gradable classification method. Remote Sens. 2012, 4, 1544–1558.

[2] Racoviteanu, A.; Williams, M.W. Decision tree and texture analysis for mapping debris-covered glaciers in the Kangchenjunga area, Eastern Himalaya. Remote Sens. 2012, 4, 3078–3109.

[3] Myint, S.W.; Lam, N.S.N.; Tyler, J.M. Wavelets for urban spatial feature discrimination: Comparisons with fractal, spatial autocorrelation, and spatial co-occurrence approaches. Photogramm. Eng. Remote Sens. 20014, 70, 803–812.

[4] Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. IEEE Trans. Syst. Man Cybern. 2015, SCM-3, 610–621.

[5] Davis, L.S. A survey of edge detection techniques. Comput. Graph. Image Process. 2014 , 4, 248–270.

[6] Huang, D.S.; Wunsch, D.C.; Levine, D.S.; Jo, K.H. Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues (ICIC). In Proceedings of the Fourth International Conference on Intelligent Computing, Shanghai, China, 15–18 September 2018.

[7] Zhang, Y. Optimisation of building detection in satellite images by combining multispectral classification and texture filtering. ISPRS J. Photograms. Remote Sens., 54, 50–60.

[8] Sahoo, M. Biomedical Image Fusion and Segmentation using GLCM. In Proceedings of the 2nd National Conference-Computing, Communication and Sensor Network (CCSN), Orissa, India, 29–30 October; pp. 34–39.

[9] Kekre, H.; Thepade, S.D.; Sarode, T.K.; Suryawanshi, V. Image retrieval using texture features extracted from GLCM, LBG and KPE. Int. J. Comput. Theory Eng. 2010, 2, 695–700.

[10] De Almeida, C.W.D.; De Souza, R.M.C.R.; Candeias, A.L.B. Texture Classification Based on Co-Occurrence Matrix and Self-Organizing Map. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 10–13 October 2010; pp. 2487–2491.

[11] Srinivasan, G.N.; Shobha, G. Segmentation Techniques for ATDR. NAUN Int. J. Comput. 2008, 2, 165–171. Sensors 2017, 17, 1474 15 of 15

[12] Tuceryan, M. Moment-based texture segmentation. Pattern Recognit. Lett. 1994, 15, 659–668.

[13] Liu, J.; Cheng, Y.; Sun, J. Modified FCM SAR image segmentation method based on GMCL feature. Comput. Eng. Des. 2012, 33, 3502–3506.

[14] Pacifici, F.; Chini, M.; Emery, W.J. A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. Remote Sens. Environ. 2009, 113, 1276–1292.

[15] Mukherjee, G.; Chatterjee, A.; Tudu, B. Study on the Potential of Combined GLCM Features towards Medicinal Plant Classification. In Proceedings of the 2nd International Conference on

Control, Instrumentation, Energy & Communication (CIEC), Kolkata, India, 28–30 January 2016; pp. 98–102.

[16] Li, Z.; Zhu, G.; Dong, T. Application of GLCM—Based texture features to remote sensing image classification. Geol. Explor. 2011, 47, 456–461.

[17] Huang, X.; Yang, W. Classification of remotely sensed imagery according to the combination of gray scale and texture features based on the dynamic windows. J. Geomat. Sci. Technol. 2015, 32, 277–281.

[18] Rao, C.N.; Sastry, S.S.; Mallika, K.; Tiong, H.S.; Mahalakshmi, K. Co-occurrence matrix and its statistical features as an approach for identification of phase transitions of mesogens. Int. J. Innov. Res. Sci. Eng. Technol. 2013, 2, 4531–4538.

[19] Teng, X.; Song, S.; Zhan, Y. Statistical Class Feature in Texture Analysis of Remote Sensing Imagery. Adv. Environ. Sci. Eng. 2012, 518–523, 5749–5753.

[20] Liu, F.; Picard, R.W. Periodicity, directionality and randomness: Wold features for image modeling and retrieval. IEEE Trans. Pattern Anal. Mach. Intell. 1996, 18, 722–733.

[21] Chen, B.; Zhang, Y.; Chen, L. RS Image classification based on SVM method with texture. Eng. Surv. Mapp. 2007, 16, 23–27.

[22] Zhao, L.; He, H.; Yin, Z. Texture direction adaptive image interpolation based on Curvelet transform. J. Optoelectron. Laser 2012, 23, 799–804.

[23] Puissant, A.; Hirsch, J.; Weber, C. The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery. Int. J. Remote Sens. 2005, 26, 733–745.

[24] Zheng, S.; Zheng, J.; Shi, M.; Guo, B.; Sen, B.; Sun, Z.; Jia, X.; Li, X. Classification of cultivated Chinese medicinal plants based on fractal theory and gray level co-occurrence matrix textures. J. Remote Sens. 2014, 18, 868–886.