

Decision Tree-based Machine Learning Algorithms to Classify Rice Plant Diseases

R. Sahith, P. Vijaya Pal Reddy, Satyanarayana Nimmala

Abstract: Rice is one of the most important foods on earth for human beings. India and China are two countries in the world mostly depend on rice. The output of this crop depends on the many parameters such as soil, water supply, pesticides used, time duration, and infected diseases. Rice Plant Disease (RPD) is one of the important factors that decrease the quantity and quality of rice. Identifying the type of rice plant disease and taking corrective action against the disease in time is always challenging for the farmers. Although the rice plant is affected by many diseases, Bacterial Leaf Blight (BLB), Brown Spot (BS), and Leaf Smut (LS) are major diseases. Identification of this disease is really challenging because the infected leaf has to be processed by the human eye. So in this paper, we focused on machine learning techniques to identify and classify the RPD. We have collected infected rice plant data from the UCI Machine Learning repository. The data set consists of 120 images of infected rice plants in which 40 images are BLB, 40 are BS, and 40 are LS. Experiments are conducted using Decision tree-based machine learning algorithms such as RandomForest, REPTree, and J48. In order to extract the numerical features from the infected images, we have used ColourLayoutFilter supported by WEKA. Experimental analysis is done using 65% data for training and 35% data for testing. The experiments unfold that the Random Forest algorithm is exceptional in predicting RPD.

Keywords: Rice plant disease, J48, Machine Learning, Classification.

I. INTRODUCTION

India and China are two major countries in the world in the production of rice [1]. Many people in the world consume rice as a primary food. The production and quality of rice depending on the kind of diseases and usage of pesticides during the lifetime of the rice plant. The accurate diagnosis and identification of rice plant disease is a major challenge for formers [2]. The natural way of identifying these diseases needs a lot of experience and knowledge. Present days many machine learning models are getting used to predict different diseases in the agricultural field [3]. So in this paper, we focused on the classification of rice plant diseases using decision tree-based machine learning models. The main motivation of using machine learning models is to enable formers to predict rice plant disease at the early stage of infection. is done using image filters [4]. The features extracted also can improve diagnostic results. Image filters

are nowadays expected to be automated feature learning The way of extracting the different features used in the prediction process also affects the performance of the system. The feature extraction for identifying rice diseases is models supported by many machine learning tools. Machine learning techniques really aid the farmers to identify the various diseases from the images of plants, the exemplar used here is to identify and classify the rice diseases based on the images of rice plant infections. We have detected three rice plant diseases namely BLB, BS, and LS. The extraction of features is based on color, shape, and texture of infected leaf using ColourLayoutFilter supported in WEKA. It supports many machine learning techniques such as supervised, unsupervised techniques, data preprocessing techniques, and feature selection techniques.

II. RELATED WORK

Although the rice plant is affected by many diseases, Bacterial Leaf Blight (BLB), Brown Spot (BS), and Leaf Smut (LS) are major diseases. In this paper, we focused on applying decision tree-based machine learning algorithms to classify rice plant diseases based on the features extracted from the rice plant images.

A. Brown Spot (BS)

BS is a fungal disease that infects leaves, leaf of the sheath, cloeoptile, glumes, and panicle branches. The infected leaves have many numbers of big spots on them. The big spots larger in number may kill the entire leaf. The disease attacks the crop from the initial stage in the nursery to the milkstage in the field [4]. BS is also called as fungal blight. The quality and quantity of rice fall down if BS is left untreated. If BS attacks the seed, Then different forms of seed such as spotted seeds, unfilled grains, or discolored seeds may be formed. This type of formation badly affects the quality of rice.



Fig. 1. Brown Spot Infected Rice Plant Image

Revised Manuscript Received on November 08, 2019.

*Correspondence Author

R. Sahith*, CSE, CVR College of Engineering, Hyderabad, India. Email: sahith.indian1010@gmail.com.

Dr.P.Vijaya Pal Reddy, CSE, Matrusri Engineering College, Hyderabad, India. Email: hodcse@matrusri.edu.in.

Satyanarayana Nimmala, CSE, CVR College of Engineering, Hyderabad, India. Email: satyauc234@gmail.com.

B. Leaf Smut (LS)

LS is a minor disease and is caused by the appearance of minute, sooty, dull, angular patches on the leaves, which which represents the sori. In susceptible varieties, the fungus maximum covers the total leaf surface of the older leaves. LS is caused by *Entylomaoryzae* [5]. The disease perennates through sori lying in infected plant debris. Spores reaching leaves near the soil is the reason for the infection. High nitrogen enhances disease incidence.



Fig. 2. Leaf Smut Infected Rice Plant Image

C. Bacterial Leaf Blight (BLB)

BLB is a serious threat to rice plants. It is caused by *Xanthomonas oryzae* pv. *Oryzae*. It causes shriveling of seedlings, drying, and yellowing of leaves. The malady is bound to be created in the zones that have weeds and stubbles of contaminated plants. It can happen in both tropical and calm conditions, especially in flooded and rainfed marsh territories. Specifically, the illness spread at his best at lower temperatures between 20 and 30 degrees, with relative stickiness above 70% [6]. It is normally seen when solid breezes and nonstop overwhelming downpours happen, permitting the ailment making microscopic organisms effectively spread through seepage beads on injuries of tainted plants. Bacterial scourge can be extreme in defenseless rice assortments under high nitrogen preparation.



Fig. 3. Bacterial Leaf Blight Infected Rice Plant Image

D. Existing Work

Harshad Kumar B.Pathiraju [7] proposed a method for identifying the rice plant diseases such as BLB, LS, and BS captured 40 images of each infected rice plant from the rice field. Extracted features like color, shape, and texture of each infected rice plant by using centroid based feeding and K-Means clustering for segmentation of each image disease area. They used the Support Vector Machine tool for multi-class classification and achieved 73.33% of accuracy on the test dataset. M.A Kahar [8] proposed early detection on paddy plant leaves with images of 28-BLB, 28-LB, 32-BSB. Canny Edge detection was used to extract the features of the and grammar. infected image. The

Neuro-Fuzzy method contains a back-propagation neural network with a combination of fuzzy logic-systems and rule-based expert systems. They achieved an average accuracy of 74.21 %. C.K.Charliepaul [8] proposed a method for classification of rice plant leaf using feature matching in which they used 6 images of each BS, BLB, LS to extract the features from infected by applying RGB mean and standard deviation to get the accuracy of 75% on IF-Then Classifier. S. Phadikar [9] proposed classification of Rice Leaf Diseases Based on Morphological Changes by taking 500 images of each disease (BS,BLB,LS) ,features extracted using Radial distribution of hue from center to boundary of the infected rice plant image spot and applied SVM classifier and achieved 68.1% accuracy in detection of the rice plant disease. Whereas in our proposed methods 65% of data is used for training and 35% of data is used for testing and the RandomForest algorithm has shown 76.19% accuracy in classifying rice plant diseases.

III. EXPERIMENTAL RESULTS

Experimental analysis is done using 120 images of infected rice plants. The data set is collected from the UCI Machine Learning Repository. The collected data set has 40 images of BS, 40 images of LS and 40 images of BLB. 78 images are used for training and 42 images are used for testing. We applied the ColorLayout filter supported by WEKA to extract the features from the images for experimental analysis. Extracted 33 features are used in conducting experiments using different decision tree-based algorithms.

Table-I: Confusion matrix generated using RandomForest

		Predicted Class		
		BS	LS	BLB
Actual Class	BS	11	1	0
	LS	2	12	1
	BLB	4	2	9

The performance of each algorithm is tabulated in Table I, II, III, IV, V, VI as given below. Confusion matrix represents in Table 1 shows that 11 BS infected images are classified as BS, 1 BS infected image as predicted as LS and no BS infected image is classified as BLB. The same is with LS and BLB. Fig. 4,5,6 represents various performance measures of classifiers such as TP Rate, FP Rate, precision, and recall.

Table -II: Confusion matrix generated using J48

		Predicted Class		
		BS	LS	BLB
Actual Class	BS	7	3	2
	LS	4	9	2
	BLB	3	4	8

Table-III: Confusion matrix using REPTree

Actual Class		Predicted Class		
		BS	BS	BS
	BS	8	4	0
	LS	2	10	3
	BLB	1	5	9

Table -IV: Performance Measures of RandomForest Algorithm

TP Rate	FP rate	Precision	Recall	F-Measure	Class
0.917	0.200	0.647	0.917	0.759	BS
0.800	0.111	0.800	0.800	0.800	LS
0.600	0.037	0.900	0.600	0.720	BLB

Table-V: Performance Measures of J48

TP Rate	FP rate	Precision	Recall	F-Measure	Class
0.583	0.233	0.500	0.583	0.538	BS
0.600	0.259	0.563	0.600	0.581	LS
0.533	0.148	0.667	0.533	0.593	BLB

Table-VI: Performance Measures of REPTree

TP Rate	FP rate	Precision	Recall	F-Measure	Class
0.667	0.100	0.727	0.667	0.696	BS
0.667	0.333	0.526	0.667	0.588	LS
0.600	0.111	0.750	0.600	0.667	BLB

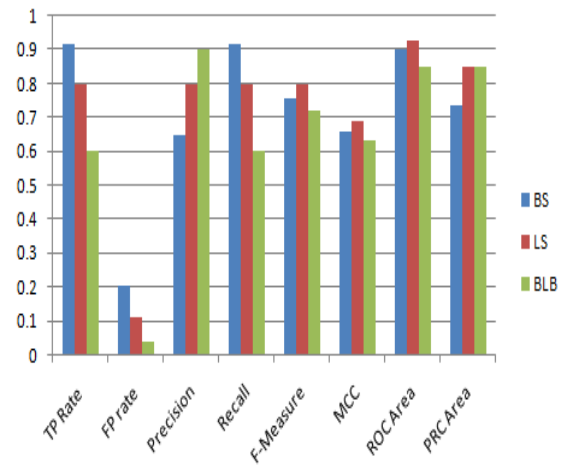


Fig. 4. Performance measures using RandomForest

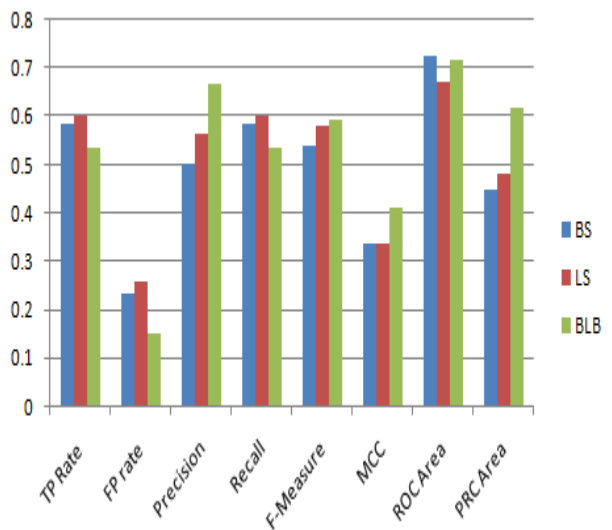


Fig. 5. Performance measures using J48

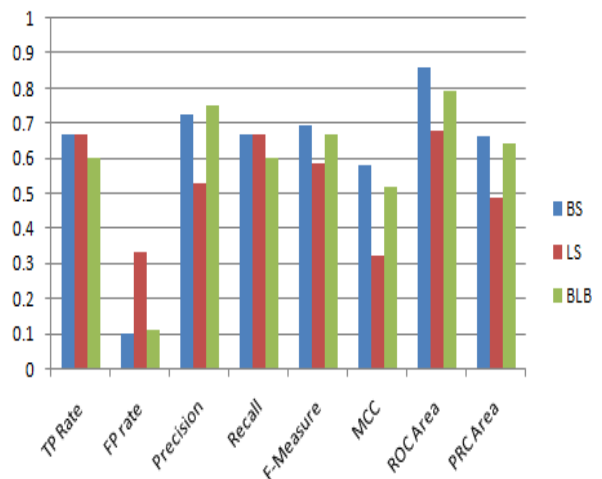


Fig. 6. Performance measures using REPTree

IV. COMPARATIVE ANALYSIS

Experiments are conducted using RandomForest, J48, and REPTree. Although the Mean Absolute error is less for the J48 algorithm but the accuracy of J48 is less than the RandomForest. RandomForest is outperformed in accuracy and also exceptional in reduced

Root Mean squared error compared with J48 and REPTree. Table VII and Fig. 7 shows the comparative analysis of different classifiers

Table-VII: Accuracy Details of Different Classifiers

Algorithm	Accuracy	Mean Absolute Error	Root Mean Squared Error
RandomForest	76.19	0.3071	0.369
J48	57.14	0.2834	0.525
REPTree	64.28	0.3	0.4138



Fig. 7. Accuracy Details of Different Classifiers

V. CONCLUSION AND FUTURE WORK

Rice Plant Disease (RPD) is one of the important factors that decrease the quantity and quality of rice. Identifying the type of rice plant disease and taking corrective action against the disease in time is always challenging for the farmers. Machine learning techniques are nowadays helping decision-makers to take accurate decisions in different fields. The primary idea of developing the machine learning models for rice plant disease classification is to provide a simplified and easily understandable system to detect rice plant infections in advance by the farmers. Secondly, extracting effective features for identifying rice diseases, this can be done using image filters. In this paper, we have used different decision tree-based machine learning techniques to classify rice plant diseases. The random forest algorithm has shown 76.19% accuracy. In future other machine learning techniques such as fuzzy-based classifiers, neural network-based classifiers and Bio-inspired techniques may be used to improve the accuracy of the classifier.

REFERENCES

1. J. P. Shah, H. B. Prajapati, and V. K. Dabhi, "A survey on detection and classification of rice plant diseases," *2016 IEEE International Conference on Current Trends in Advanced Computing of Rice Leaf Diseases Based on Morphological Changes*, International Journal of Information and Electronics Engineering, vol. 2, pp. 460-463, May 2012 (ICCTAC), 2016.
2. S. Phadikar, J. Sil, and A. K. Das, "Classification of Rice Leaf Diseases Based on Morphological Changes," International Journal of Information and Electronics Engineering, vol. 2, pp. 460-463, May 2012.

3. H. B. Prajapati, J. P. Shah, and V. K. Dabhi, "Detection and classification of rice plant diseases," *Intelligent Decision Technologies*, vol. 11, no. 3, pp. 357-373, 2017.
4. M. S. T. Kumbhare, "Detection and Classification of Diseases on Rice and Sugarcane: A Review," *International Journal for Research in Applied Science and Engineering Technology*, vol. 7, no. 4, pp. 2927-2929, 2019.
5. J. W. Orillo, T. M. Amado, N. M. Arago, and E. Fernandez, "Rice Plant Disease Identification And Detection Technology Through Classification Of Microorganisms Using Fuzzy Neural Network," *Jurnal Teknologi*, vol. 78, no. 5-8, 2016.
6. V. K. Shrivastava, M. K. Pradhan, S. Minz, and M. P. Thakur, "Rice Plant Disease Classification Using Transfer Learning Of Deep Convolution Neural Network," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-3/W6, pp. 631-635, 2019.
7. H. B. Prajapati, J. P. Shah, and V. K. Dabhi, "Detection and classification of rice plant diseases," *Intelligent Decision Technologies*, pp. 1-18, 2018.
8. M.A. Kahar, S. Mutalib, A. Rahman, "Early Detection and Classification of Paddy Diseases with Neural Networks and Fuzzy Logic", *Recent Advances in Mathematical and Computational Methods*, pp. 248-257.
9. C.K.Charliepaul, "classification of rice plant leaf using feature matching", *International Journal On Engineering Technology and Sciences* vol. 1, pp. 290-295, 2014.

AUTHORS PROFILE



R. Sahith received his M.Tech in Computer Science and Engineering from JNTUH, India and currently pursuing his Ph.D. in Informatics, Department of Computer Science and Engineering from Osmania University, Hyderabad, India. He is currently working as an Assistant Professor in the Department of Computer Science and Engineering at CVR College of Engineering, Hyderabad, India. His research interests are programming and machine learning.

Email: sahith.indian1010@gmail.com



Dr. P. Vijaya Pal Reddy currently working as a professor in the department of CSE, Matrusri Engineering College, Hyderabad, India. His research interests include text mining, natural language processing, data mining, and optical character recognition. Currently, he is guiding research scholars at OU, JNTUA, India.

Email: hodcse@matrusri.edu.in



Satyanarayana Nimmla received his B.Tech from Kakatiya University, India and his M.Tech from JNTUH, India. He is currently pursuing his Ph. D. from Osmania University, Hyderabad, India. His research intrests include programming, and machine learning.

Email: satyauce234@gmail.com