IDENTIFICATION OF HEALTHY AND DISEASED PADDY LEAVES USING kNNCLASSIFIER

Suresha M
Assistant Professor,
Department of P.G. Studies in Computer Science,
Kuvempu University, Karnataka, India.

Shreekanth K N
Research Scholar,
Department of P.G. Studies in Computer Science,
Kuvempu University, Karnataka, India.

HarishaNaik T
Research Scholar,
Department of P.G. Studies in Computer Science,
Kuvempu University,Karnataka, India.

Abstract: In agricultural field recognition of healthy or diseased plant was the major constrain. The control period was habitually missed because the traditional approaches are taking long time to analyse. Becausean expert has to visit agricultural field repeatedly, it may take more time to analyse, and it may be burden to the farmer. In this proposed work, recognition of healthy and diseased leaves is done using LBP features. Global thresholding Otsu methodis used for segmentation andkNN classifier is used to classify the data,and found a success rate of 84.43%. Interpolation with scalar values is done for smoothing an image.

Keywords: Disease, kNN classifier, Segmentation, Texture Features**.**

I INTRODUCTION

Rice is a monocotyledonous angiosperm and it belongs to *Oryza*. It contains more than 20 species, yet just two species are alluded to as development ie*Oryzasativa* and *Oryzaglaberrima* developed in south-east Asia and West African nations. Basically rice was developed in the tropical locale of Asia, and from the past record rice is from 5000 years BC, and there after reached out to calm areas [23]. Rice is a fundamental staple nutrition food in Asia. In an Asiancountry rice is the major leading food which isgrown all over the worldie 90%. Three billion Asianpeople were intake caloric in the ratio of 35 – 60% [7]. Plant development was suggestively decreased with compared to un-inoculated plants and inoculated plants. Blast disease indications of Rice are also produced in inoculated plants [10]. Rice Brown Spot is the evolving disease in low fertility and marginal areas which causes substantial yield loss [2]. Initially brown spot lesions and blast diseased lesions are difficult to distinguish.Blast lesions turn to diamond or spindle shape after maturity [4].Rice (*Oryzasativa*) blast, instigated by Magnaporthegrisea, which is one of the most harmful agent of rice. Many studies have been conducted to resist mechanisms of rice to the blast fungus and thus several antifungal substances have been isolated from rice leaves [14]. In particular, seven parasites Curvularialunata,Magnaportheoryzae,Fusariummoniliforme, Helminthosporiumoryzae, Nigrosporaoryzae, Aspergillusniger and Alternariaalternata were disconnected from influenced leaves and seeds contaminated leaves of five rice assortments viz., IRRI-8, IRRI-6, KS-282, DR-82, and Shua. N. Oryzae is disconnected just from IRRI-8 and IRRI-6 rice assortments. Among these growths, M. Oryzae was, for the most part, disengaged both from leaves and

seeds of the rice assortments considered. M. Oryzae was detected with maximum frequency from seeds and leaves of IRRI-6 [10].Brown Spot diseases is one of the most significant diseases in rice, Brown spot is caused by the fungus Bipolarisoryzae (earlier identified as Helminthosporiumoryzae)can affect all plant parts of the paddywhich will affect the quality of grain and yield losses. Brown spot symptoms primarily appear as small circular to oval spots on the leaves. Yield loss, shape, size, and color these dependon and and these are estimated based on environmental conditions and rice variety, 16 to 40% of yield loss observed in Florida. Large spots have a light greycentre enclosed by a dark to reddish-brown margin or reddish-brown while small spots are dark brown to reddish brown. Elder spots may have a bright yellow halo surrounding the lesion. Spots on the hulls and leaf sheath are similarto those on the leaves [4].Diseases are most important causes for quality and quantity reduction in agricultural products. Plant diseases also can harm and wreck the agricultural fields. The tungro is the most important rice virus disease; it is transmitting through green leafhopper [5] and it is the one of most destructive diseases of rice at the early stage of the tungro.Disease symptoms are absent and symptoms appear only in the later stages [6]. Rice plants infected with tungro symptoms including stunting growth which leads to yellow to orange discoloration and reduced tillering[9]. Leaves may discolour and shows irregular shaped dark-brown blotches. The younger leavesshow striping or mottling along with intervernalchlorosis[8].Bacterial blight is one of the most devastating diseases of rice worldwide and is found both in tropical and temperate regions [19]. The disease is accountable for a loss of 20-30% in different countries. In India losses in yield varied from 6-60% in different states

depending upon stage and severity of infection and type of cultivars [21]. Leaf blightisone of the most common syndrome, it begins like water-soaked stripes on the leaf blades. Then stripes enlarged in area and become yellow to white and enclosed entire leaf blade. Water-soaked margins are small circular lesions, they appear in severe infections, and infected plants produce poor quality of grain. Bacterial leaf blight is produced by Xanthomonasoryzaepv.*oryzae*. It is Gram-negative, non-spore forming rod [1].Bacterial Brown Stripe also known as bacterial stripe occurs in nurseries and also distributed widely in rice growing countries [20].The symptoms of this disease in seedling stage are inhibition of germination occurring brown stripes interveinally or margins of leaf sheath is curving and abnormal elongation of mesocotyl. Manifestations were covered after the contaminated seedlings were transplanted to paddy fields. The regular event of the illness was seen where rice plants were submerged by overflowed water [19]. Bacterial dark colored stripe is caused by Pseudomonas syringaepv. panici. is a Gram-negative, non-spore-shaping, non-embodied pole [20].

This paper is organized as follows: section I is aboutintroduction to leaves diseases,section II gives brief literature review it includes techniques pre-processing, segmentation, feature extraction and classification, section III explains the proposed work for our approach, section IV explains experimental result obtained in our approach, and section V summarizes our contribution and focuses on possible directions for further research.

II   LITERATUREREVIEW

In [13] identified causes of diseases in cotton plant. Diseases are commonly found in all agricultural plant by nutrient deficiency. In this work author identified alternaria leaf spot disease by using template matching and color histogram model. Color histograms were used to identify nutrient deficiency like potassium and sulphur in cotton plant leaf.Image segmentation is carried out by using statistical region merging algorithm.In [12] proposed a method for image segmentation in two ways using global thresholding and local adaptive thresholding. Author converted the RGB images to YCbCr component.Otsu segmentation technique with global thresholding and adaptive thresholding has been applied for segmentation. Authors conclude that this type of segmentation is used for blurred images using local thresholding's.In [17] proposed a novel method for identification of agricultural leaves and obtained 77.96% accuracy. Author used Arahar, Bathua, Wheat, Bearseem Clover and GajarGhas leaves for classification. Otsu's segmentation method is used to threshold the gray level images of leaves and normalization of leaves are done to avoid the shape disorder like size, orientation of the leaves and shape features. Using boundary box properties images are rotated towards vertical position based on major axis points. Binary classifiers are used based on SVM for an individual class and also feature vectors are generated using PCA. Author generate the eigen leaf similar to eigenface from individual class and Euclidian distance were calculated using eigen levels, the minimum and maximum values of this Euclidian distance assist as feature vector for their respective classes of leaves. By calculating same weights for unknown sample, classification of leaves has been done by using SVM classifier. In [3] proposed a method for recognition of plant using leaves color and shape features, using k-Nearest Neighbour (kNN), Support Vector Machines (SVM), Naïve Bayes, and Random forest classifier are used and obtained 96% of accuracy with random forest classifier. In first stage shape features are extracted and in second stage color feature like mean and standard deviation features are extracted for individual component of R,G,B component and average of these channels are also taken. These features are measured as first set and second set consists of color histograms of R,G, and B component. Author provides efficient representation of color distribution with respect to plant leaf. Random forest classifier has given good result among the different classifier used by the authors. Only shape, a combination of shape and color features of the first and second set have been used for random forest classifier.In [24] proposed a method for identification of the leaf diseases in cucumber plant and obtained 80% accuracy using Back Propagation Neural Networks (BPNN). A pre-processing median filter has been applied for denoising. Extraction of background has been done, quad tree a region segmentation method which is used prewitt operator for processing an image. Otsu segmentation method is used for segmentation of gray image, and morphological operation like expansion, erosion, opening and closing operation were done for better binary image. RGB individual component are considered for feature extraction, shape features like area and euler number were extracted, texture feature like Energy and Contact were extracted and using hue component of HSV color space feature is extracted in two moments for this features Back Propagation Neural Networks (BPNN) is used to classify the disease.

III   PROPOSEDWORK

Plant also suffers from various diseases by disease causing agents like viral, bacterial and fungal,Itreduces the yield loss and quality of rice grain. The disease gives the impression on leaf surface. In this proposed work identification of healthy and diseased leaves are carried out. Segmentation of healthy and diseased part is the major task. Global thresholdingotsu segmentation technique is used to segment unhealthy part. LBP Features are extracted from segmented region and the features are normalized using equation (1) and kNNclassifier is used to classify the data. Proposed methodology work flow is shown in figure 1.
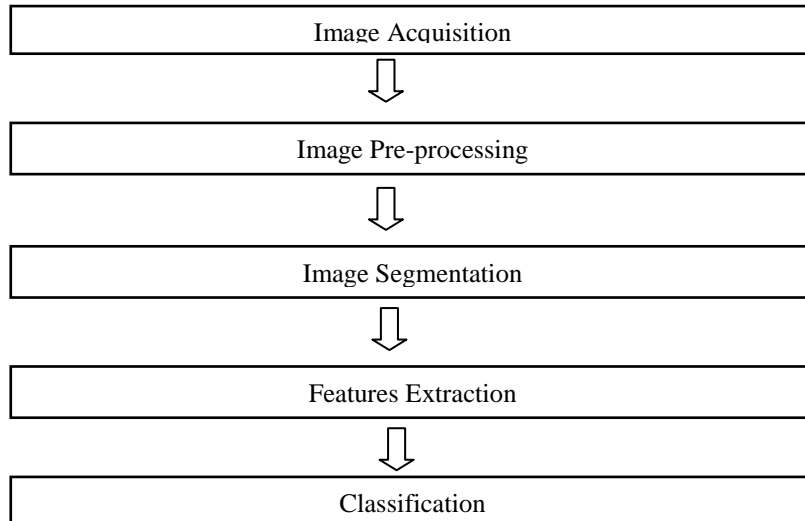
$$\sum_{i=1}^{n}\sum_{j=1}^{n}\left(f(x,y)\right)/N \Big/ N \qquad (1)$$

Figure 1. Typical work flow for proposed methodology

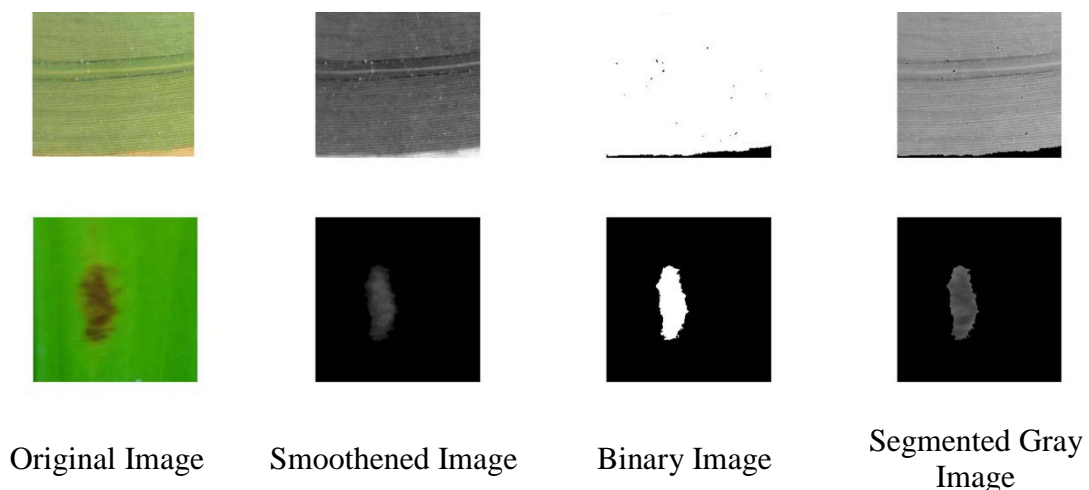| Original Image | Smoothened Image | Binary Image | Segmented Gray Image |

Figure 2. Sample and Segmentation results

A. *Image acquisition*:

Sampled Images are captured using sony digital camera with 18.1 megapixels in day light because clear view of leaf object is required for analysis. Sample diseased and healthy paddy leaves are shown in figure 2.

B. *Image Pre-processing:*

Pre-processing is the preliminary step before segmentation and feature extraction. In this proposed work images are resized to 255 X 255 to improve the accuracy of an algorithm. Individual RGB components are extracted for image smoothening. Here the G and B components are interpolated with scalar values and the resultant value is subtracted with R component and multiplied with a value shown in equation (2). Images are smoothened and then enhanced. In resultant image the healthy part and diseased part can be easily identified because the diseased part appears darker compared to other part of the leaf using equation (2). Same type of other calculation for R, G, B components are conducted [16].

$$SIM = 2 \ X \ (R - (0.75 \ X \ G - \ 0.25 \ X \ B)) \qquad (2)$$

C. *Image segmentation:*

In image analysis and pattern recognition segmentation is the major task, partitioning of an image into groups of pixels based on similarity measurement is known as segmentation. It helps in the feature extraction from the segmented image based on the interested area or based on the objective which is considered for the experiment [11,18]. Pre- processing of images is significant task. In proposed methodology image smoothening were carried out using equation (2). For resultant images segmentation are carried out by using global thresholding method of Otsu segmentation. Experimental segmented images are shown in figure 2.

D. *Feature extraction:*

The local binary pattern texture feature information extraction was proposed by Ojala et al., [15].Based on theory,texture has locally two corresponding parts like a pattern and its strength with the goal of texture

classificationand then it is extended to face detection, face recognition, face expression recognition, disease identification [22] etc.

The advantages of LBP are it is invariance to monotonic gray-scale modifications, convenient multi-scale extension and minimum computational. Traditional and statistical structural methods are well designed in LBP. In an input image each processing pixel is matched with its eight neighbours and ones whose intensities exceed the processing pixels are marked as 1, else as 0, in resultant a circular point features consisting of only binary bits were measured. Naturally the feature ring is considered as row vector, and binomial weight assigned to each bit. Row vector is transformed to decimal for further work. Local Binary Pattern using circular neighbourhoods and linearly incorporating the pixel values allows the choice of any radius, R, and number of pixel in the neighbourhood, P, to form an operator, which can model large scale structure.Basic LBP is noise sensitive, because the threshold value is the processing pixel, so average of pixels including processing pixel that encompass by an LBP operator as a threshold. The mathematical model for LBP is shown in equation (3)

$$LBP_{P,R}(x,) = \sum_{p=0}^{p-1} s\left(g_p - g_c\right)2^p \quad (3)$$

Where g_c is the average gray value of the pixels and g_p is the intensity value of pixels in eight neighbourhood.

$$h(i) = \sum_{x,y} B\left(LBPP, R(x,y) = i\right) \mid i \in [0, 2^{p-1}], \quad (4)$$

$$B(v) = \begin{cases} 1 & v > T \\ 0 & Otherwise \end{cases}$$

For texture analysis histogram is a descriptor *h(i),* of the LBP shown in equation (4) and its advantages is, it is invariant to image translation.

In the proposed work the smoothen image obtained by equation (2) are passed to extract LBP features with normalized block size [32 X 32] and [255 X 255] for normalization of images and histogram is generated, after reconstruction of feature of LBP which is used for classification of the data.

E. Classification:

In this proposed work supervised classifier is used for classification of healthy and diseased paddy leaves. Training data set has a combination of order set of features vector and corresponding label vector. Let's say t_x is a feature vector and l_i is the corresponding label vector. Testing data set has an ordered set of features without label. In testing data set, label is unknown whereas training data set has labels with respect to corresponding features. kNN classifier at first fixes k nearest neighbors and then fixes the labels for the test sample based on neighbor weight. Consider a training set T has $x_i$training samples. The training features and testing features are normalized to [0,1]. The training data samples are labeled with class label y_o. The objective is to classify testing sample q for each $x_i \varepsilon D$ and the distance between q and x_i is calculated using equation (5).

$$d(q, x_i) = \sum_{f \varepsilon F} w_f \delta(q_f, x_{i f}) \quad (5)$$

The equation (6) showsfor finding k nearest neighbors is nominated for continuous and discrete attributes.

$$\delta(q_f, xi_f) = \begin{cases} 0 & f \quad discrete\ and\ q_f = x_{if} \\ 1 & discrete\ and\ q_f \neq x_{if} \\ |q_f - x_{if}| & f \quad continuous \end{cases} \quad (6)$$

These are different variety of approaches to identify k nearest neighbors to fix the class. The most general method is majority rule for assigning a label to unknown sample.

LBP features have been extracted from the difference of different components of RGB with block size 255 X 255 and 32 X 32. The equations for these are given in equation (7) to (9).

$$d(GRAY, R) = GRAY(x, y) - R(x, y) \quad (7)$$
$$d(GRAY, G) = GRAY(x, y) - G(x, y) \quad (8)$$
$$d(GRAY, B) = GRAY(x, y) - B(x, y) \quad (9)$$

Where   GRAY – Gray component,
           R – Red component,
           G – Green component,
           B – Blue component.

IV   EXPERIMENTAL RESULTS

The collected leaves images are resized to 255 X 255 pixel resolution to improve the algorithm accuracy. Image smoothening have been done by using equation (2), segmentation of leaves is carried out by using otsusegmentation method. In this proposed work LBP features are extracted for segmented images of individual component of RGB color space and also for gray images. The data set difference is calculated by subtracting the R,G and B components with gray component(Here the similarity between LBP Features are computing the squared error between them). For this resultant data set classification processes arecarried out individually to observe the accuracy of the feature for difference data set and also for individual component of R, G, B and gray. Testing and Training feature data set arepassed to kNN classifier and obtained result is shown in table 1.

Table 1: classification of Results for different data set with different normalized block size.

Normalized Block Size	Sample data set	Accuracy
255 X 255	Gray images	77.71%
	Similarity difference of Gray with Red component	70.48%
	Similarity difference of Gray with Green component	75.30%
	Similarity difference of Gray with Blue component	78.91%
	Red component	79.57%
	Green component	79.51%
	Blue component	76.50%
32 X 32	Gray images	84.93%

Similarity difference of Gray with Red component	86.14%	
Similarity difference of Gray with Green component	83.13%	
Similarity difference of Gray with Blue component	80.72%	
Red component	84.93%	
Green component	84.93%	
Blue component	84.93%	

In this proposed work different data set is considered with kNN classifier, in this experimentation, the features extracted from red component is used for training and the features extracted from green component is consider for testing which has given a success rate of 84.93%. when the similarity difference resultant is considered for the same type of components for training and testing which reduces the accuracy rate. So, by this observation 84.93% of accuracy given by individual component would be the better result. Comparatively, when the block size is small, our proposed work will give better results.

V   CONCLUSION

In this paper, classification of healthy and diseased paddy leaves is identified using kNN classifier with Local Binary Pattern (LBP) features. There are three types of diseases called viral, bacterial and fungal. In this paper, we have considered only identification of healthy and diseased leaf. In the future work, we will consider mentioned three types of diseases.

REFERENCE

[1] Ansari, M. M., & Sridhar, R. (2012). Iron nutrition and virulence in Xanthomonasoryzaepv. oryzae. Indian Phytopathology.

[2] Aryal L, Bhattarai. G, SubediA Subedi .M , Subedi .M &Sah. G.K. (2016, April-June) Response of Rice Varieties to Brown Spot Disease of Rice at Paklihawa, Rupandehi, Global Journal of Biology, Agriculture and Health Sciences, ISSN: 2319-5584, G.J.B.A.H.S., Vol.5(2) pp: 50-54.

[3] Caglayan, A., Guclu, O., & Can, A. B. (2013, September). A plant recognition approach using shape and color features in leaf images. In International Conference on Image Analysis and Processing (pp. 161-170). Springer, Berlin, Heidelberg.

[4] Datnoff, L. E., &Lentini, R. S. (1994). Brown spot in Florida rice.

[5] Daud, S. M., Jozani, H. J., & Arab, F. (2013). A review on predicting outbreak of tungro disease in rice fields based on epidemiological and biophysical factors. International Journal of Innovation, Management and Technology, 4(4), 447.

[6] Gnanamanickam, S. S. 2009. Major diseases of rice, in *Biological Control of Rice Diseases,* Springer publications, ch. 2, pp. 32-34.

[7] Guyer, D., Tuttle, A., Rouse, S., Volrath, S., Johnson, M., Potter, S., ...& Ward, E. (1998). Activation of latent transgenes in Arabidopsis using a hybrid transcription factor. Genetics, 149(2), 633-639.

[8] Hibino, H. (1996). Biology and epidemiology of rice viruses. Annual review of phytopathology, 34(1), 249-274.

[9] Hibino, H., Roechan, M., &Sudarisman, S. (1978). Association of two types of virus particles with penyakithabang (tungro disease) of rice in Indonesia. Phytopathology, 68(10), 1412-1416.

[10] Jamal-u-ddinHajano, M. A., Pathan, Q. A., &Mubeen, A. L. (2011). Rice blast-mycoflora, symptomatology and pathogenicity. International Journal for Agro Veterinary and Medical Sciences, 5(1), 53-63.

[11] Jayaraman, S., Esakkirajan, S., &Veerakima, T. (2009). *Digital image processing*, ed iii.

[12] Mandal, A. K., &Baruah, D. K. (2013). Image Segmentation Using Local ThresholdingandYcbcrColor Space. Int J Eng Res Appl, 3(6), 511-514.

[13] Miyatra, A., &Solanki, S. (2014). Disease and Nutrient Deficiency Detection in Cotton Plant. International Journal of Engineering Development and Research, 2(2), 2801-2804.

[14] Ohta, H., Shida, K., Peng, Y. L., Furusawa, I., Shishiyama, J., Aibara, S., & Morita, Y. (1991). A lipoxygenase pathway is activated in rice after infection with the rice blast fungus Magnaporthegrisea. Plant Physiology, 97(1), 94-98.

[15] Ojala, T., Pietikainen, M., &Maenpaa, T. (2002). Multiresolutiongray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence, 24(7), 971-987.

[16] Phadikar, S., Sil, J., & Das, A. K. (2012). Classification of Rice Leaf Diseases Based onMorphological Changes. International Journal of Information and Electronics Engineering, 2(3), 460.

[17] Puja, D., Saraswat, M., &Arya, K. V. (2013). Automatic Agricultural Leaves Recognition System. In Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012) (pp. 123-131). Springer, India. DOI: 10.1007/978-81-322-1038-2_11, Springer India

[18] Rafael C. Gonzalez, Richard E. Woods & Steven L. Eddins. (2013 ninth reprint). *Digital Image Processing Using MATLAB.* McGraw Hill Eduction (India) Private Limited, ISBN 978-0-07-070262-2.

[19] Saha, S., Garg, R., Biswas, A., &Rai, A. B. Bacterial Diseases of Rice: An Overview.

[20] Shakya, D. D., Vinther, F., &Mathur, S. B. (1985). Worldwide distribution of a bacterial stripe pathogen of rice identified as Pseudomonas avenae. Journal of Phytopathology, 114(3), 256-259.

[21] Singh, G. P., Srivastava, M. K., Singh, R. V., & Singh, R. M. (1978). Variation in quantitative and qualitative losses caused by bacterial blight in different rice varieties [India]. Indian Phytopathology.

[22] Suresha, M., Danti, A., & Narasimhamurthy, S. K. (2014). Classification of Diseased Arecanut based on Texture Features.

[23] Watanabe, Y., Matsuo, T., Futsuhara, Y., Kikuchi, F., & Yamaguchi, H. (1997). Genomic constitution of genus Oryza. Tokyo: Food and Agriculture Policy Research Center.

[24] Wei, Y., Chang, R., Wang, Y., Liu, H., Du, Y., Xu, J., & Yang, L. (2011, October). A study of image processing on identifying cucumber disease. In International Conference on Computer and Computing Technologies in Agriculture (pp. 201-209). Springer, Berlin, Heidelberg.