WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Identification of Paddy Leaf Diseases Using Machine Learning Approaches

P. Prasanth[1], D. Saraswathi[1], K. Rajesh[1], G. Sai Pavan Kumar[1], Dr. Harihara Santosh Dadi[2].

[1]Student, Department of Electronics and Communication Engineering, AITAM College, Tekkali, Srikakulam, Andhra Pradesh, India, 532201.
[2]Professor, Department of Electronics and Communication Engineering, AITAM College, Tekkali, Srikakulam, Andhra Pradesh, India, 532001.

ABSTRACT

In India, 70% of the Economy depends upon agriculture. Out of 70%, rice contributes more than 40% of the country's total food grain production. The weather conditions and leaf disease will affect the yield of the paddy crop. In the early stage leaf disease identification system improves the quality and yield of the paddy. Machine Learning approaches are used to identify the leaf disease in the early stage. The proposed machine learning approach includes image augmentation techniques and is followed by Machine Learning algorithms.

The used Machine learning approaches are SVM(Support Vector Machine), KNN(K- nearest neighbour), DT(Decision tree), Naïve Bayes and Random forest.

So we brought an innovative project for the welfare of farmers and a healthy society.

Keywords: *Augmentation, Machine Learning Algorithms(SVM, KNN, Naïve Bayes, Random forest, Decision tree).*

1.INTRODUCTION:

In India maximum of the economy depends on agriculture itself, India is the second largest producer of wheat and rice crops. India's Agriculture consists of various crops like wheat, rice, vegetables, pulses, fruits etc.. The growth of the crops needed more advanced technologies similar in a proper manner how a person can lead to getting a healthy life. The major drawback is that the farmers are not getting proper quality seeds for the growth of the crops, and due to that reason, the crops get damaged. The farmers were unable to identify the damage to the crops which affects the growth of the crops and also damages human health. Even the farmers get identified in the beginning stage of the damage to the field, but they cannot identify the disease of the crop. Finding the disease and providing the best pesticides needs an expert opinion. In order to control the disease leaf disease plays a major role. This is why we are going to find the disease of the leaf of a particular crop. Leaf production loses its quality by the means of various types of diseases and lots of the diseases cannot be identified with our naked eye.

We are considering a paddy leaf disease for better growth of rice production. This project proposes a method that solves the issue and helps in identifying and classifying leaf disease by applying various techniques. Machine Learning is one of the best-proposed systems for the paddy crop. There are various types of machine learning techniques we will be using like

1. SVM(support vector machine)
2. KNN(K-nearest neighbour)
3. DT(decision tree)
4. Naïve Bayes
5. Random forest

Paddy is frequently affected by Bacterial Blight, Brown Spot, and Leaf Smut thus leading to a decrease in paddy crop production. Recent searches are saying around 37% of annual rice production tends to become decreased due to these rice plant diseases. These techniques can be performed only on the external appearance of the infected plants. Our contribution to this work includes as follows: From the farm field in real-world circumstances we captured the rice plant leaves images and prepared the dataset.

2. PROBLEM DEFINITION

For detecting leaf diseases, the conventional methods are human vision-based approaches. The case finding expert advice is time-consuming and very expensive and also human vision cannot gives a lot of drawbacks. The accuracy and precision of the human vision approach are dependent on the eyesight

of the person or the expert to identify the types of disease in an efficient manner and very soon. We use various types of algorithms in finding this type of disease. Very few recent developments were recorded in the field of paddy(plant) leaf disease detection using a machine learning approach and that too for the paddy leaf disease detection and classification is the rarest thing. However, the productivity of paddy crops is often hampered by various leaf diseases that can cause significant yield losses. Early detection of these diseases are critical for effective disease management.

Therefore, to overcome the drawbacks of the conventional methods there is a need for a new ML-based classification approach. Very few recent developments were recorded in the field of plant leaf disease detection using a machine learning approach and that too for the paddy leaf disease detection and classification is a rare thing. To face this problem we use various Algorithms and find out the best and most effective results that will be considered for the prediction of the leaf disease on the Machine Learning approach.

3. Types of paddy leaf Diseases and data sets

Bacterial Blight:

Rice bacterial blight is one of the deadly bacterial diseases, which is the most destructive affliction of cultivated rice. It is also known as the bacterial blight of rice. The bacterial leaf blight of rice is caused by a species of Xanthomonas. Almost 75% of the crops would be destroyed in severe epidemics and millions of hectares of rice are infected manually.

Bacterial leaf blight of rice is most commonly observed in rice-growing regions of Asia, the Western coast of Africa, Australia, Latin America and the Caribbean regions. In places like the United States, bacterial blight is not found but a bacterial strain related to Xoo has been listed as a select agent by the US Department of Agriculture. It is a designation that places it under strict regulations. The common name of the rice bacterial leaf is also known as rice leaf blight.

Effect of Bacterial Blight

Symptoms of bacterial blight of rice are caused by Xanthomonas oryzae, which can survive on Grass weeds or the stubbles of infected plants. These pathogens are spread through the wind, a Splash of rain, or irrigational water. This disease spreads more than 70 per cent and is warm with Temperatures ranging from 25 to 34 degrees Celsius. More nitrogen fertilizers in the field favour. The diseases, particularly in susceptible varieties. Rice blight disease is found in both tropical and temperature environments, particularly in lowland rainfed areas.

Fig 3.1 Bacterial Blight

Brown Spot

Brown Spot is called a sesame leaf spot or Helminthosporiose or fungal blight The fungus attack. The crop from seeding in the nursery to the milk stage in the main field. The disease appears first as minute Brown dots, later becoming cylindrical or oval to circular. (Resemble sesame seed). Spots measure 0.5 to 2.0mm in breadth-coalesce to form large patches. Then several spots coalesce and the leaf Dries up. Infection also occurs on the panicle, and neck with a brown colour appearance. Seeds are also infected(black or brown spots on glumes spots are covered by their brownish scorched appearance. Dark Brown or black spots also appear on glumes. The inflexion of the seed causes the failure of the seed

Germination, seedling mortality and reduces grain quality and weight. 50% yield reduction in severe cases.

Mode of spread and survival:

The infected seeds are the most common source of primary infection. The fungus also survives. On collateral hosts. The fungus can survive in the seed for more than 4 years. Infected seeds, Volunteer rice, infected rice debris, and several weeds are the major sources of inoculums in the Field. Infected seeds give rise to infected seedlings. The fungus can spread from plant to plant And in the field by airborne spores. The disease is common in nutrient-deficient soils and Unflooded soil but rare on rise grown on fertile soils. Abnormal soils, which are deficient in Nutrient elements, or soils in a much-reduced condition in which toxic substances accumulate Favour the development of the disease.

Fig 3.2 Brown spot

Leaf smut

Leaf Smut is one of the most commonly observed leaf diseases worldwide. But is considered Minor and causes little yield loss. Light yield losses have been reported when severe Leaf Smut Has caused premature desiccation of the leaves. The pathogen survives between crops as teliospores on diseased leaf debris in the soil. As the Teliospores germinate, the sporidia are spread to rice leaves where they germinate and infect in The late growth stages after booting.

Fig 3.3 Leaf Smut

4. ALGORITHMS IN ML

K-Nearest Neighbour (KNN)

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on the Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. This algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a good suite category by using the K-NN algorithm. K-NN is a non-parametric which means it does not make any assumption on underlying data. This algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

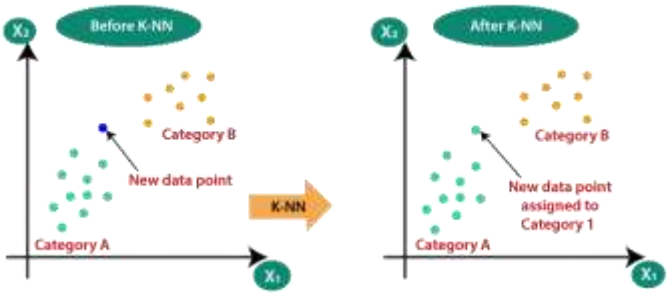


Fig 4.1 Classification of KNN

Decision Tree Classification Algorithm

A Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and

each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of that decision and do not contain any further branches. It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions. It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.

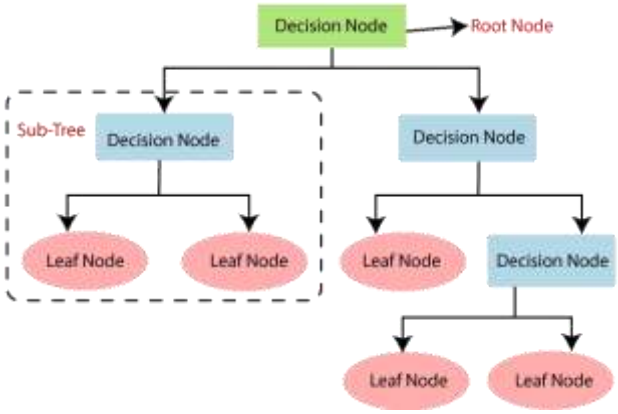


Fig 4.2 Decision making of a Decision tree

Support Vector Machine Algorithm

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for classification as well as Regression problems. However, primarily, it is used for classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called support vectors, and hence algorithm is termed a Support Vector Machine.

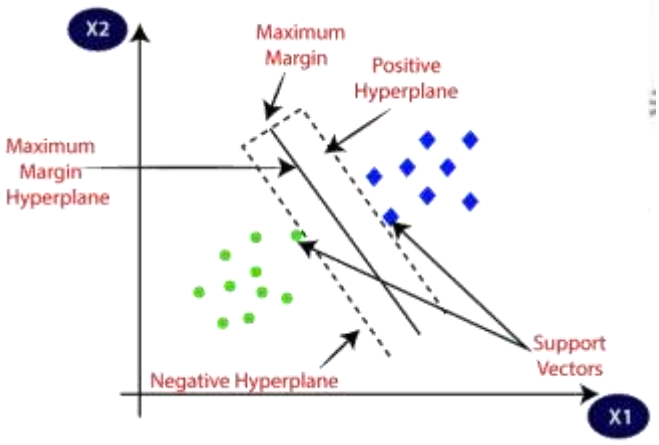


Fig 4.3 Classification of SVM

Types of SVM

SVM can be of two types

Linear SVM: Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

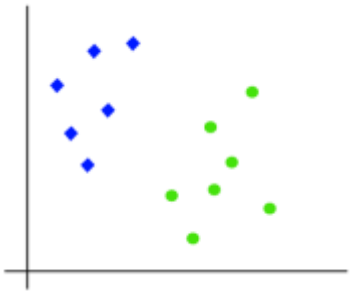


Fig 4.4 Linear SVM

Non-linear SVM: Non-linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and the classifier used is called a Non-linear SVM classifier.

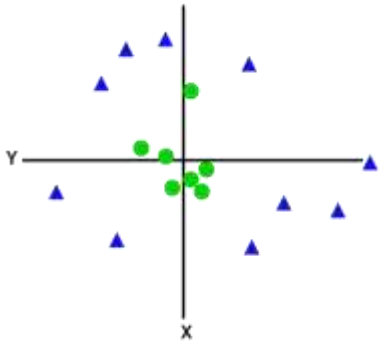


Fig 4.5 Non-Linear SVM

Naive Bayes Classifier Algorithm

The naive Bayes algorithm is a supervised algorithm, which is based on the Bayes theorem and used for solving classifier problems. It is mainly used in text classification that includes a high-dimensional training dataset. Naive Bayes classifier is one of the simple and most effective classification algorithms which helps in building fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. Some popular examples of the naive Bayes algorithm are spam filtration, sentimental analysis, and classifying articles.

Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and improve the performance of the model. As the name suggests, "Random forest is a classifier that contains a number of a decision tree on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

5. METHODOLOGY:

The proposed system is developed with many ML algorithms including SVM, Decision tree, Random forest, KNN, and Naïve Bayes. We will check the results for all these algorithms. All these algorithms are used for classification purposes. And further, the Augmentation technique is used. After Augmentation, we will once again verify the ML algorithms. Here we will find there is an increase in accuracy results in some algorithms. For this paddy leaf data set, we will conclude that this particular algorithm predicts more accuracy.

Here we will be performing various algorithms like SVM(support vector machine), KNN(k- nearest neighbour), Naïve Bayes, Random forest and Decision tree. Then we will be observing which is giving better results and further Augmentation process is implemented and checked again then the best results will be considered in the case. Here we are taking three types of leaf diseases i.e. Bacterial Blight, Brown spot and leaf smut. Testing and training is to be done with a different percentage and then the output is calculated. For training, we are considering 20% of the images and for testing the remaining 80% of the images are taken.

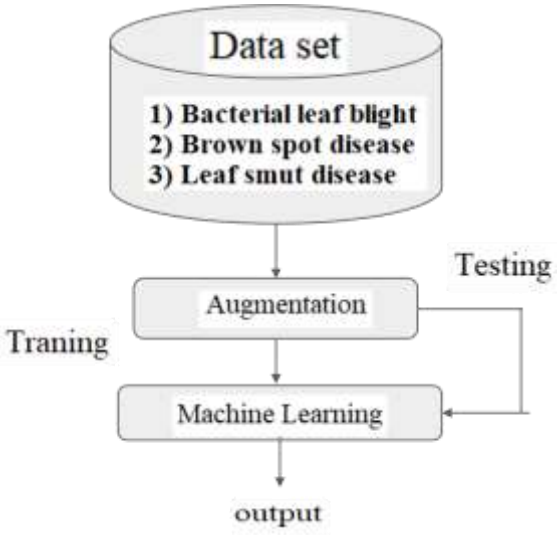


Fig 5.1 Methodology

Augmentation:

In the context of machine learning, augmentation refers to the process of increasing the size of the training dataset by generating new data samples from the existing ones. The purpose of augmentation is to improve the generalization and robustness of the machine learning models. By increasing the diversity of the training data, the model becomes less likely to overfit and better able to recognize and generate patterns in new, unseen data.

Types of Augmentation:

Image Augmentation: This includes rotating the Image by a certain angle, flipping the image horizontally or vertically, cropping the image to a larger or smaller size, changing the brightness, contrast etc.. and Adding some noise to the image.

Text Augmentation: Replacing certain words in the text with synonyms, randomly deleting and adding words from the text and finally swapping of text which are on adjacent sides.

Audio Augmentation: Shifting the pitch of the audio up or down, changing the speed and adding random noise to the audio signal, and finally changing and adjusting the volume of the audio signal.

6. RESULTS:

The evolution of accuracy is an important step in building effective models. There are various methods involved in finding the accuracy of machine learning algorithms, we use a confusion matrix which is used to evaluate the performance of a classification algorithm and is then trained on k-1 folds and tested on the remaining fold. This process is repeated k times, with each fold being used as a testing set once and finds the high accuracy.

Finding accuracy by using the confusion matrix :

True Positive(TP): The number of instances that were correctly predicted as positive

False Positive(FP): The number of instances that were predicted as positive but were actually negative.

True Negative(TN): The number of instances that were correctly predicted as negative.

False Negative(FN): The number of instances that were predicted as negative but were actually positive.

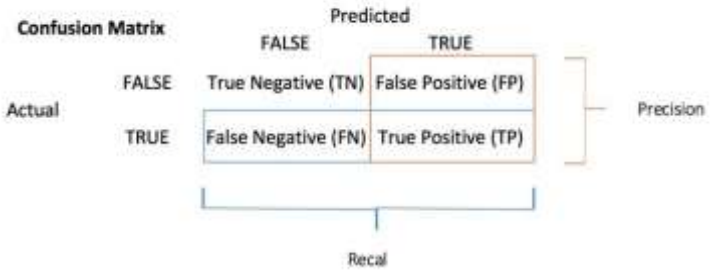
Accuracy=(TP + TN)/(TP+TN+FP+FN)

The precision and recall can also be calculated from the confusion matrix as:

Precision = TP/(TP + FP)

Recall = TP/(TP+FN)

F1 score = 2* (Precision * Recall)/(Precision + Recall)



EXPERIMENTAL RESULTS:

Algorithms	**Before Augmentation**	**After Augmentation**
SVM	69.57 %	48.571428571%
Random Forest	73.91304347%	60.0%
Decision tree	**56.521739130%**	**68.571428571%**
Naïve Bayes	**43.478260%**	**48.571428571%**
KNN	47.8260%	42.8571428571%

7. REFERENCES:

1. *Jitesh P. Shah et al. (2016) plant disease diagnosis and classification using machine learning and image processing*
2. *Geraldin B. Dela Cruz(2019) introduced automatic image processing using a smartphone device*
3. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering Vol. 7, Issue 2, February 2019 Copyright to IJIREEICE DOI 10.17148/IJIREEICE.2019.7220 97. Paddy Leaf Disease Detection Using Image Processing and Machine Learning(Dr. Neha Mangla1, Priyanka B Raj, SoumyaG Hegde3, Pooja R4)*
4. *Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with algorithm (S. Ramesh, D. Vydeki)*
5. *Plant Infection Detection Using Image Processing(Dr. Sridhathan C, Dr. M. Senthil Kumar) Department of ECE, Nalla Malla Reddy Engineering College.*
6. *Deep Learning Utilization in Agriculture: Detection of Rice Plant Diseases Using an Improved CNN Model Ghazanfar Latif 1,2, (Sherif E. Abdelhamid 3, Roxane Elias Mallouhy 1, Jaafar Alghazo 4 and Zafar Abbas Kazimi)*