*R*esearch Article

Application of machine learning in detection of blast disease in South Indian rice crops

S. Ramesh*, D. Vydeki

Department of Electronics and communication Engineering, VIT Chennai, India

ABSTRACT

It is a well-known fact that the quality and quantity of the rice crop is reduced due to plant disease. This paper proposes rice blast disease detection mechanism using Machine learning algorithm, to identify the disease in the early stage of the crop cultivation. The proposed method would find the blast disease and reduce the crop loss and hence increase the rice agriculture production in an effective manner. The images of the paddy field are captured and eight features are extracted to distinguish the healthy and the disease affected leaves. The proposed machine learning based classification methodology includes KNN and ANN. The performance of these two classification techniques is compared using an appropriate confusion matrix. The simulation results show that KNN based classification method provides an accuracy of 85% for the blast affected leaf images and 86% for the normal leaf images. The accuracy is improved to 99% and 100% respectively for the ANN based classification mechanisms.

INTRODUCTION

As Indian population is growing fast, the production of agriculture has to be increased in rapid manner. Rice is the most important food crop in India [1, 2]. However, it is one crop which tends to be much affected by disease causing agents resulting in reduced yield. Although several challenges affect the crops, such as pest, environmental conditions and natural diseases, crop getting infected by diseases is the major problem in rice cultivation [3,4]. The diseases are caused by bacterial or fungal and can affect the crop in any stage, such as main field stage or nursery stage.

Blast, Bacterial leaf blight, tungro, sheath blight, sheath rot, and brown spots are few of the bacterial and fungal generated diseases that severely affect the rice crop cultivation process [5]. As per the report of Tamilnadu Agriculture University (TNAU), expected grain loss is 70 to 80% of the damage in the rice crop is due to the blast disease [6]. The blast disease may affect very small portion of the leaf during the nursery stage and even spread to entire cultivation resulting in huge loss to the farmer.

This paper presents a novel machine learning algorithm based blast disease detection in rice crops using KNN and ANN Classification techniques.

BACKGROUND OF RESEARCH WORK

In rice crops, the problems can occur by pests, diseases and environmental conditions. The pests can be controlled by applying the proper chemical Pesticides in the field. The environmental parameters like water irrigation, soil moisture and animal invasion in the field are monitored and controlled by automated technology. Several diseases can be controlled by monitoring and implementing proper farming methods. Blast disease is the common factor affected by the rice crop wherever it is cultivated [6]. This disease is caused by bacterial or fungal and it could affect the crop either in nursery stage or main field. Inspite of applying pesticides, it is difficult to control the rapid spreading of this infectious disease.

The sample images shown in Figure 1 and 2 indicate the severity of the rice plant infection due to various diseases [6, 1].

As shown in Figure. 1.a, blast disease might appear as small spots on leaf lamina and gradually develop to spindle shaped spots of size 0.5 to 1.5cm length, 0.3 to 0.5cm 0.5 width, with ashy spot in centre.

Figure.1.b illustrates the leaf of rice plant infected by the bacterial leaf blight; this appears as water-soaked yellowish stripes on leaf. It starts at tips and enlarge in size with the wavy margin.

The rice crop affected by rice tungro diseases is shown in Figure. 1.c. The plant affected by this disease exhibits stunting and less tillering. The leaves turn yellow or orange-yellow in color and contain rust colored spots.

The three images depicted in Figure. 2 represent the rice crop affected by diseases during main field stage.

Figure. 2.a indicates the discolouration in the leaf tip that could extend down to the blade which is created by the sheath blight infection.

The irregular spots are lesions with dark reddish brown margin and gray center shown in the leaves of Figure. 2.b, represents the sheath rot diseases.

Another disease that affects the rice plant is the brown spot. It appears as isolated brown, round to oval spots measuring 0.5 to 2.0mm in breadth and increase to form large patches, Figure. 2.c shows the brown spot diseases

METHODOLOGY

Among the various diseases discussed in section 3, blast disease is considered to be major yield cutting disease, as it could infect the rice crop at any stage beginning from nursery to main field. Hence it is proposed to develop methodology based on machine learning techniques to detect the infection at an early stage [11].

In the proposed system, the images of rice crop are captured from the field on a regular basis. These images are processed using suitable Pre-processing, Segmentation and Enhancement techniques. The appropriate features which are required for classification are extracted from the treated images. Out of the collected images, a fraction is used for training the classifier, which is either a K-Nearest Neighbour (KNN) or Artificial Neural Networks (ANN).These classifiers will process the images, based on the training and classifies the image as either

Figure 1 : Nursery stage rice crop diseases

Figure 2: Main stage rice crop diseases

normal or blast infected one. The various steps involved in the proposed system explained in the subsequent section.

Image Acquisition

The images from the paddy field are captured using digital camera with high resolution (512×512) for better accuracy and [15,18] the images are resized to 256×256 pixels for ease of processing as shown in Figure.3.The resized color image is converted to gray scale image for further processing.

Image pre-processing

During the image pre-processing phase, the RGB image is converted to HSV. Extracted saturation and Hue part of the image is shown in Figure 3a and 3b.

Image Segmentation

K-means Clustering is used for Image segmentation after treated the preprocessed images using K-Means Clustering for various K values. From experimental results, been found that K=3 produces more clear and accurate image [12, 13]. A sample result of the K-Means clustering applied to the disease affected images is shown in Figure.4

Feature Extraction

The segmented images are used to extract the features related to the disease infection. Statistical features such as mean and standard deviation are extracted in addition to other features like GLCM, Entropy, Skewness and RMS [14]. The GLCM texture feature indicates the Gray level co-occurrence matrix to calculate the disease affected portion. In the GLCM algorithm, following features were extracted for blast affected and normal portions and their relations are given in Eq. (3) - Eq. (6).

a) Mean Value for Image

The average color value of the image can be calculated by the following equation

$$\text{Mean} = Ei = \sum_{j=1}^{N} Pij \tag{1}$$

b) Standard Deviation

$$\text{S.D.} = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (Pij - Ei)2} \tag{2}$$

c) Energy

$$\text{Energy} = \Sigma_{(i,j=0)}^{(N-1)} (Pij)^2 \tag{3}$$

d) Homogeneity

$$\text{Homogeneity} = \sum_{i,j=0}^{N-1} \frac{Pij}{1 + (i-j)^2} \tag{4}$$
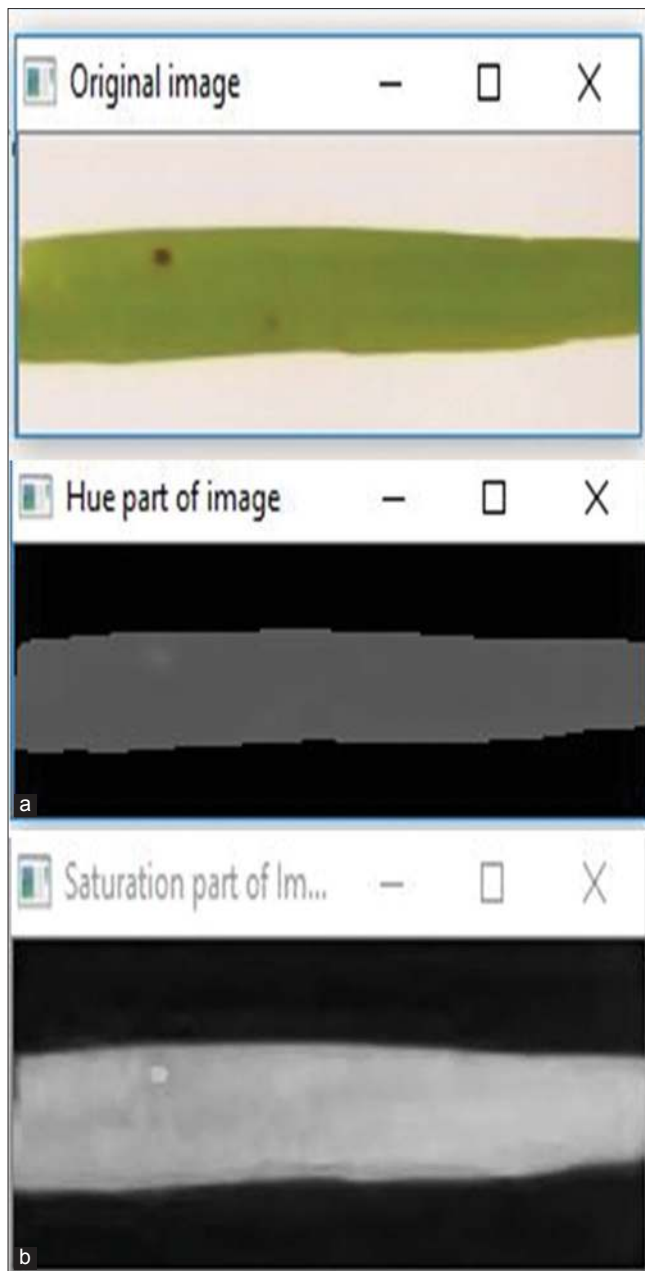
Figure 3: Original image (a) Hue part of image (b) Saturation part of image
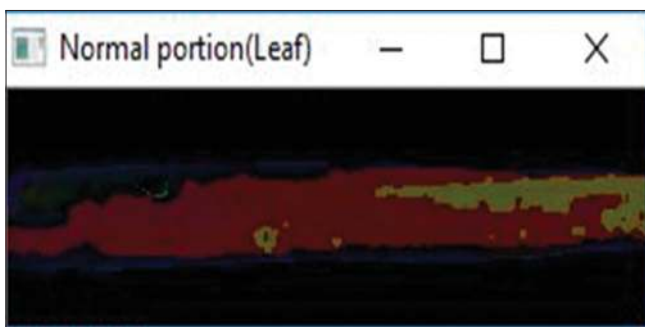

Figure 4: K-means clustering image

e) Contrast

$$\text{Contrast} = \sum_{i,j=0}^{N-1} Pij(i-j)^2 \qquad (5)$$

f) Correlation

$$\text{Correlation} = \sum_{i,j=0}^{N-1} \frac{Pij(i-\mu)(j-\mu)}{\sigma^2} \qquad (6)$$

Classification

The extracted features are applied to a classifier to determine the type of image, whether it is infected or uninfected. Two types of classifiers are used in this research work: KNN and ANN, which are based on Colour and Texture features. [15, 18, 19]. These two classifiers are discussed in the following divisions.

KNN Classifier

K- Nearest Neighbour Classifier (KNN) is a simple classifier that provides better result in identifying the blast disease. The KNN classifier is used over disease affected leaf and normal portion of the leaf images. In this method, it is proposed to take the K value as 3, because the selection of parameter K values is crucial in KNN based classification, as it determines the accuracy of the result. After carrying out multiple simulation experiments, it is inferred that K=1 shows only the background of the leaf part and K=2, shows leaf part alone. When K=3 the image with background and leaf part is seen. Hence this K=3 is selected as the parameter value. This selected K Values provides an accuracy of 85% for blast affected images and 86% for normal images [16].

ANN Classifier

It could be seen from the previous division that the accuracy using KNN Classification still needs improvement and hence it is proposed to use ANN Classifier to improve the accuracy. ANN Classifier uses the standard layer architecture, in the input layer of ANN Classifier, the extracted features are applied. One hidden layers with five neurons is used to get more accurate result. The output layer indicates whether input image is normal or infected one. The ANN works on the basis of initialization of weights, feed forward back propagation and updation of the weights. The ANN Classifier used in this research work resulted in 99% accuracy for normal image and 100% for blast infected images.

Loss function

Loss function is an important part in artificial neural networks, which is used to measure the inconsistency between predicted value ($\hat{y}$) and actual label (y). It is a non-negative value, where the robustness of model increases along with the decrease in the value of loss function. It is also known as error. The learning rates are tested for various values such as 0.2, 0.1, 0.02 and 0.25.

These values are used in a trial and error basis. The number of epochs range from 100 to 1500. The loss function Graph is shown in the Figure 5.

RESULTS

Performance Analysis

The performance measures of the KNN and ANN Classifiers is derived using confusion matrix through the following four outcomes True Positive, True Negative, False Positive and False Negative. [17]

From the confusion matrix various performance factors such as error rate, accuracy, sensitivity, specificity and precision are computed.

The confusion matrix consists of the following four outcomes.
1. True Positive (TP): correct positive prediction
2. False Positive (FP): incorrect positive prediction
3. True Negative (TN): correct negative prediction
4. False Negative (FN): incorrect negative prediction

Table 1 indicates the confusion matrix generated for the data sets used in this research.

Table 2 Shows the Confusion Matrix Performance Measures.

Accuracy

Accuracy can be calculated from the confusion matrix as the number of correct predictions over the total number of images in the dataset. An accuracy value of 1 indicated best performance and zero represents the worst behaviour.

The computational formula to calculate the accuracy in the Eq. (7). The accuracy of the proposed methodology is plotted in Figure. 6. It could be inferred from Figure. 6 that the accuracy of ANN Classifiers is better than that of the KNN Counterpart.
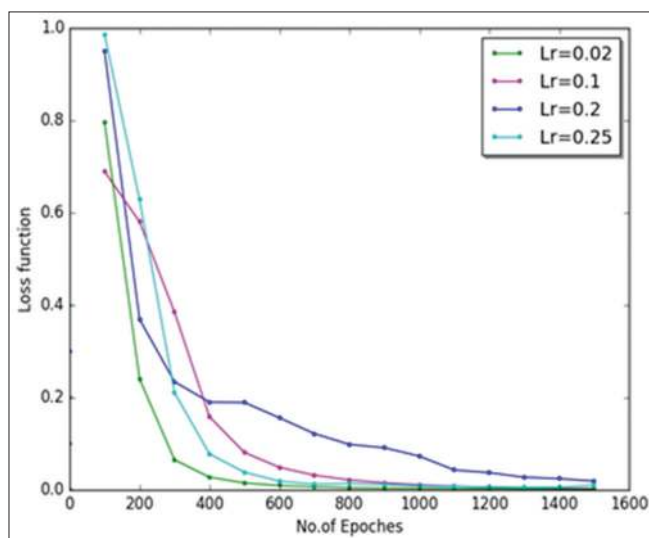

Figure 5: Loss function graph

$$ACC = \frac{TP + TN}{P + N} \qquad (7)$$

Sensitivity (SN)

Sensitivity is also known as True positive rate (TPR) or recall. It is calculated using the number of correct positive predictions over the total number of positive images. The methodology is considered to be more sensitive if the sensitivity value as 1, and insensible if it zero. The sensitivity factor for the two classifiers used in this research work over the set of images is presented in Figure. 6. It is seen that ANN classifier is comparatively more sensitive than KNN.

Specificity (SP)

The specificity is calculated from the confusion matrix as the number of negative predictions over the total number of negative images. This performance measure for the proposed system is presented in Figure. 6 from which it is reinforced that ANN algorithm performs better than the KNN classifier. It is also called as True Negative Rate.

Precision

The precision of the classifier algorithm is computed as the ratio of number of correct predictions over the total positive prediction. It is also referred to as positive predicate value (PPV). The formula to compute PPV is given in Eq. (8). The precision value obtained for the simulated results are shown in the Figure. 7.

$$PREC = \frac{TP}{TP + FP} \qquad (8)$$

Negative Predicated value

The NPV is calculated as given in Eq. (9) and the value from the data set images used in this paper is given in Figure. 7.

$$NPV = \frac{TN}{TN + FN} \qquad (9)$$

Table 1: Confusion matrix

S.No.	Prediction	KNN	ANN
1	True positive	132	146
2	False positive	22	2
3	True negative	182	202
4	False negative	14	0

Table 2: Confusion matrix performance

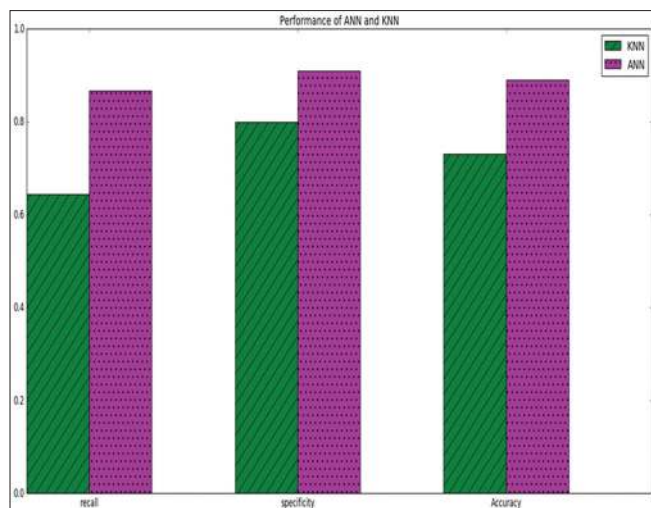| S.No.	Performance measure	KNN (%)	ANN (%)
1	Accuracy	70	90
2	Recall	65	88
3	Specificity	78	90
4	Precision	72	98
5	NPV	72	99
6	F1 Score	65	97

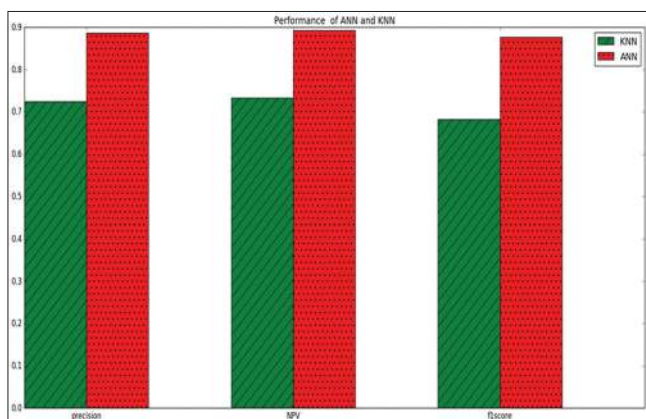**Figure 6:** Performance for recall, specificity and accuracy

Figure 7: Performance for Precision, NPV and f1 score

F1 Score

F1 score is calculated by the formula

$$f \, 1 \, score \, = 2^{*(precision*recall/precision+recall)} \qquad (10)$$

The f1 score performance is shown in Figure. 7.

False Positive Rate

False positive rate is calculated by the number of incorrect positive predictions divided by number of negatives. The False positive rate obtained in this work is shown in Figure 7. 0.0 is considered to be the best false rate and 1.0 as worst. FPR is given in Eq. (11).

$$FPR = 1 - SP \qquad (11)$$

False Discovery Rate

The false discovery rate is calculated by the formula shown in Eq. (12).

$$FDR = \frac{FP}{(TP + FP)} \qquad (12)$$

The plot of FPR and FDR for the two classifiers is shown in Figure. 8.

Overall accuracy analysis

The main objective of this paper is to identify the blast affected part of the rice leaf using Machine Learning Techniques.

From the total datasets 78% of images are used for training the classifier and the balance 22% is utilized to analyse the performance of proposed mechanism.

Table 2 presents the complete details about the training and testing image samples.

Total training images consists of 350 which includes blast affected as 204 images and normal 146 images.The total testing images consists of 101 images which consists of blast affected as 56 and normal as 45 images.

Figure. 9a presents the accuracy of the classification algorithms over the blast affected and normal images. It can be seen from fig. 9a that the blast affected images are 99% accuracy identified as the affected images by ANN Algorithm, where as KNN algorithms produces less accurate results.In the traing phase the KNN and ANN algorithm show the accuracy of 90 and 99% for normal images. Subseuqtly the blast affected shows the result of 99 and 100% which is evident from Figure. 9a.

Figure. 9b. presents the accuracy of the classifier mechanisms over test images is given in fig 9b. It is proven by the test images that ANN is more accurate in identifying the blast infected images. In the traing phase the KNN and ANN algorithm show the accuracy of 63 and 88% for normal images. Subseuqtly the blast affected shows the result of 79 and 90% which is evident from Figure. 9b.

DISCUSSION

Archana Chaudhary et al. [7] proposed the work for multi-class disease classification problem. The performance is improved using random-forest algorithm. The accuracy is increased up to 97.80% for multi class groundnut disease dataset.

Juan Ignacio Arribas et al. [8] proposed the concept to find the sunflower crops using automatic neural networks. The system consists of four stages: first segmentation is based on RGB color space, second many different features are detected and extracted from segmented images, third select the set of features and fourth is neural networks algorithm to differentiate the sunflower and non-sunflower crops. The experimental results show that correct classification rate of 85% and an area under the receiver information curve over 90% for the test set.

The author Vijai Singh et al. [9] proposed the work for automatic plant disease detection using image segmentation techniques.
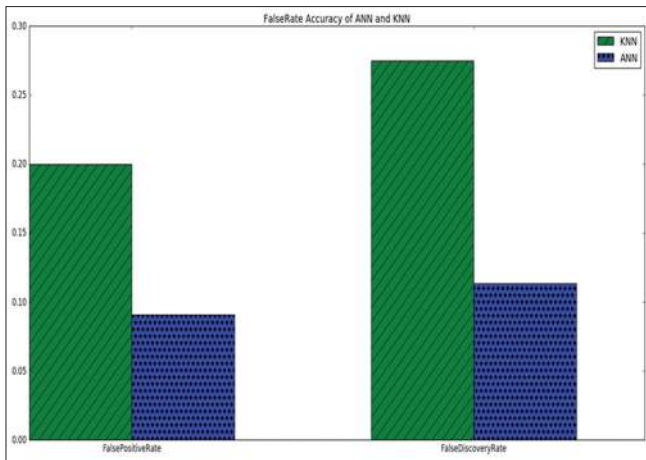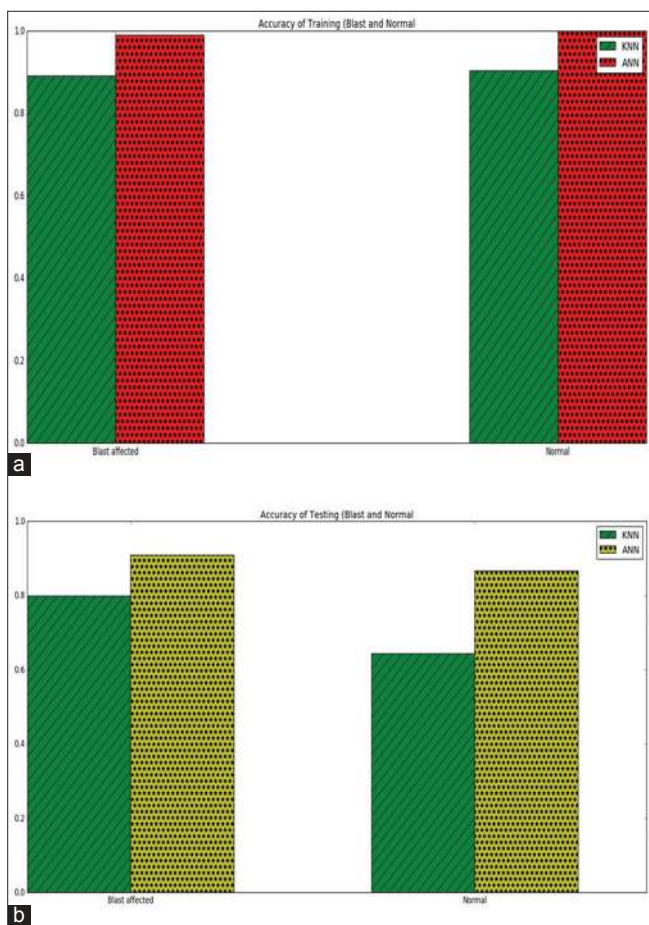
Figure 8: - Performance for FPR and FDR

Figure 9: (a) Training accuracy for KNN and ANN (b) Testing accuracy for KNN and ANN

Several diseases are taken and compared with proposed work. Image segmentation is important aspect of automatic plant disease detection and is carried out using genetic algorithm. When compared with existing algorithms the proposed algorithm obtained better accuracy.

The author Yang Lu et al. [10] proposed the work for automatic plant disease identification using deep convolution neural

Table 3: Complete data set

Total training images-350	Total testing images-101
Blast affected-204	Blast affected-56
Normal-146	Normal -45

networks. The data set consists of 500 healthy and infected leaves, captured from the rice field. Ten common diseases are trained using Convolution Neural Network (CNN) Model. A high accuracy is obtained as 95.48%. The results show that it is feasible for the rice disease detection for the proposed method. These studies outlines that, machine learning algorithm will find the disease in earlier time and farmers can save the crop from diseases.

Conclusion and Future Work

Onsite images were collected from the paddy fields of south India and various features were extracted to train the classifier. The segmentation of images was done using k-means clustering algorithm. Once the classifiers are trained with both blast affected and normal rice leaves they are tested for the performance accuracy. The simulated results were analysed in details in this study and it is inferred that the ANN Classifier performs better in all aspects than the KNN Method. In future the same Machine Learning Methodology could be applied to other rice plants affecting diseases or other crops.

ACKNOWLEDGEMENT

REFERENCES

1. Galhano R, Talbot NJ. The biology of blast: Understanding how Magnaporthe oryzae invades rice plants. Fungal Biology Reviews. 2011;5(1):61-7.
2. Gowda M, Shirke MD, Mahesh HB, Chandarana P, Rajamani A, Chattoo BB. Genome analysis of rice-blast fungus Magnaporthe oryzae field isolates from southern India. Genomics data. 2015;5:284-91.
3. Yan X, Talbot NJ. Investigating the cell biology of plant infection by the rice blast fungus Magnaporthe oryzae. Current Opinion in Microbiology. 2016;34:147-53.
4. Srivastava D, Shamim M, Kumar M, Mishra A, Pandey P, Kumar D, Yadav P, Siddiqui MH, Singh KN. Current status of conventional and molecular interventions for blast resistance in rice. Rice Science. 2017;24(6):299-321.
5. Martin-Urdiroz M, Oses-Ruiz M, Ryder LS, Talbot NJ. Investigating the biology of plant infection by the rice blast fungus Magnaporthe oryzae. Fungal Genetics and Biology. 2016;90:61-8.
6. Agricultural crops: Cereals: Paddyhttp://agritech.tnau.ac.in/crop_protection/crop_prot_crop%20diseases_cereals_paddy.html (Last accessed December 2018).
7. Chaudhary A, Kolhe S, Kamal R. An improved random forest classifier for multi-class classification. Information Processing in Agriculture. 2016;3(4):215-22.
8. Arribas JI, Sánchez-Ferrero GV, Ruiz-Ruiz G, Gómez-Gil J. Leaf classification in sunflower crops by computer vision and neural networks. Computers and Electronics in Agriculture. 2011;78(1):9-18.
9. Singh V, Misra AK. Detection of plant leaf diseases using image segmentation and soft computing techniques. Information processing in Agriculture. 2017;4(1):41-9.

10. Lu Y, Yi S, Zeng N, Liu Y, Zhang Y. Identification of rice diseases using deep convolutional neural networks. Neurocomputing. 2017;267:378-84.
11. Singh V, Misra AK. Detection of plant leaf diseases using image segmentation and soft computing techniques. Information processing in Agriculture. 2017;4(1):41-9.
12. Dyrmann M, Karstoft H, Midtiby HS. Plant species classification using deep convolutional neural network. Bio systems engineering. 2016;151:72-80.
13. Nidhis, A. D., et al. "Cluster Based Paddy Leaf Disease Detection, Classification and Diagnosis in Crop Health Monitoring Unit." Computer Aided Intervention and Diagnostics in Clinical and Medical Images. Springer, Cham, 2019. 281-291.
14. GLCM Texture Feature, https://support.echoview.com/WebHelp/Windows_and_Dialog_Boxes/Dialog_Boxes/Variable_properties_dialog_box/Operator_pages/GLCM_Texture_Features.htm#Energy

15. Lee SH, Chan CS, Mayo SJ, Remagnino P. How deep learning extracts and learns leaf features for plant classification. Pattern Recognition. 2017;71:1-3.
16. Ferentinos KP. Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture. 2018;145:311-8.
17. Basic evaluation measures from the confusion matrix https://classeval.wordpress.com/introduction/basic-evaluation-measures (Last accessed December 2018).
18. Borhan MS, Panigrahi S, Satter MA, Gu H. Evaluation of computer imaging technique for predicting the SPAD readings in potato leaves. Information Processing in Agriculture. 2017 Dec 1; 4(4):275-82.
19. Dhingra G, Kumar V, Joshi HD. A novel computer vision based neutrosophic approach for leaf disease identification and classification. Measurement. 2019;135:782-94.
(Last accessed December 2018).