Early Identification of Rice Plant Diseases using Machine Learning Algorithms

1) Lakshmana Rao Rowthu
   Research Scholar
   Department of Computer Science and Engineering
   Centurion University of Technology and management
   Andhra Pradesh, India
   laxman.rowthu@cutmap.ac.in

2) M Naveen Kumar
   Research Scholar
   Department of Computer Science and Engineering
   Centurion University of Technology and management
   Andhra Pradesh, India
   mnaveenkumar@cutmap.ac.in

3) Ch Narayana Rao
   Research Scholar
   Department of Computer Science and Engineering
   Centurion University of Technology and management
   Andhra Pradesh, India
   chnarayanarao@cutmap.ac.in

ABSTRACT

As a precaution and to achieve better quality of crops early detection of rice plant is mandatory. To do this, first disease on the rice plant leaves is identified which helps the farmers to identify type of disease. Different types of diseases may infect rice plants which decrease the yield. Most common diseases are which infect the rice is Rice Blast, leaf smut, bacterial leaf blight and brown spot. Here we have applied machine learning algorithms to detect most common diseases. Different machine learning algorithms are used to train the dataset. First, The data is preprocessed after this step we use different machine learning algorithms which includes Decision Tree algorithm-J48, K-Nearest Neighbor, Naive Bayes. Among these algorithms J48 Decision Tree Algorithm achieved 96% of accuracy. To increase the accuracy CNN Classifier is applied on the dataset. Using this Approach the accuracy level is increased up to 97.58%.

Keywords: Rice plant disease, CNN, Machine learning, Decision Tree, Naive Bayes.

I. INTRODUCTION

Some of the south Asian countries consumed staple food is Rice and it is most staple food for people of Indian country. Rice is most cultivated crop Indian agricultural system. Rice plant Crop diseases due to Environmental problems and floods suffers the farmers in every season. Some of the farmers are illiterate and lack of knowledge on crop diseases. Most of the farmers identify the diseases manually and consult the expertise people. The Manual detection takes more time and some time it gives wrong results. The lack of awareness about diseases it effects yield production. The early detection system is need for detect and identify the rice crop diseases early. There are different types of rice plant disease which include sheath blight, bacterial blight, rice blast, sheath rot, brown spot, narrow brown spot etc. These diseases show impact on rice plants yield production.

The most basic issue for the board of plant ailments is acquiring a right finding. Factors impact illness improvement in plants including different genetics, age of the plant at the hour of infection, condition (e.g., soil, atmosphere), climate (e.g., temperature, downpour, wind, hail, and so on.), single versus blended diseases, and hereditary qualities of the microorganism populaces. Because of this factors, analysis of plant diseases can be troublesome at the beginning phases of infection on plants. Be that as it may, for some illnesses indications do get symptomatic at some phase of infection advancement and a sensible degree of
certainty can be set in analyze dependent on these side effects. Most of the occurred diseases are

1. **Pests:**
   - a) Insects:
     1) Leafhoppers & plant hoppers,
     2) Mole cricket, 3) Rice bug, 4) Rice case worm,
     5) Rice gall midge, 6) Stem borers

2. **Deficiency:**
   - a) Zinc, b) Iron, c) Potassium

3. **Diseases**
   - a) Bacterial
     1. Bacterial leaf streak, 2. Leaf scald
   - b) Fungal
     4. Narrow leaf spot, 5. Rice blast,
     6. Sheath blight, 7. Stem rot

In this work, we propose an early identification of rice plant disease by image processing techniques and machine learning algorithms.

**II. RELATED WORKS**

Different works that have been already done by researchers in classification of plants, fruits disease detection and classification.

Study by [1] a rice disease image database is established for the fast detection of rice blast, bacterial blight, and blight, with 2DFM-AMMF noise reduction and Faster 2D-Otsu segmentation used. The final average accuracy rate is 97.2%. Study by [2] classifies and quantifies the desired two types of diseases from the rice crop and satisfactory results were obtained using combination of K-means clustering algorithm, SVM classifier and CNN based algorithms.

Study by [4] proposed a suitable frame work where enhancement, filter, color segmentation and color feature for classification steps were incorporated for identification. CNN classifier was applied to increase the identified accuracy rate. Study by [5] worked with three different rice plant leaves diseases and proposed an image processing approach where a number of steps including segmentation and classification were incorporated. However; only a small sample of 50 images were tested and accuracy rate was under 75%

Study by [6] proposed system detects the disease accurately for the rice crop. The leaf area was captured from a distance of 25cm of the visible region. The leaf was captured for both normal and diseases part and applied the machine learning algorithm to detect crop diseases.

Study by [7] using image growing, image segmentation techniques to detect infected parts of the plants. Then the infected part of the leaf has been used for the classification purpose using neural network. The zooming algorithm extracts features of the images using simple computationally efficient technique, which results satisfactory classification for image.

Study by [8] discussed the management of biological features on plant in terms of image growth analysis and regular height measurement for huge field by using image masking technique and contour based object detection technique.

Study by [9] worked with three different rice plant leaves diseases and proposed an image processing approach where a number of steps including segmentation and classification were incorporated. The Back propagation Neural Network was used in this project to enhance the accuracy and performance of the image processing.

Different types of classification approaches including neural network, SVM [10] were incorporated as well. Other studies by [11, 12] proposed an image analysis approach where threshold based segmentation and classification for identification. Therefore, it is a need to utilize a image processing techniques to detect the rice plant
diseases. This study proposed an early detection system for rice plant disease detection which will help farmers to convenient analysis the rice plants, to take the fitting measures to expand the nature of harvests and diminish the advancement cost.

III. PROPOSED METHOD

The design of the proposed framework includes obtainment, preprocessing operation, segmentation and identification. It illustrate in Figure 1.

1. **Image Obtainment**: image dataset was collected from real fields and online data sources. The data set contains 3 types of rice plant disease images around 120 images, each type 40 images and set each image into 228x228 size. A sample images are

2. **Image preprocessing**: To get better results reduce the noise from the images. Noise includes blur image, image background, insect excrements, image size and etc.. For this apply image processing techniques to reduce noise and increase the image quality. To reduce color issues of rice plant disease image dataset, apply enhancement operation as convert RGB images were converted to HSV color model for better capability and then apply the median filter and local histogram equalization. To eliminate noise and resize the shape of the regions, hat filter operation and median filter [3 x 3] kernel applied to hold the region edges

3. **Image Segmentation**: It is the process of divide the image into segments of objects. the main object of this process is extract features from the image data to analyze the data. these features are shape, color, Texture. Otsu’s method is applied based on threshold for binary segmentation. It is quicker method as it takes less time. Then , apply techniques to extract features.

4. **Image classification**: In the plant disease identification and classification, the diseases are classified according to the features extracted from the images. Machine learning algorithms were applied on Rice Leaf Disease Dataset to detect three diseases of rice leaf. In this work, 3 classification algorithms and CNN classifier were applied to detect the diseases. They are Decision Tree algorithm-J48, K-Nearest Neighbor, Naive Bayes.

1. **Decision Tree**: Decision tree is one of the most commonly used machine learning classifiers. Taking the best suitable attribute at the root, this algorithm breaks the dataset into partitions. The goal of the partition is to unmix the dataset. The splitting iterates until eventually the partitions group the data such that they are homogeneous. Iterative dichotomiser 3 (ID3), which uses a greedy approach, is the core algorithm for decision tree. In this approach, entropy and information gain, concepts borrowed from information theory, are used for constructing the tree. Entropy measures the impurity of arbitrary attributes; zero entropy means all instances belong to the same class. As entropy becomes more and more positive, the instances become more and more heterogeneous. Using selected features decision tree algorithm was able to correctly classify 94.9074% data on training set where 10-fold cross validation is performed. The model achieved 96% accuracy on test data.
2. K-Nearest Neighbor: KNN works well for discrete target classes. It calculates the distances of the query point from each of the instances and finds the K minimum distances that is, it determines the K nearest neighbors for the query point from which it can predict the class of the query point. The value of K needs to be chosen by inspecting the data; in case, we found when K = 1 the accuracy is 98.8426% on training set and 91.6667% on testing set after performing 10-fold cross validation. And when K=3 the accuracy is 85.6481% on training set and 72.9167% on testing set after performing 10fold cross validation. We found, if the value of K is increased then accuracy is decreased.

3. Naive Bayes Classifier: Naive Bayes algorithm is a probabilistic algorithm that is based on Baye’s theorem. Based on this theorem, the best hypothesis \( \hat{y} \) is chosen based on equation

\[
\hat{y} = \arg\max_{y} P(y) \prod_{i=1}^{n} P(x_i|y)
\]

In this work Naive Bayes algorithm achieved the lowest accuracy to correctly classify three diseases.

4. CNN Classifier:
To escalate the accuracy rate CNN was integrated our test dataset the accuracy rate found only 97.58%.. in this study but it helps to increase the accuracy rate.

IV. EXPERIMENTAL RESULTS:
In this experiment we use machine learning algorithms and CNN classifier. Among these CNN classifier gives higher accuracy rate 97.58%. The following table shows the results of algorithms.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Tree</td>
<td>96%</td>
</tr>
<tr>
<td>K-means Nearest Neighbor</td>
<td>85.12%</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>73.12%</td>
</tr>
<tr>
<td>CNN Classifier</td>
<td>97.58%</td>
</tr>
</tbody>
</table>

V. CONCLUSION:
In this paper, we follow different machine learning algorithms which are KNN, Decision tree, Naive Bayes and CNN classifier. the differentiation between the results of the these algorithms are found. Among these CNN Classifier achieves higher accuracy rate with 97.58%. In this work, we use small dataset which includes 120 images of various rice plant diseases. For our future work, we plan to add more no of images to detect diseases of rice plant early for effective system.

REFERENCES

2. Guoxiong Zhou, Wenzhuo Zhang, Aibin Chen, Mingfang He, “Rapid Detection of Rice Disease Based on FCM-KM and Faster R-CNN Fusion” on IEEE journal, volume xx, 2019


