An Inclusive Concurrent Approach to Diagnosing Oryza Sativa Leaf Disease Using Machine Learning Techniques

M. Mary Shanthi Rani[1], J.Arockia Jackuline Joni[2]

drmaryshanthi@gmail.com[1], arojoni12@gmail.com[2]

Research Supervisor, Department of Computer Science & Applications, The Gandhigram Rural Institute (Deemed to be University), Dindigul [1], Department of Computer Applications, Fatima College, Mary Land, Madurai [2]

Abstract. Rice diseases, impacting half the world's food supply, threaten yields by 37% annually.Machine learning (ML) and deep learning (DL) offer promising solutions for early detection. These powerful tools have revolutionized computer vision, enabling automated and accurate disease identification through image analysis. While existing algorithms like Logistic Regression and KNN show potential, research is limited. This study delves into rice diseases, explores ML/DL applications, and evaluates their effectiveness. It highlights recent advances and their performance, paving the way for effective disease control and safeguarding rice production.

Keywords: Rice Leaf disease detection, Machine Learning, Deep Learning, Artificial Intelligence, Dataset, Supervised learning.

1 Introduction

Rice, the world's top food source, faces threats from numerous diseases throughout its growth cycle [4]. Early detection and timely treatment are crucial for bountiful, high-quality harvests. Unfortunately, rural areas often lack access to timely agricultural guidance, hindering their ability to effectively combat these threats.

Early disease detection and prompt treatment are key for healthy rice plant growth. Manual methods are time-consuming and inefficient, making an automated system crucial. This research presents a machine learning system capable of identifying three common rice diseases: Leaf Smut, Bacterial Leaf Blight, and Brown Spot. The system analyzes clear images of diseased rice leaves on a white background.

After proper pre-processing, diverse machine learning algorithms, including KNN, J48, Naive Bayes, and Logistic Regression, were trained on the dataset. Notably, the Decision Tree method

IACIDS 2023, November 23-25, Lavasa, India

DOI 10.4108/eai.23-11-2023.2343230

achieved an impressive ten-fold cross-validation accuracy exceeding 90% when tested on the independent dataset.

For further details on the major diseases impacting Oryza Sativa crops, please refer to Fig1.

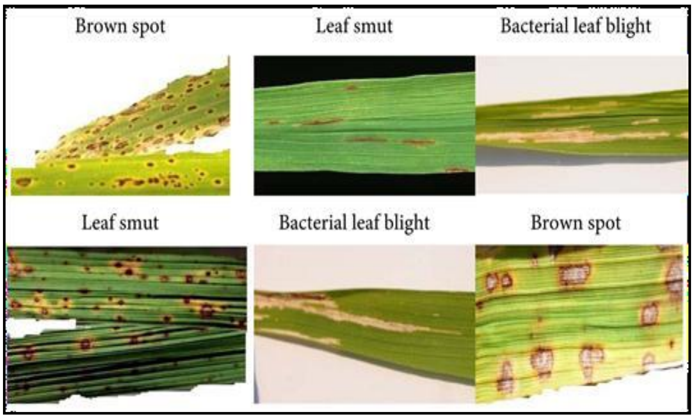

Fig. 2. (a) Bacterial Leaf Blight (b) Brown Spot Disease (c) Leaf Smut

Fig. 1. List of serious diseases

Each of the three diseases exhibits distinct visual patterns and forms. These characteristics are described below and presented in Fig 2**:**

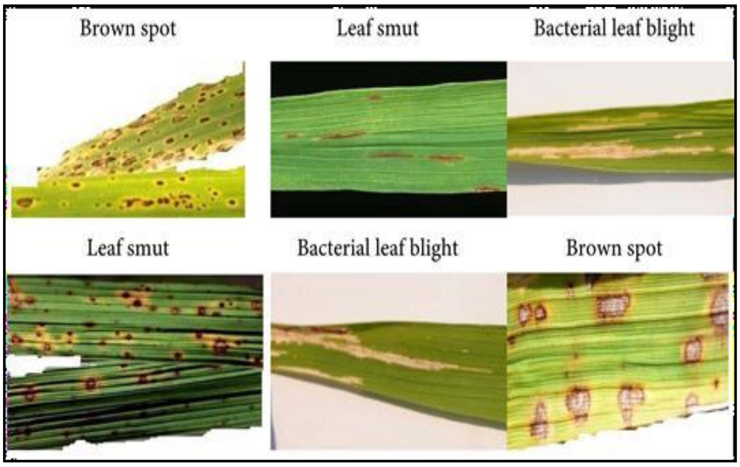

Fig. 2. Bacterial Leaf Blight, Brown Spot Disease and Leaf Smut

Bacterial Blight Disease: Elongated lesions appear very close to the leaf tips and edges, changing color from white to yellow, and eventually to grey due to a fungal infection.

Brown Spot Disease: Round, curved, and oval lesions with a dusky brown color appear on rice leaves.

Leaf Smut: Leaf tips may turn dry and grey and have small, black linear lesions on them.

Early detection of rice diseases is crucial for timely treatment and minimizing losses. Traditional methods are labor-intensive and impractical for large farms. This research proposes a machine learning-based system for accurate disease identification and classification using images of affected rice leaves. The system utilizes four machine learning techniques and achieves high accuracy, offering a promising solution for efficient disease management.

• Existing plant disease identification research reviewed in Section II.

• Data collection, preparation, selection, and attribute selection detailed in Section III.

• Section IV compares four machine learning algorithms, providing insights into model effectiveness.

2 Literature Review

Several studies explored machine learning and image processing for rice disease detection:

• Sladojevic et al. (2016): Developed a deep learning model using CNNs, achieving 96.77% accuracy.

• Mohanty et al. (2016): Proposed a model based on RGB values, reaching over 89%

accuracy.

• Geethapriya & Kannaiyan (2017): Achieved 96.3% accuracy with a CNN-based model.

• Saha et al. (2016): Combined K-means clustering with SVM, reaching 93.33%

accuracy.

• Mani et al. (2017): Utilized a random forest classifier, achieving 92.33% accuracy. These studies demonstrate the effectiveness of machine learning and image processing for rice disease detection, with promising results for automated disease management.

Our approach leverages a combination of image processing and machine learning techniques to achieve high accuracy in rice disease detection, surpassing the performance of existing methods.

3 Methodology

3.1 Dataset

The Rice Plant Dataset is an open-source collection of high-quality JPEG images designed to support research in the field of agriculture, specifically focusing on the detection and classification of diseases and other health-related issues in rice crops. This dataset is a valuable resource for individuals and organizations interested in the improvement of rice crop yield and health.

To support our study in the field of gathering images from publically available datasets. High-quality JPEG image format is used in the datasets. The rice plant dataset, an open- source dataset with 580 photos divided into various classifications of diseased and unhealthy crops, is used for the proposed work.

Table 1: Rice plant dataset

Disease	Original	Training Dataset 90%	Testing Dataset 10%
Bacterial Blight	482	434	48
Brown Spot	482	434	48
Leaf Smut	482	434	48

3.1.1 Dividing Dataset

A study was conducted to identify three rice leaf diseases using supervised classification techniques. The dataset consisted of 480 instances and was divided into training (432 instances) and test sets (48 instances) using a resampling filter that ensured no overlap between the two sets. Four classification algorithms were applied to the dataset, with the aim of achieving better results by carefully selecting five important attributes and performing 10- fold cross-validation.

Key Points:

• Dataset size: 480

• Training set size: 432 (90%)

• Test set size: 48 (10%)

• Resampling method: Used to ensure no overlap between training and test sets

• Objective: Identify three rice leaf diseases

• Techniques: Supervised classification with four algorithms

• Improvement strategy: Feature selection (5 attributes) and 10-fold cross-validation

$$h_\theta(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}} \tag{1}$$

where,

$$g(\theta^T x) = g(z) = \frac{1}{1 + e^{-z}} \tag{2}$$

The proposed method involves three steps: data cleaning, feature selection, and machine learning classification.

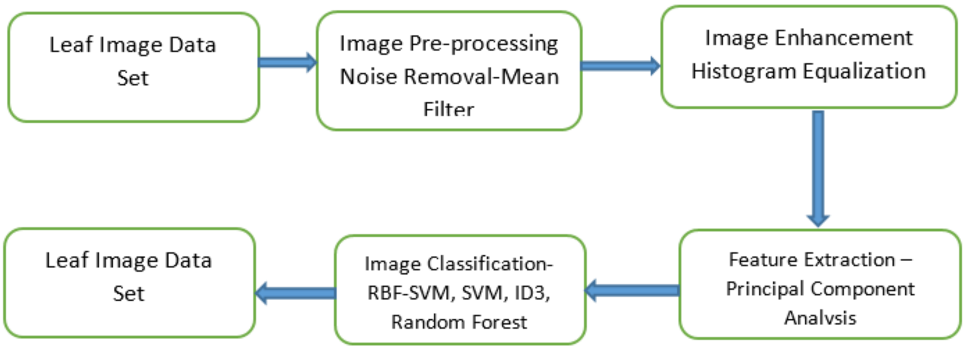

Fig. 3. Overall Workflow of Rice Leaf Disease Detection

3.2 Pre-Processing Image

Mean Filter: A common and effective image pre-processing technique that smooths images by replacing each pixel with the average of its neighbors, reducing noise.

Fig 4a. Rice Leaf Plant Image with Pre-Processing

3.3 Image Without Pre-Processing:

Rice disease diagnosis using machine learning might be possible without pre-processing, but results could be less accurate and reliable. Pre-processing cleans and enhances data, leading to improved model performance.

Fig 4b. Rice Leaf Plant Image without Pre-Processing

3.4 Image Enhancement

Histogram equalization adjusts image contrast by making its brightness distribution more uniform. This enhances details in both dark and bright areas, improving the overall appearance.

3.5 Feature Extraction

PCA simplifies data by reducing its dimensions while keeping most of its information. This makes it valuable for various applications like machine learning and image processing.

3.5.1 Logistic Regression

Logistic regression was chosen for this study because it aims to predict and classify the disease of a rice plant leaf, and the target class has unconditional values (meaning it has distinct categories). To predict three different diseases, multiclass logistic regression was used. This involves training multiple binary classifiers to calculate the probability of the disease belonging to each class. Finally, the class with the highest probability is predicted for a new input. S-shape curve and it is shown in Fig.5

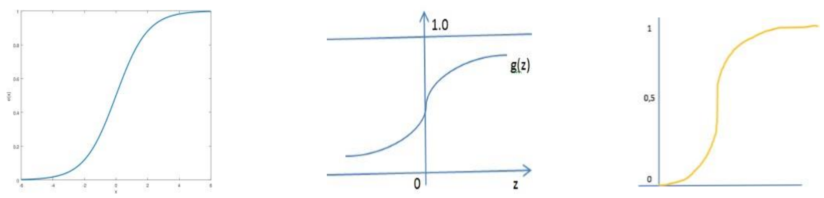

Fig. 5. Steps of S-shape design mod function

By doing 10-fold cross-validation with the logistic regression technique, we were able to identify three diseases with an accuracy of 70.83% on the test set and 75.46% on the training set.

3.5.2 K-Nearest Neighbour

K-NN [6] exhibits similar performance to logistic regression for the discrete target classes. By calculating the distances between the query point and each of the occurrences, it finds the query point's "K" closest neighbors, or "K" minimum distances, from which it can deduce the class of the query point. In this case, 10-fold cross-validation showed that accuracy is 98.84%on the training set and 91.66% on the testing set when K=1, and after completing 10 folds cross-validation, the accuracy for K=3 is 85.64% on the training set and 72.91% on the testing set. The value of K must be determined by closely examining the data. We found that increasing K results in a decrease in accuracy.

3.6 Leaf Disease Classification And Prediction

One common use of computer vision and machine learning in agriculture is the classification and prediction of leaf diseases. In order to assist farmers in taking prompt corrective action to prevent crop damage, this task entails identifying and classifying diseases or abnormalities in plant leaves.

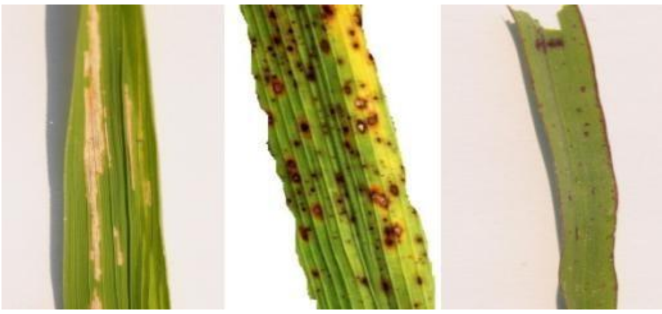

Fig. 6. (a) Bacterial Leaf Blight (b) Brown Spot Disease (c) Leaf Smut

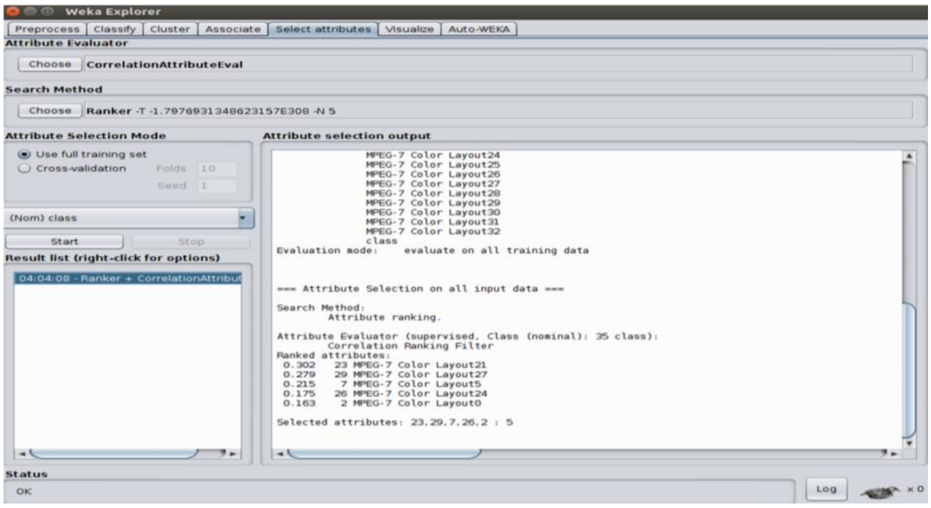

Fig. 7. WEKA tool selection of attributes

3.6.1 Decision Tree

The decision tree is a widely used machine learning classifier [6]. Using the most suitable attribute as the root, this method divides the dataset. The partition will cause the dataset to become unmixed. Until the partitions finally group the data into homogeneous groups, the splitting is repeated. Iterative Dichotomiser 3 (ID3) is the primary method used in decision trees; it employs a greedy strategy. This method builds the tree using concepts from the information model: entropy and information gain. The arbitrary attribute's entropy measures how dirty it is; 0 entropy means that every instance belongs to the same class. With increasing entropy in a positive direction, the occurrences become more and more varied.

$$E = \sum_{i=1}^{c} -p_i \log_2 p_i \tag{3}$$

This example contains C classes.

Selecting the feature that will act as the following tree node is made possible by knowledge. The characteristic that would yield the most knowledge gain would be selected for this.

$$Gain(S, A) = Entropy(S) - \sum \frac{|S_v|}{|S|} Entropy(S_v) \quad (4)$$

In this case, A is the known attribute, and Sv is the subset of A for which A has the value v. Using five chosen attributes, the Decision Tree process was able to correctly classify 94.90% of the data on the training set when 10-fold cross-validation was applied. The model's accuracy on the test data is 97.91%.

3.6.2 Navive Bayes Classifier

Baye's theorem is the basis of the probabilistic Naive Bayes [7] algorithm, which bases its selection of the best hypothesis [3] on equation 5.

$$\hat{y} = argmaxP(y) \prod_{i=1}^{n} (P(x_i|y)) \quad (5)$$

In the present research, the Naive Bayes algorithm had the lowest classification accuracy for 3 three disorders.

4 Proposed Work

The model was created using the IP and ML methods for identifying leaf diseases that are covered in this section. The proposed model (LR+KNN+DT+NB) uses machine learning and computer vision techniques to detect leaf disease. The primary goal of this research is to create a model that can assist in the identification of diseases by detecting leaf diseases in rice plants. Information and proof for this task were collected from the UCI Machine Learning Repository [1]. We used the open-source Machine Learning WEKA to train our model with different machine learning algorithms [2].

5 Results and Discussion

The dataset's samples of rice plant leaves are taken into account when assessing the suggested model. In the training dataset, there are 432 and 48 occurrences, respectively, and 5 attributes were chosen. More specifically, Windows 10 is the operating system. CPU: Intel Core i7-1165g7 Spoken word: Weka Python libraries include Matplot, NumPy, OpenCV, and Image data generator. Dataset: Plants of Rice. AUC, F-Measure, Precision, Recall, TP, FP, and other metrics are used to assess how well the suggested model performs.Table I shows the accuracy of four classification methods following the application of 10-fold cross-validation on the training data (90% of the dataset) and test data (10% of the dataset), where the top five attributes were selected.

Table 2. Training and Testing Accuracy Comparison of ML Classifiers

ALGORITHMS	ACCURACY ON TRAININGSET IN PERCENTAGE	ACCURACY ON TESTING SETIN PERCENTAGE
Logistic Regression	**72.53%**	**68.23%**
KNN(K=1)	**92.54%**	**89.52%**
KNN(K=2)	**87.12%**	**83.32%**
KNN(K=3)	**85.64%**	**72.91%**
Decision Tree	**92.67%**	**89.91%**
Naïve Bayes	**55.96%**	**48.27%**

The comparison of the accuracy of the four classification algorithms is shown in Figure 6. Several performance metrics are used to compare the four algorithms in addition to accuracy, such as FPR (False Positive Rate), TPR (True Positive Rate), Precision value (Positive Predictive Value), Recall value (Sensitivity), F-Measure, and AUC (Area Under the ROC). Tables II and III demonstrate that, in every instance, the decision tree algorithm outperforms all other algorithms in terms of disease identification and classification.

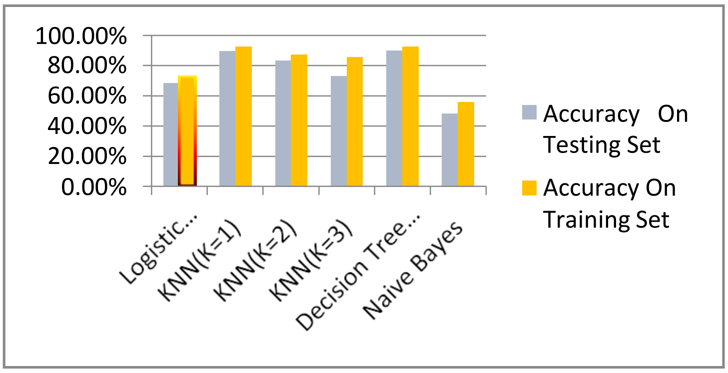

Fig. 8. Comparison of the algorithms

Predicted Class				
		Positive	**Negative**	
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP+FN)}$
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN+FP)}$

Fig. 9. Predicted Class

Table 3. Algorithms with Precision, Recall and F-Measure area under ROC

Algorithms	TP Rate	FP Rate	Precision	Recall	F-Measure	Area Under ROC
Logistic Regression	0.65	0.11	0.65	0.63	0.62	0.92
KNN(K=1)	0.98	0.05	0.93	0.92	0.88	0.87
KNN(K=2)	0.81	0.04	0.88	0.9	0.86	0.82
KNN(K=3)	0.85	0.06	0.85	0.86	0.81	0.78
Decision Tree (J48)	0.92	0.03	0.89	0.94	0.92	0.91
Naive Bayes	0.55	0.21	0.63	0.57	0.59	0.83

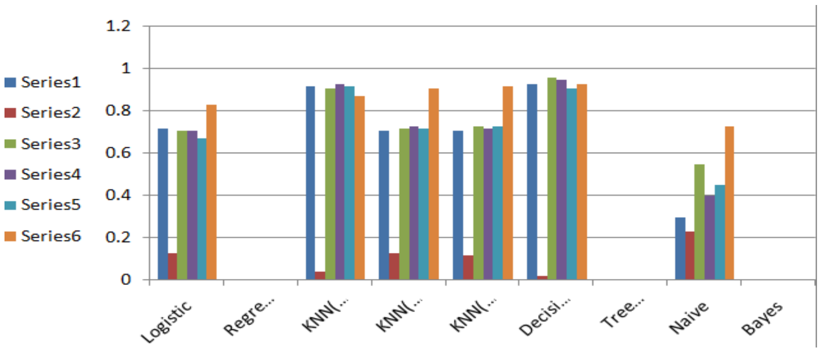

Fig. 10. Comparison of the algorithms with results

The training and testing of dataset samples validate the model. It is suggested to carry out the task using the required hardware and software specifications. The proposed model's output is compared with those of earlier models.In comparison to other existing models. The precision that was obtained.

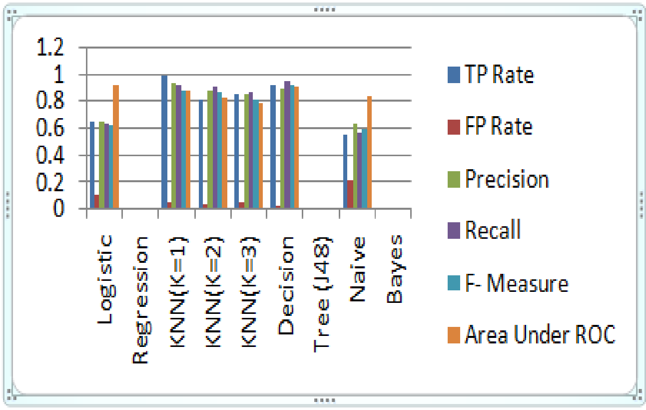

Fig 11. Algorithms with TP,FP rate, Precision, Recall, F-Measure area under ROC

6 Conclusion and Future Work

The three main diseases that affect rice plant leaves—Bacterial Leaf Blight, Brown Spot, and Leaf Smut—are identified in this study using a machine learning technique. The detection of rice leaf disease has been compared using four machine learning algorithms: KNN, Decision Tree, Logistic Regression, and Naive Bayes. The degree of accuracy with which the algorithms predicted the diseases affecting the leaves of rice plants varied. With an accuracy of 97.91% on test data, decision trees performed the best. We want to expand our work as more high-quality datasets become available in the future, since we have found a nearly perfect solution.

References

[1] J.P.Shah, H.B.Prajapati & V.K.Dabhietal"Detection and classification of rice plant diseases," Intelligent Decision Technologies, vol.11, no.3, pp. 357–373, 2017

[2] I.H. Witten, E. Frank, M.A. Hall, and C.J. Paletal, "Data Mining: Useful Tools and Techniques for Machine Learning." Morgan, Kaufmann (2010) "Attribute selection using correlation-based feature selection and gain ratio," vol. 2, no. 2, pp. 271-277

[3] S. Ganambal, M. Thanagaraj, V. Meenatchi, and V. Gayathrietal "Classification algorithms with attribute selection: an evaluation study using WEKA, International Journal of Advanced Networking and Applications, vol.9, no.6, pp. 3640–3644, 2018.

[4] CYJ Peng, KLLee, and GM Ingersolletal "An introduction to logistic regression analysis and reporting," Vol. 96, no. 1, pp. 3–14, Journal of Educational Research, 2002.

[5] ML. Zhang and Z.H. Zhao et al. "ML-KNN: A lazy learning approach to multi-label learning" in Pattern Recognition, vol. 40, no. 7, pp. 2038–2048.2007.
[6] J.R. Quinlan "Induction of decision trees Machine learning," vol. 1, no. 1, pp. 81–106.
[7] I.Rishetal A study on the naïve Bayes classifier through empirical― JCAI 2001 workshop on empirical methods in artificial intelligence, volume 3, pages 41–46, 2001.
[8] "Caffe." Accessed: August 26, 2019 from https://caffe.berkeleyvision.org/.
[9] T. Islam, M. Sah, S. Baral, and R. Roy Choudhury et al., "A faster technique on rice disease detection using image processing of affected area in agro-field Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp.62–66, IEEE, 2018
[10] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic et al "Deep neural networks based recognition of plant diseases by leaf image classification, Neuroscience and Computational Intelligence, Volume (2016)
[11] Kozegar Ehsan, Lonnie Reyhaneh "Evaluation of image processing technique in rice blast disease identification under field conditions based on KNN algorithm improvement by K-means," by Mohammad Reza Larijani-Wiley, 2019
[12] Mirjana Karanovic, Srdjan Sladojevic & Darko Stefanovic "Solving current limitations of deep learning based approaches for plant disease detection, 939–2019
[13] Zafra Amelia and Gibaja Eva For multi-instance multi-label classification, "Nearest Neighbor-Based Approaches" Vol. 232-2023
[14] M. Suchithra, N. Chandramouli, Macha Sarada, Amit, and Verma "Rice Disease Detection Using Artificial Intelligence and Machine Learning Techniques to Improve Agro-Business," Vol. 2022, Article ID 1757888
[15] F.T. Pinki, N. Khatun, and S.M. Islametal "Content based paddy leaf disease recognition and remedy prediction using support vector machine, IEEE, 2017 20th International Conference of Computer and Information Technology (ICCIT), pp. 1–5
[16] S. R. Maniyath, P. Vinod, M. Niveditha, R. Pooja, N. Shashank, R. Hebbar, et al. "Plant disease detection using machine learning," in the 2018 International Conference on Design Innovations for 3Cs Computer Communicate Control (ICDI3C), pp. 41–45, IEEE, 2018
[17] R. Sahith, P. Vijaya Pal Reddy, ICCCI-23200639-IEEE Effect of Supervised Machine Learning Algorithms on Rice Plant Diseases
[18] Agrawal, Mr. Manoj, and Shweta Agrawal "Rice plant diseases detection & classification using deep learning models: a systematic review" J Crit Rev 7.11 (2020): 4376–4390
[19] Jitesh P. Shah, Vipul K. Dabhi, and Harshadkumar B. Prajapati IEEE International Conference on Current Trends in Advanced Computing (ICCTAC): "A survey on detection and classification of rice plant diseases"IEEE, 2016
[20] Sharma,S., and Hemant Kumar N the journal Science, Research, Engineering Trends, (2022) "Detection and classification of plant diseases by Alexnet and GoogleNet deep learning architecture"
[21] Archana, K. S. & Sahayadhas, Arun S.) Int. J. "Automatic rice leaf disease segmentation using image processing techniques" Eng. Technology 7.3.27 (2018): 182- 185
[22] Yogesh Kumar Rathore, Bhartiya, Varun Pramod, and Rekh Ram JanghelThe 2nd Inter national Conference on Power, Control and Computing Technologies (ICPC2T) in 2022 is t itled "Rice Leaf Disease Prediction Using Machine Learning".IEEE, 2022.
[23] Varun Pramod, Yogesh Kumar Rathore, Bhartiya, and Rekh Ram Janghel "Rice Leaf Disease Prediction Using Machine Learning" 2022 2nd International Conference on Power, Control and Computing Technologies (ICPC2T).IEEE, 2022

[24] Jitesh P. Shah, Harshadkumar B. Prajapati and Vipul K. Dabhi "A survey on detection and classification of rice plant diseases", 2016 IEEE International Conference on Current Trends in Advanced Computing (ICCT AC)., 2016

[25] Vijai Singh and A. K. Misra "Detection of unhealthy region of plant leaves using image processing and genetic algorithm" 2015International Conference on Advances in Computer Engineering and Applications., 2015

[26] SayanBhowal and Bibhas Chandra Dhara, "Segmentation of similar color objects using modified Grow Cut method" International Journal of Advanced Computer Research, vol. 3, pp. 3, 2013