**Detection and Classification of Rice Plant disease by using Machine Learning**

**Abstract**

Agriculture is the most important sector of Indian economy. Detection of plant diseases plays an important role in agricultural field using image processing automatic techniques. It reduces the large amount of work and time for monitoring big forms of crops. The Image segmentation and Neural networks are used for automatic detection and classification of plant leaves. It takes the image as input, and apply techniques to that image like pre-processing, segmentation and then the input is given to the Convolution neural network in order to classify the disease.

**Keywords:** Image processing, HOG, Otsu’s threshold , CNN

**1. Introduction**

Rice is a prime food for majority of population in the world. The mass production of rice is increased quantity and quality then it will generate additional income sources. But in today’s environment the farmer is suffering various types of problems in growing the crops. The farmer does not suffer from environmental problems only he suffer a great loss in agriculture due to various kind of plant leaf disease that effect the relent of crops. Rice crops suffer from infection of many bacterial, viral and fungal diseases. It decreases the annual production of rice in high society. Rather huge production we are still fall back behind, because existing compositions does not obtain any comprehensive technique which can deal with the complete identification of the plant disease. Plant disease has been a major factor influencing food production and human societal development over thousands of years and for major economic losses in the agricultural industry worldwide. Detecting plant health and pathogen early are essential to reduce disease spread and easy effective management practices. We should develop a framework for disease identification . When some diseases are not visible to naked eye but actually they are present, then it is difficult to detect it with the naked eye. And when it is visible it will be too late to detect disease and can’t help anymore. Earlier, microscope is used to detect the disease, but it become difficult as to observe each and every leaf and plant. So, the fast and effective way is a remote sensing technique. Automatic detection of plant diseases is an important research topic these days as it may prove benefits as automatically detect the diseases from the symptoms that appear on the plant leaves. Detection and recognition of diseases in plants using machine learning is very fruitful in providing symptoms of identifying diseases at its earliest.

Some of the different types of rice plant diseases:

**1.1 Bacterial Blight of Rice**

**Pathogen:** Xanthomonas Oryzae

**Lesion Colour:** Yellow to white due to the effect of bacteria

**Symptoms :**

* These lesions coalesce and become yellowish-white with wavy edges.
* Gradually whole leaf affected and then the drying & twisting of leaf tip.
* Streaks coalesce filling vascular bundle with bacteria.
* In older plants, the leaves become yellow and then die.
* Margin blight- symptoms appear on the leaves of young plant as pale-green to grey-green, water- soaked streaks near the leaf tip and margin.



Figure1:Example of Bacterial blight of rice

**1.2Leaf smut of rice:**

**Pathogen:** Entyloma Oryzae

**Lesion Colour:** Reddish brown

**Symptoms :**

* The fungus produces slightly raised angular black spots on both the side of leaves.
* It also produces on leaf sheaths.
* Many spots remain distinct from each other.
* Epidermis breaks open when wet and releases the black spores.
* Heavily infected leaves turn yellow and leaf tips die and turn gray.



Figure2:Example of leaf smut of rice

**1.3Brown Spot:**

**Pathogen Name:**Helminthosporium oryzae

**Lesion Colour:** Reddish brown to dark brown

**Symptoms:**

* Start as dark brown,irregular spots on both upper and lower leaf surfaces.
* Seedlings have yellow brown lesions at the starting stage which are dark brown in color.
* Later became purple brown. Fully developed lesions are surrounded by reddish brown margin
* It can also develop on stem and pods of plants approaching maturity. stem and pods lesions have indefinite margins, are dark in appearance and range in size from flecks to larger areas.



Figure3:Example of Brown spot

**2 Related Work:**

**2.1 [1]**This article presents a prototype system for detection and classiﬁcation of rice diseases based on the images of infected rice plants. Here they propose centroid feeding based K-means clustering for segmentation of disease portion and SupportVectorMachine(SVM) for multi-class classiﬁcation . They achieve 93.33% accuracy on training dataset and 73.33% accuracy on the test dataset.

**2.2[2]**This paper presents an overview on various types of plant diseases and different classification techniques in machine learning that are used for identifying diseases in different plant leaves . The intended to focus on increasing the recognition rate and classification accuracy of severity of leaf diseases by using hybrid algorithms.

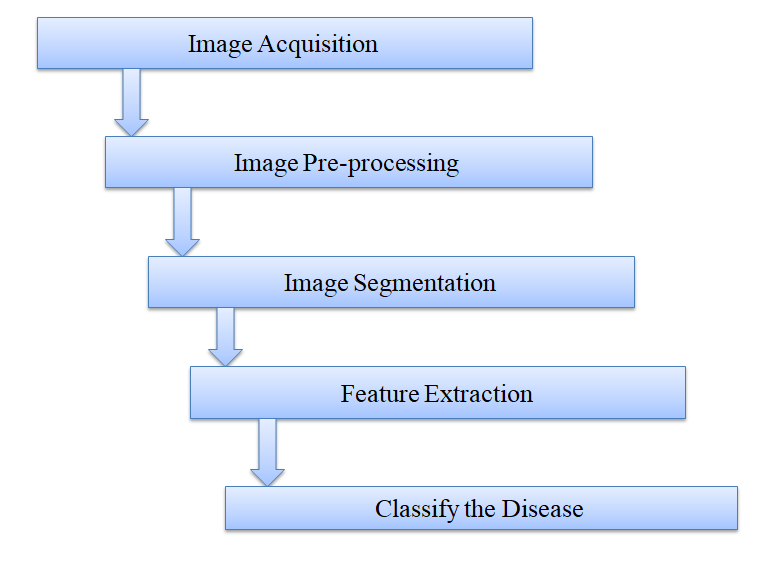
**2.3[3]** The proposed system identifies diseases of rice plant leaf by extracting features from the infected regions of the rice plant leaf images. Fermi energy based segmentation method used to segment the infected region from its background region. Symptoms of the diseases are characterized using features like color and shape of the infected portion and extracted feature used for identifying diseases.

**2.4[4]**The paper presents a comprehensive view on the various researches done in contemporary domain of Crop diseases. The collection of analysis is done dealing with various plant diseases. Then the identification of exact disease in rice such as Blast, Bacterial Leaf Blight, Brown spot, Sheath Blight and False smut are to be achieved through the image processing.

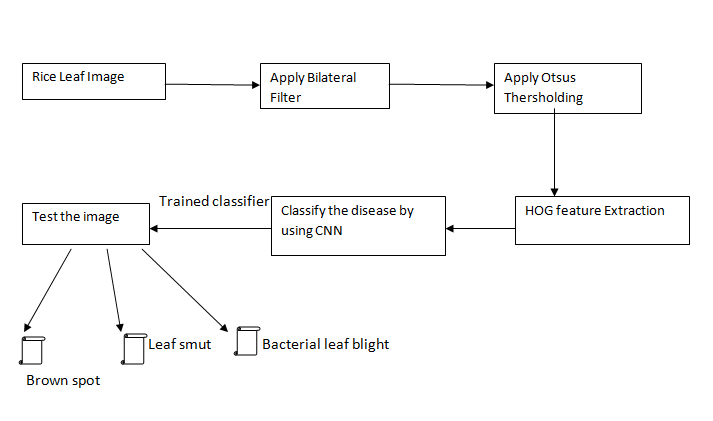
**2.5[5]**This paper presents a survey of different image processing and machine-learning techniques used in the identification of rice plant diseases based on images of disease infected rice plants. This paper presents not only survey of various techniques but also concisely discusses important concepts of image processing and machine learning applied to plant disease detection and classification.

**3 Proposed Work**

The basic steps required for image processing deals with image acquisition, image preprocessing, disease segmentation, feature extraction and disease classification[1][4]



**Figure 4: Block diagram**



**Figure 5: Proposed block diagram**

**3.1 Image Acquisition :**

Image acquisition is the start of any image processing technique, in which images are digitalized and stored. For this system, leaf sample images with visible disease spots can be either captured with a digital color camera with a uniform background on site, or can be retrieved from any online database. [6] It is recommended that images retrieved from the internet have a uniform background. This will allow for more accurate segmentation results as the background area can be easily distinguished from the leaf area. The resolution of images retrieved from an external database does not need to be set to a specific value, but this will be amended in the pre-processing stage .The real time images are fed directly from the camera. For further analysis, proper visibility and easy analysis of images, white background is created because most of leaves colour varies from red to green for exact segmentation.

**3.2 Image preprocessing:**

In this method, the image is converted into a standard binary format from other format like color or grayscale. This preprocessed binary image is then subjected to identification wherein the vital parameters of the leaf are extracted for its comparison. In this stage images are resized to a fixed resolution to reduce the computational burden, and any other image adjustments can be achieved in this stage such as cropping, contrast enhancement, and angle correction.

**3.2.1Bilateral filter**

Bilateral filter is a technique can be traced 1995 by Aurich and Weule in non-linear guassian filters.Later proposed by Tomasi and Manduchi in 1998 combines domain and range filtering of smooth images while preserving edges.The filtering specializations such as texture editing,tone management,stylization,relighting,denosing,demosaicking and optical flow estimation.

Image smoothing with Gaussian convolution:

Gaussian convolution : GC [I] p = q∈S Gσ(||p − q||) Iq ……eq(1)

Edge preserving filtering with the bilateral filter:

The idea of bilateral filter[BF] is pixel to influence another pixel and its not occupy by a nearby location but also have similar value.

BF[I] p = 1 /Wp q∈S Gσs (||p − q||) Gσr (|Ip − Iq|) Iq …….eq(2)

where normalization factor Wp ensures pixel weights sum to

Wp = q∈S Gσs (||p − q||) Gσr (|Ip − Iq|) ……..eq(3)

Terminology:

where

σ is a parameter defining the neighborhood size

p and q is a pixels of an image

Gσs is a spatial Gaussian weighting

Gσr is a range Gaussian

||p − q|| is the Euclidean distance between pixel locations p and q

Ip and Iq are intensity of p and q pixels

Gσ(||p − q||) is normalized Gaussian function

σs is a spatial extent kernel

**3.3 Image Segmentation:**

Image segmentation is the process in which the digital image is partitioned into constituent regions, so that the different regions can be easily distinguished and analyzed. Segmentation can be achieved by various techniques such as clustering methods, compression-based methods, and histogram-based methods. The first step is to accurately segment the region of interest from the background. In this case the leaf area should be recognized and distinguished from the uniform background. Different thresholding techniques will be tested and compared to select the most appropriate for this system.

**3.3.1 Otsu Thresholding**

[7] Threshold segmentation is achieved by setting a proper threshold value to distinguish the foreground from the background. The greyscale image is then converted into a binary image according to the threshold set value. The output binary image replaces all pixels in the input image with effulgence greater than the threshold value with the value 1 (white) and replaces all other pixels with the value 0 (black). The threshold is set as a value between 0 and 255 where values closer to 0 signify a threshold value closer to lower grayscale values (black) and vice versa. Otsu's thresholding method uses histogram information to iterate through all possible range of threshold values and selects the threshold that:

* Minimizes the within-class variance
* Maximizes the between-class variance

**3.4 Feature extraction:**

[3]Color and texture information is often enough to make this promience, as against to more general applications, where innumerable shape features must be attain. More recently, several groups have detain the problem of automatic leaf classification.

**3.4.1 Histogram of oriented gradients**

Histogram of oriented gradients (or) HOG is a feature descriptor i.e., used to extract features from image and mainly focuses on the structure or shape of an object .It is widely used in computers vision tasks for object detection .

It decomposes an image into small squared cells computes an histogram of oriented gradients in each cell. The normalizes using a block wise pattern and return a descriptor for each cell. It is mainly able to provide the edge direction as well.HOG is done by extracting the gradient and orientation (magnitude and direction) of the edges. The “localized portions” is called orientation it means completely image is splitted into smaller regions and for each region of gradients and orientation are calculated .Finally, histograms for each of these regions separately they created using the gradients and orientations of the pixel values. Hence the Histograms of oriented gradients.

### Step 1: Preprocess the Data (64 x 128)

### Step 2: Calculating Gradients (direction x and y)

Step 3: Calculate the Magnitude and Orientation

Total Gradient Magnitude =  √[(Gx)2+(Gy)2]

 the value of the angle would be:

Φ = atan(Gy / Gx)

**3.5 Classification:**

[6] Machine learning techniquescompared to statistical models, machine learning me­thods focus on data themselves and foreground the perfor­mance of certain tasks. Machine learning can be applied to four areas based on the problems to be solved: 1) iden­tification/detection 2) classification; 3) quantification; 4) prediction. Machine learning is divided into two catego­ries: 1) supervised learning and 2) unsupervised learning. In the recent decade, machine learning has been used in various disciplines, such as computing, bioinformatics, marketing, medical diagnosis, game playing, etc. Some of machine learning algorithms are frequently used in researches, such as Naïve Bayes Classifier, K Means clustering, support vector machine (SVM), artificial neu­ral networks (ANN), convolution neural networks ,decision trees and random forests. A typical process of employing machine learning includes data collection, dataset preparation, feature extraction, preprocessing, feature selection, choosing and applying machine learning algorithms and performance evaluation. Up to now, machine learning methods are mainly applied to molecular biology and agriculture related to plant di­seases.

**3.5.1 Convolutional Neural Network**

Convolutional neural network is a feed forward neural network i.e., generally used to analyze visual images by processing data.CNN is also known as ConvNet. A CNN uses a system much like a multilayer perceptron that has been designed for reduced processing requirements.In CNN every image is represented in the form of arrays in pixel values.The input image is scanned multiple times to generate the input feature map.Pooling layers connected one after another that carry out feature extraction.The rectified feature map now goes through a pooling layer.pooling is a down sampling operation that reduces the dimensions of the feature map pooling layer uses different filters to identify different parts of the image like edges,corners,body..etc., Pooling is a Subsampling pixels will not change the object ,We can subsample the pixels to make image smaller fewer parameters to characterize the image .

A CNN compresses a fully connected network in two ways:

* Reducing number of connections
* Shared weights on the edges
* Max pooling further reduces the complexity

**4 Result analysis**

**4.1 Image Preprocessing:** In figure 6 the image is converted into applying the bilateral filter image it shows the figure 7.

** **

Figure6 : Original Image Figure7: Apply Bilateral filter

**4.2 Image Segmentation:** After applying the bilateral filtered image its converted into otsu’s segmented image.

****

Figure 8 :Otsus segmented image

**4.3 Feature Extraction:** In segmented image we can obtain the features by using HOG .

****

Figure 9:Obtain features by using HOG

**4.4 Classification:** Based on the features it can extract to classify the disease name.

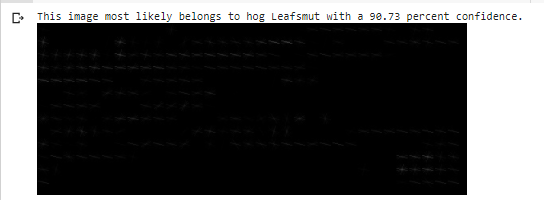
****

Figure 10:Classify the disease name

|  |  |  |
| --- | --- | --- |
| **Image Number** | **Disease Name** | **Confidence of detection** |
| 01 | Blight | 99.81 |
| 02 | Leaf smut | 86.81 |
| 03 | Leaf smut | 100.00 |
| 04 | Blight | 100.00 |
| 05 | Brown spot | 99.29 |
| 06 | Brown spot | 96.23 |
| 07 | Brown spot | 98.34 |
| 08 | Leaf smut | 89.36 |
| 09 | Leaf smut | 100.00 |
| 10 | Blight | 99.68 |
| 11 | Blight | 100.00 |
| 12 | Leaf smut | 99.99 |
| 13 | Blight | 88.24 |
| 14 | Brown spot | 98.32 |

**Table 1:** Detection and classification of disease name and confidence percentage

**5 Conclusion**

This paper reviews and summaries various techniques used or classifying and detecting various plant leaf diseases.The classification techniques helps in automating the detection of plant leaf diseases. Rice plant diseases can sustain huge misfortune in farming if enough notice is not given. Using computer and conveying technologies, an automated system can be built which can deliver early informing of disease. A system for identifying the Rice diseases like Brown spot, Bacterial blight, Leaf smut are detected. It is mainly based on the classification using CNN algorithm.The techniques in image processing for detecting, diagnosing, recognizing of leaf diseases HOG algorithm procedure detection is used for obtain the features and CNN classify automatically the disease name.

**References**

1. Harshadkumar B. Prajapati∗, Jitesh P. Shah and Vipul K. Dabhi , Detection and classiﬁcation of rice plant diseases Intelligent Decision Technologies 11 (2017) 357–373
2. L. Sherly Puspha Annabel, Member, IEEE, T. Annapoorani and P. Deepalakshmi , “Machine Learning for Plant Leaf Disease Detection and Classification” – A Review April 4-6, 2019
3. Dr.C.Kumar Chaliepaul “Classification of Rice Plant Leaf Disease Using Feature Matching”

Volume I,Issue VII,November-2014

1. Shashank Chaudhary, Upendra Kumar, Abhishek Pandey , “A Review: Crop Plant Disease Detection Using Image Processing” ISSN: 2278-3075, Volume-8 Issue-7 May, 2019
2. Jitesh P. Shah , Harshadkumar B. Prajapati , Vipul K. Dabhi , “A Survey on Detection and Classification of Rice Plant Diseases” 978-1-5090-1936-6/16/$31.00 ©2016 IEEE
3. Xin Yang1\*, Tingwei Guo2 , “Machine learning in plant disease research” ISSN :2428-5544
4. G. Prem Rishi Kranth M.Hema Lalitha ,Laharika Basava , and Anjali Mathur, Plant “Disease Prediction using Machine Learning Algorithms” Volume 182 – No. 25,November- 2018
5. Anuradha Badage , “Crop Disease Detection using Machine Learning” : Indian Agriculture Volume: 05 Issue: 09 | Sep 2018 ISSN: 2395-0056 ISSN: 2395-0072
6. Dr. Neha Mangla1, Priyanka B Raj2, Soumya G Hegde3, Pooja R4 , Paddy Leaf “Disease Detection Using Image Processing and Machine Learning” Vol. 7, Issue 2, February 2019