Disease Prediction in Paddy Crop Using Machine Learning

**Swathi1, Sushma2, Sneha Nayak3, Rakshitha M.4, Deepak Rao M.5**

Department of Computer Science and Engineering,

Shri Madhwa Vadiraja Institute of Technology and Management, Bantakal, Udupi, India.

1 swathi.16cs096@sode-edu.in, 2sushma.16cs095@sode-edu.in, 3sneha.16cs091@sode-edu.in, , 4 rakshitha.16cs070@sode-edu.in, 5deepakraom.cs@sode-edu.in

***Abstract:* India is a country mainly based on agriculture. Paddy is one of the major crops of India. Millions of people in India depend on paddy so as to leave by means of farming and later on processing. As the population of our country goes on increasing the starvation and demand for the food is also getting increasing day by day. In this time of crisis, the leaf disease of crops is the biggest problem in the agriculture industry. These plant diseases will decrease the productivity of plant and directly impact on economy of the nation. Most of farmers in our country are illiterate and there is a lack of knowledge for them about how to overcome with these types of plant diseases. They are unaware of new technologies and advance developments that are taking place in the field of science and technology and there is a lack of some professional specialist in order to bring awareness about preventive measures to avoid disease infecting plants, technology to detect these diseases and actions to be performed after the identification of disease. In this paper, an automated system is proposed for diagnosis of three common paddy leaf diseases (Brown spot, Leaf blast, and Bacterial blight). K-means clustering is used for separating affected part from paddy leaf image. Visual contents (color, texture, and shape) are used as features for classiﬁcation of these diseases. The type of paddy leaf diseases is recognized by Support Vector Machine (SVM) classiﬁer**

**Keywords- Machine Learning, image acquisition, median filtering, K-means clustering, ANN - Artificial Neural Network, Naïve Bayes Classifier, SVM - Support Vector Machine, KNN - K- Nearest Neighbor, Gradient Classifier.**

# I. INTRODUCTION1

Rice is one of the major crop in India and it takes major part in agricultural economy. But now a days these crops are facing many problems such as disease which may results in poor quality and quantity of the crops. The diseases found in the crops results in major loss to the farmers. It will cause grater economic loss to the farmer, apart from the other problems like pest and environment factors.

Leaf blast, brown spot, and leaf sheath blight are some of the diseases that occurs in the rice crop. The early detection of these diseases will prevent the grater loss to the farmers. By their experience, they identify the type of the disease that found on the crop but their predictions may not be correct always. Most of farmers in our country are illiterate and there is a lack of knowledge for them about how to overcome with these types of plant diseases. They are unaware of new technologies and advance developments that are taking place in the field of science and technology and there is a lack of some professional specialist in order to bring awareness about preventive measures to avoid disease infecting plants, technology to detect these diseases and actions to be performed after the identification of disease. Most of the farmers can manually identify and classify the disease but it takes lot of time. The disease that occurs mainly on stem and leaves. The disease on crop may be bacterial, viral, fungal, due to insects, rust etc. Identifying these diseases in their early stage is very important.

These are the steps used to identify diseases by using image processing and machine learning techniques. At first the images are taken from the computer, later are fed into preprocessing step, then by using segmentation, affected regions from leaf are taken, Color, texture, shape etc. features are extracted from the segmented images, and finally using machine learning technique disease is classified. The appropriate features of the images are extracted and the images are classified either as unhealthy or healthy. The procedure is divided into five steps. Firstly, it performs Image Acquisition. Second stage consists of Image Pre-processing stage III has Image segmentation techniques, Stage IV defines the Feature selection and Feature extraction of segmented images. and finally, the classification of the diseases takes place and performs the analysis and conclusion.

The image processing has the following objectives:

* To identify the infected leaf or stem of the plants.
* To quantify infected region by disease.
* To identify the structure or shape of affected region.
* To recognize color of infected region.
* To identify the size and shape of the leaf.

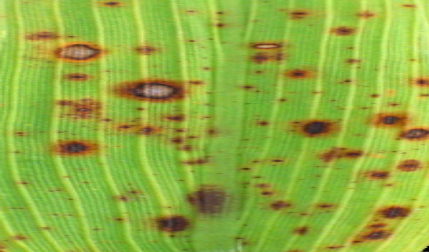
Classification is performed based upon their distinct features. This paper presents different techniques using which detection of rice crop disease can be done. Considering the shortcomings in existing surveys, the paper presents the different classification of crop diseases including preprocessing, segmentation, feature extraction and feature selection, classiﬁcation. By using this techniques, farmers can automatically recognize the leaf diseases at early stages.

# II. DESCRIPTION OF RICE DISEASE

Leaf blast, Brown spot, and Leaf sheath blight are the diseases that are found commonly in paddy leaves. These diseases not only affect leaves but also root, stem and other different plant parts. Identification of these in their early stage will stop greater loss to the farmer. Some people who are experienced by working in the field of agriculture can guess disease but that is not the right way to do even the people who are specialist in identifying disease takes a lot of time to correctly predict. So as to prevent greater loss of crops farmer should take actions as quick as possible.

* + **Brown spot**

Brown Spot disease is caused by Bipolarisoryzae named fungi which belongs to class Deuteromycetes,



**Figure1: Paddy brown spot disease**

***Symptoms:***

* The leaves regions affected by this disease have small, elliptical and brownish colored wounds in the early stages.
* It is the disease which occurs commonly in rice plants and causes a greater loss to farmers.
* Brown spot appear initially small circular to oval spots on seedling leaves.
* Some spots are small which are of dark to reddish brown and big spotsare not dark and are brown with reddish color or may be gray centered covered bymargin of dark brown to red brown and spots that occurred early are yellowhalo surrounding the spot lesion.
* Brown spots symptoms ofleaves also similar symptoms on the leaf sheath and hulls*.*
* **Bacterial leaf blight**



**Figure 2: Paddy bacterial leaf blight disease**

***Symptoms****:*

* Bacterial leaf blight as name suggest it is a bacterial disease which is a commonly occurring disease in paddy.
* Here we can identify the disease affected regions at the tip of leaf or on the lines that are observed on the surface of leaf.
* It is usually having golden stripes on blades of leaves or at the tip.
* The main symptoms are leaves having yellowish to golden colored lesions on the blades of leaves and at the tip, leaves back getting dry more early from the tip, leaves will curl because of lack of nutrients.
* The leaves that are completely prone to this disease are showing their first symptoms by getting dry early.
* **Leaf blast**

Oryzae fungi is the one which cause this disease in paddy crops. It not only affects leaves but also on stem, root and sheath of leaves. Whenever spores are found on the surface of plant in that spot blast disease can occur in any stage of plant growth. Initially disease symptoms are white to gray green lesions or brown border spots on all parts.



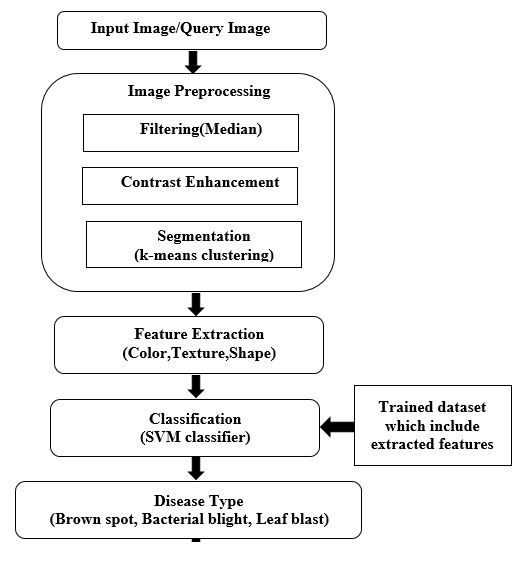
**Figure3: Paddy Leaf blast disease**

***Symptoms:***

* Here we can identify the disease affected regions at the tip of leaf or on the lines that are observed on the surface of leaf.
* It is another paddy disease which is caused mainly due to growth of fungus on the leaves of paddy.
* Here we can see round or spindle shaped wounds on the surface of leaves.
* The lesions are having brown colored border, in the middle it will be white or gray.

# III. PROPOSED METHODOLOGY

The proposed methodology has concentrated on detecting the paddy leaf diseases which helps farmers to take correct measurements and increase the yield.



**Figure 4: workflow of plant disease prediction**

**Algorithm for plant disease prediction**

Step 1: Start

Step 2: Take the infected leaf image from farmer field using a digital camera and then input image to the system.

Step 3: Median filter is used to remove the noise and smooth the image.

Step 4: The contrast of image is enhanced for better understanding of different part of the image.

Step 5: k-means clustering is used to segment the image so that we can obtain the disease affected part of paddy leaf image.

Step 6: Features are extracted such as color, texture and shape. After which a feature vector is created.

Step 7: This feature vector is sent to SVM classifier to recognize the diseases. It classifies the query image to one of the types of paddy leaf disease.

Step 8: Finally the percentage of disease affected area of image is measured and accuracy of the disease is predicted.

Step 9: Stop

**1)Image Acquisition:** Somedisease affected paddy leaf images are collected from the internet and captured by camera. Then these images are preprocessed. The digitized images are about 200-400 KB size .

**2)Preprocessing:** Preprocessing involves resizing the image, filtering image contrasting and image segmentation.

**Image Filtering:** When the images are captured, they may contain dust, sediment, dew drops etc. The median filter is used to remove noise and smooth the images. The median filter is a technique to smooth the images by replacing the value of a pixel by median of the intensity value of the neighborhood of the pixel so that the noise are replaced by the image intensity value. The median is calculated by first sorting all the pixel being considered with the middle pixel value.

**Contrast Enhancement:** The contrast of the image is enhanced for better understanding of different part of the image such as normal object and lesion present in the leaf part.

**Image Segmentation:** Segmentation is an important step while processing an image. Segmentation divides an images into its constituent region or objects. Segmentation will terminate when an object or interested region is encountered in an application. Segmentation of nontrivial images is difficult task. Based on accuracy of segmentation, success or failure of computer analyzed procedures is found. That’s why, correct majors should be preferred to improve the probability of segmentation accuracy. The process of dividing digital image into many segments with similar attributes is segmentation. It is to simplify the representation and more helpful for interpretations and analysis. In simpler terms, Segmentation is the process of dividing something into parts or segments.

In this we use K-means clustering is used to segment the images. The images are segmented into 3 clusters depending upon the color variation. Among the three clusters, the disease affected part is chosen to extract features. K-means is one of the simplest unsupervised learning algorithms that is mainly used to solve the clustering problem. It is mainly used for the segmentation purposes. It clusters or groups the huge amount of pixels.

**Algorithm For K-means Clustering**

Step1: Start

Step2: Read image.

Step 3: Convert the image from RGB Color Space to L\*a\*b\* Color Space.

Step 4: Classify the Colors in a\* b\* Space using K- Means Clustering.

Step 5: Label every pixel in the image using the results from k-means.

Step 6: Create images that segment the image by color.

Step 7: Separate infected part and uninfected part.

Step 8: Stop

**3) Feature Extraction:** In image processing, features plays an important role in dimensionality reduction. When the size of the input data becomes too large but not containing so much information then the input data can be transformed into set of features. In this system, the visual based features i.e. color, texture and shape features are extracted. Here we use 13 features to identify the disease those are Mean, Standard Deviation, Entropy, RMS, Variance, smoothness, Kurtosis, Skewness, IDM, Contrast, Correlation, Energy, Homogeneity.

## 4) Classification: In machine learning, we mainly have three types among which supervised learning is one of the type and classification is coming under this type. Classification is the best one to use when we get the finite and discrete values as the end results. Classification will take the data as inputs and will group those data to the specific group for which they are belongings to. For the input variables provided it predicts the class as well. Here different patterns are processed and later they are grouped together into one class based on the similarities identified with their group members. Different Classification types are used on the features that are extracted from the morphological features that may be color, shape of leaf, stem, root and later are processed by taking the images of disease affected area in order to predict the disease type. We have different classification techniques they are Naïve bayes classification, SVM classification (support vector machine), Decision Tree Classification, KNN classification (K-nearest neighbors), Deep learning based techniques, Artificial neural network (ANN), probabilistic neural network (PNN) and Random forest classification.

**Support vector machine**

For nonlinear classification SVM can be utilized. This is explained mainly for the factors with double classes. However, it is possible to increase technique of support vector machine with factors having many classes using one for all the type. For decision making this can be made use of. SVM works mainly on two phases. The ﬁrst one is oﬄine phase, in which using the cells that are grouped together into different sets that are need to be sprayed or not be based on which training process is presented and performed completely the operations upon the decision. Another phase is online phase, here they presented the operations of making the decisions depending upon the cells that are entering recently, which in turn controlled by the operations of making decisions that are finished throughout the course of oﬄine phase. For the classification and training process there is classifier named MSVM which has a collection of support vector machine.

The benefits of support vector machine are that it only be made use for when we are making classification linearly and non-linearly. In terms of recognition it is easy and provides a perfect outcomes compared to other types. During training if the images are having lot of unwanted noise, support vector machines are best one to use. When it is about drawback, the process of training is slow. The algorithm on which support vector machine works is not easy for to understand. It works only on two classes which is another drawback. The very difficult thing is selecting the parameters and functions for kernel.

**IV. CONCLUSION**

Paddy leaf diseases are the major problem in the agriculture sector all over the world which causes a greater economic loss. In order to prevent this loss a proper system should be built so that it is possible to overcome the loss. This paper is used to develop a system which will be more reliable and gives more accurate prediction of disease. Now a days there is more scope for image processing and machine learning and detection of the plant disease and identifying the type of disease and finding the treatment for these disease is one of the interesting research that is going on in the line of computer science and agriculture. In order to predict accurate result k-means is used for clustering and svm classifier for disease identification. Color, shape, texture features are extracted and are used to identify the diseases. Later on after the prediction the farmers can take the correct measures to overcome the disease.

**REFERENCES**

1. Divya Srivastava, Rajesh Wadhvani and Manasi Gyanchandani “Color Feature Extraction Methods for Content Based Image Retrieval” International Journal of Computational Engineering & Management, Vol. 18 , May 2015
2. Ravindra S. Hegadi , “Image Processing: Research Opportunities and Challenges” National Seminar on Research in Computers, Coimbatore, India
3. S.Padikar J.Sil,and A.K Das “Classification of Rice leaf diseases based on Morphological changes”, International Journal of Information and Electrical Engineering,Vol.2,Num 3,May 2012
4. Basavaraj S.Anami, Surendra palaiah, and Naveen n palavade “classification of yield affecting biotic and abiotic paddy crop stress using field images.
5. Dr. Neha Mangla1, Priyanka B Raj2, Soumya G Hegde3, Pooja R4 "Paddy Leaf Disease Detection using Image processing and Machine Learning” International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control EngineeringVol. 7, Issue 2, February 2019
6. Farhama Tazmim Pinki,Nipa Khatun,S.M. Mohidul Islam “Content based Paddy Leaf Disease Recognition and Remedy Prediction using Support Vector Machine”International conference of computer and information technology ,22-24 December 2017
7. Er. Jasdeep Kaur1 and Dr. Sanjay Singla2 “Detailed review and Classification of Segmented Image for Paddy Plant Disease” International Journal of Latest Trends in Engineering and TechnologyVol.(7)Issue(3), *pp. 419-426*
8. Toran Verma and Sipi Dubey “Paddy Disease Recognition using Image Processing and Radial basisFunction Network” Indian Journal of Science and Technology,Vol 10(46), DOI: 10.17485/ijst/2017/v10i46/116025, December *2017*
9. Shital Bankar, Ajita Dube,Pranali Kadam, Prof. Sunil Deokule “Plant Disease Detection Techniques Using CannyEdge Detection & Color Histogram in Image Processing” International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1165-1168
10. Simranjeet kaur, Geetanjali Babbar, Gagandeep“Image Processing and Classification, A Method for Plant Disease Detecion” International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue-9S, July 2019
11. Darshana A.1, Dr. Jharna Majumdar2, Shilpa Ankalaki3 “Segmentation Method for Automatic Leaf Disease Detection” International Journal of Innovative Research in Computerand Communication EngineeringVol. 3, Issue 7, July 2015.
12. MuhammadAsraf,H.,Nooritawati,M.T.,Rizam,MSBShah,2012.Acomparativestudyin kernel-based support vector machine of oil palm leaves nutrient disease. Procedia Eng. 41, 1353–1359.
13. Sannakki, Sanjeev S., Rajpurohit, Vijay S., Nargund, V.B., Kulkarni, Pallavi, 2013. Diagnosis and classiﬁcation of grape leaf diseases using neural networks. In: 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT). IEEE, pp. 1–5.