

Rice Plant Leaf Disease Detection using Machine Learning

Sonu Singh

Department of Computer Science and Applications, Dr. Vishwanath Karad MIT World Peace University, Pune India.
Email: sonusingh112@gmail.com

Abstract— Recent advancement in A.I. have significantly extended the smart farming and the early invention of plant syndrome came with a new way of efficient and accurate practices resultant in diagnosis of leaf and fruit problems. There are so many efforts taken in the type of plant disorder. This research examination basically used on image analysis which is frequently done by the hand of scholars. There are several image related databases of information regarding plant syndrome. Both the data researcher (scientists) and agriculture specialists are interested about secondary use of the photos for M2 -based analysis and farming. Advances in M.L. such as deep learning has produced very accurate outcomes in the field of plant syndrome. Current article discusses the fascinating contributions made by the numerous researchers working on this topic and draws attention to the difficulties that are thought to be crucial to this time of enquiry. Additionally, a few popular ML techniques employing KNN, Random Forest and Logistic regression, produced classification of accuracy with 96.2%, 98.6% respectively.

I. INTRODUCTION

Agriculture is the crucial support of livelihood on earth and the economies of many countries are dependent on agriculture on large scale. 60% of people in India depends upon farming as occupation which is 17% of our GDP so plant syndrome recognition is very necessary in the field of agriculture [1]. Leaf disease is a big hindrance in the growth of a plant. Various viruses like fungi, bacteria are also harmful for plants. Prevention is added by the application of technologies like ML and computer vision. The research paper enlists the significant contribution of many effective approaches to diagnose leaf syndrome bases on ML and intellectual property technologies and provides the overview of these approaches. This research study explores methods such as segmentation, features extraction, classification and image preprocessing [2].

ML and other recent developments have led to extremely accurate results in the analysis to recover a variety of plant illnesses. The challenges believed to be essential to this field of study are highlighted in this article which also explores the intriguing contributions made by numerous scholars works on the subject.

The main contributions of this study are:

- Development of a novel architecture for rice leaf disease detection utilizing machine learning.
- Utilization of diverse algorithms, including decision tree, k-NN, random forest and Linear Regression
- Conducting simulations both with selected features for comprehensive result analysis.
- Finding the most effective model from a set of candidate models through rigorous evaluation.

The structure of the paper is as follows: section 2 details an overview of related research; Section 3 outlines the proposed methodology employed for leaf disease detection using machine learning. Section 4 presents the results and discussions, followed by the comparative analysis. Lastly, section 5 offers the conclusion of the study.

II. LITERATURE REVIEW

Many Machine Learning and Deep Learning techniques have been employed to automate the diagnosis of leaf diseases. Some of the related works are as follows:

Using DWT, PCA, GLCM, and CNN algorithm, Sunil S. Harkannavar, et al [3] assessed 600 samples in total from the tomato leaf disease dataset in order to assess the proposed model with the help of fusion methodologies to extract significant attributes, this approach achieved a greater accuracy of 99.5%. going forward, the model can be evaluated and enhanced using new dataset leaf samples in this SGD+DL+SVD+RF method is used. Sarfaraz Newroj, et al [4] made SGD+DL+SVD+RF method is used to analyse citrus plant to test this model. Umangeshwari A. [5] using FCM+SIFT+LSTM algorithm, S. Deepa and K. Raja developed an improved idea for recognizing leaf illness through this application by DL. Marwan Mustafa, et al [6] provides a deep learning based plant leaf disease classification and detection method. The convolutional neural network (CNN) was used to analyze the pics, taken from the Plant Village dataset website. The CNN produced great results for training and testing; for all used dataset. The accuracy was 98.029 percent for testing and 98.29 percent for training. To add more varieties of plants in the future to increase the variety of plants. Create an expert system that is capable of classifying and identifying plant leaf diseases. Gemini Kishore et al [7] suggested utilizing segmentation by image processing to diagnose the tomato plant leaf disease. Nilay Ganatra and Atul Patel [8] proposed a model on A Multiple class Disease diagnosis using Image Processing and M L Techniques. This study suggested a prediction model that makes use of Machine Learning and techniques for the recognition and categorization of plant leaf syndrome. Preprocessing and segmentation are used to extract many aspects from a leaf's raw image, including shape, color, texture, and vein with the use of RF 73.38 %, SVM 67.27 %, ANN 65.68% KNN 63.20% respectively.

Mohamed, M et al [9], using CNN+BRBFNN+AI+ML, produced an accuracy is of 98.66%, 98.5%, and 98.73%, respectively. On a rice plant leaf dataset. Microcontrollers with cameras integrated into them were used in this application to track performance. Subsequent endeavors would delve far further. Annette B, G Geetharamani, and Arun Pandian J [10] have taken GAN,NST dataset and augmented it . Using SVM, K-NN, and CNN, their corresponding accuracy values are 98.66%, 98.5%, and 98.73%. Other image augmentation methods including autoencoders, random erasure was also used for this study. To identify and classify leaf diseases, A. Francis S [11] developed image-based leaf disease detection using DL. The evaluation states that CNN will identify during testing phase. According to the literature evaluation, the CNN is a helpful method for categorizing illnesses, and its accuracy rate is 91.37%.

Yiannis, A et al [11] employed machine learning and autonomous K-means clustering Using SVM, they achieved 98.97% accuracy. Random forest model performed well in admission prediction system developed by Dhresh K et al. in their paper [12]. In this study about individual explanatory features which can be used by the students to analyse and improve their performance are not provided [13]. ML gives the expertise to anticipate the features impacting the company based on traditional data[14] The deep CNNs offers promising prospects for efficient and accurate auto object identification in biodiversity research [15]. The algorithms comparison conducted in the paper primarily focuses on the results of validation testing to make decisions regarding the selection of algorithms for detection purposes and which is a unique[16].

Machine learning models have become extremely effective tools for managing and forecasting agricultural crop diseases such as tomato, corn, grape etc. [17].

III. METHODOLOGY

The following steps are employed and explained individually in this table:

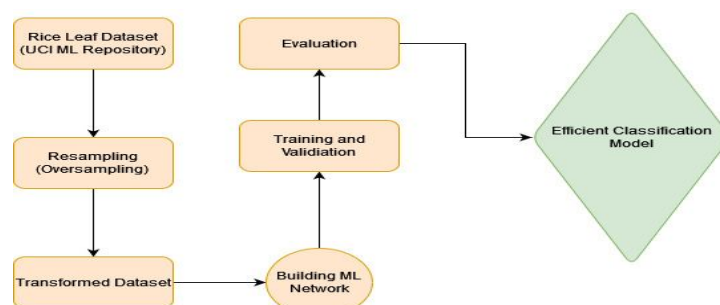


Figure 1. The flow diagram

A. Dataset

The dataset used from UCI ML repository contains rice disease leaves. The data was collected by doing the three repetitions of sustained phonations. The description of a dataset used is entailed in Table 1.

TABLE I. THE DESCRIPTION OF THE DATA

Number of instances	240
Bacterial leaf blight	40
Brown Spot	40
Leaf Smut	40
Healthy Control	120

B. Description of Algorithms

The brief description of the three algorithms namely KNN, Random Forest and Decision Tree is given below as follows.

Logistic Regression

One statistical technique for issues involving binary classification is logistic regression. In contrast to linear regression, which forecasts continuous values, logistic regression forecasts the likelihood that a given input falls into a specific class, usually 0 or 1. Any real-valued number may be mapped into a range between 0 and 1 using the sigmoid function, allowing for probability-based predictions. The model minimizes classification errors by determining the best-fitting coefficients for input characteristics. Because logistic regression can efficiently divide data points into groups by modifying these coefficients during training, it is frequently used for applications like sentiment analysis, spam identification, and illness diagnosis.

KNN

A supervised machine learning technique called k-NN is applied to both regression and classification tasks. The process of solving issues based on solutions for previously identified, comparable problems is known as instance-based learning. Lazy learners, which include instance-based learning techniques, do not compute the data until a query is submitted to the system. The various distance was used to classify the data. One hyper-parameter that varies from problem to problem is the value of k. Manhattan, Euclidean and Minkowski distance are the distance measures that are employed.

Random forest

A well-liked ensemble M L technique for classification is Random Forest. During training, a random subset of the data and characteristics are fed into each of the several decision trees that are constructed. Through majority voting for classification or average for regression, these trees' separate predictions are combined to produce the final output. The trees are guaranteed to be diverse due to the randomization created by sampling features and data, which lowers the possibility of overfitting. This improves how well the model generalizes to new data. Random Forest can handle big datasets with high-dimensional characteristics and is noise-resistant. It also offers feature importance scores, which aid in determining which variables have the greatest predictive power.

Despite its strength, the technique may have trouble processing very high-dimensional sparse data and can be computationally demanding. Nonetheless, because of its efficacy, scalability, and simplicity, it is still widely utilized.

V. RESULTS AND DISCUSSION

Table 3 displays the results of the evaluation of the accuracy metrics for each of the three machine learning models that were run separately.

TABLE: PERFORMANCE METRICS OF VARIOUS ALGORITHMS

S. No	Algorithm	Accuracy
1	Decision Tree	92.8
2	K-NN	96.2
3	Random Forest	98.6%

The three diseases were accurately predicted by Random Forest with a 98.6% accuracy rate. This demonstrates the random forest algorithm's dominance over the decision tree and k-NN algorithms, which demonstrated accuracy of 96.2 and 92.8, respectively. When accurately differentiated, the log loss graph is easy to understand

and provides a precise perspective for evaluating the model's performance. A model that performs well on the classification job is indicated by lower loss values and greater accuracy values. Our experiments' findings indicate that this model is noticeably superior to every other model that has lately been used in literature.

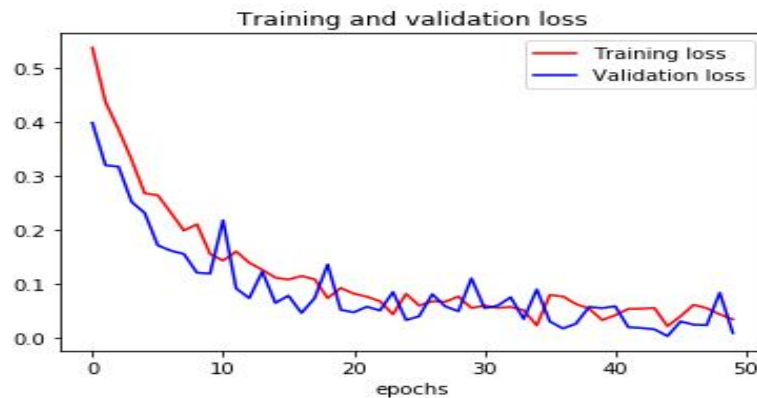


Figure 4. Graph plot showing loss with on Random Forest Algorithm

On the dataset used in our work, the outcomes of our suggested model were examined and contrasted with the most advanced methods. The findings demonstrate a notable improvement in the three leaf diseases' detection.

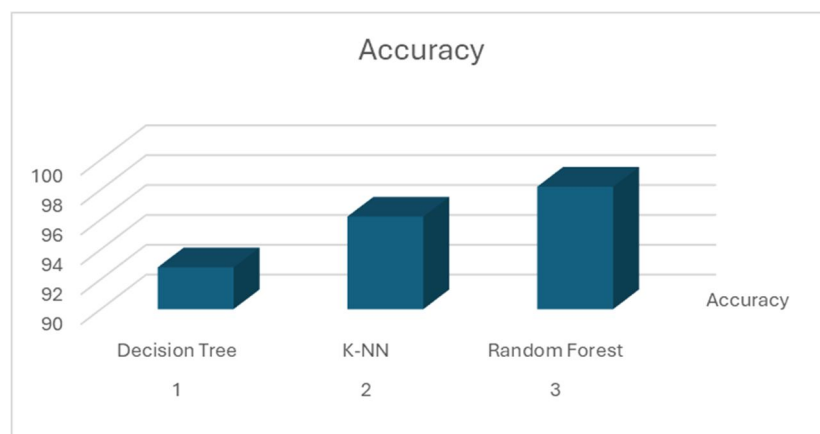


Figure 3: Classification Accuracy of various algorithms

REFERENCES

- [1] Ahmed, K., Shahidi, T. R., Alam, S. M. I., & Momen, S. (2019, December). Rice leaf disease detection using machine learning techniques. In 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI) (pp. 1-5). IEEE.
- [2] Tulshan, A. S., & Raul, N. (2019, July). Plant leaf disease detection using machine learning. In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.
- [3] Harakannavar, S. S., Rudagi, J. M., Puranikmath, V. I., Siddiqua, A., & Pramodhini, R. (2022). Plant leaf disease detection using computer vision and machine learning algorithms. *Global Transitions Proceedings*, 3(1), 305-310.
- [4] Umamageswari, A., Deepa, S., & Raja, K. (2022). An enhanced approach for leaf disease identification and classification using deep learning techniques. *Measurement: Sensors*, 24, 100568.
- [5] Sujatha, R., Chatterjee, J. M., Jhanjhi, N. Z., & Brohi, S. N. (2021). Performance of deep learning vs machine learning in plant leaf disease detection. *Microprocessors and Microsystems*, 80, 103615.
- [6] Jasim, M. A., & Al-Tuwaijari, J. M. (2020, April). Plant leaf diseases detection and classification using image processing and deep learning techniques. In 2020 International Conference on Computer Science and Software Engineering (CSASE) (pp. 259-265). IEEE.
- [7] Ganatra, N., & Patel, A. (2020). A multiclass plant leaf disease detection using image processing and machine learning techniques. *International Journal on Emerging Technologies*, 11(2), 1082-1086.
- [8] Chowdhury, M. E., Rahman, T., Khandakar, A., Ayari, M. A., Khan, A. U., Khan, M. S., ... & Ali, S. H. M. (2021). Automatic and reliable leaf disease detection using deep learning techniques. *AgriEngineering*, 3(2), 294-312.

- [9] Pandian, J. A., Geetharamani, G., & Annette, B. (2019, December). Data augmentation on plant leaf disease image dataset using image manipulation and deep learning techniques. In 2019 IEEE 9th international conference on advanced computing (IACC) (pp. 199-204). IEEE.
- [10] Amritavarshini, S., Ganesh, K. S., Arvind, M., Teja, K. R., & Vijayan, D. (2024, July). Image-Based Plant Leaf Disease Detection using Deep Learning. In 2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IconSCEPT) (pp. 1-6). IEEE.
- [11] Javidan, S. M., Banakar, A., Vakilian, K. A., & Ampatzidis, Y. (2023). Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning. *Smart Agricultural Technology*, 3, 100081.
- [12] Dhruvesh K, Rashmi P, Rushabh S , Sachin B (2019) Engineering college admission preferences based on student performance. *Int J Comput Appl Technol Res* 8(09):379–384. ISSN: 2319-8656
- [13] Shubham Khandale and Sachin Bhoite, "Campus Placement Analyzer: Using Supervised Machine Learning Algorithms", *IJCATR*, vol. 8, no. 09, 2019, ISSN 358-362.
- [14] A. Kulkarni, D. Bhandari and S. Bhoite, "Restaurants rating prediction using machine learning algorithms", *International Journal of Computer Applications Technology and Research*, 2019
- [15] T. Surabhi, B. Sachin and C. Advait, "Deep Convolutional Neural Networks for Automated Butterfly Species Recognition and Classification", 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 778-783, 2023.
- [16] Vishwakarma S., Bhoite S., "Classification of Corn Leaf Disease using Resnet18, Alexnet and VGG16" (ACT2024), 2, pp. 1333 – 1339
- [17] Amondkar M., Bhoite S. "Machine Learning Approach for Onion Leaf Disease Detection: A Case Study in Maharashtra, India"(ACT2024), 2, pp. 1249 – 125