ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR444

Utilizing Machine Learning Techniques forthe Detection of Plant Leaf Diseases

Enhancing Plant Leaf Identification using Machine Learning Techniques with the Random ForestAlgorithm

[1]Sri M.Chilaka Rao M.Tech (PhD)Assistant Professor
Department of Information Technology Sagi Rama
Krishnam Raju Engineering College
Bhimavaram, Andhra Pradesh, India

[2]Sharun Kumar; [3]Kishore; [4]Rakesh Kumar; [5]SravanKumar
Department of Information Technology Sagi Rama
Krishnam Raju Engineering College
Bhimavaram, Andhra Pradesh, India

Abstract:- Identification of plant diseases is crucial for preserving crops and ensuring food security. Analysis of detectable chemicals in plants is essential to understand transmission mechanisms and develop effective strategies for disease control measures to conserve agricultural products and prevent losses. However, manual monitoring of plant health is labor-intensive and time-consuming, requiring specialized skills and knowledge. To overcome these challenges, random forest systems are emerging as a powerful tool for disease detection and classification in plants. The process involves several steps, including image acquisition, preprocessing, and segmentation, followed by feature extraction, model training, and testing. Leveraging machine learning techniques, the random forest algorithm enables accurate classification of healthy and diseased leaves based on selected features. Image classification techniques are utilized to extract color information, while global features such as size and texture are captured through annotation. The dataset used for model training and testing comprises diverse samples, encompassing healthy and diseased plants. The random forest model is trained on 70% of the data to ensure robust learning, while the remaining 30% is reserved for testing, facilitating the exploration of model performance and overall feasibility.

I. INTRODUCTION

Agriculture plays a vital role in our country, with crop control being essential not only for cultivation but also for disease prevention. Crop diseases inflict extensive damage and pose threats to food safety, underscoring the necessity for routine plant health monitoring and preventive actions. This project aims to utilize the Random Forest algorithm, a branch of machine learning, for identifying diseased plants. Early disease detection holds paramount importance in agriculture as it impacts crop yield, quality, and nutritional value. Traditional disease detection methods are labor-intensive, prompting the adoption of machine learning for automated detection. Machine learning models analyzing plant leaf images can effectively differentiate between healthy and diseased leaves, contingent upon high-quality data and algorithm selection. Furthermore, machine learning aids in yield prediction and pest management, enhancing productivity and profitability. India's agricultural progress underscores the significance of innovation. Despite providing essential resources, plants encounter disease challenges necessitating early detection for sustainable development. Various methods, ranging from manual surveys to comprehensive assessments, have been employed. Effective management of common fungal infections in plants is imperative to prevent spoilage. Balanced pesticide application is crucial to avert environmental harm and economic losses. Hence, precise antibiotic utilization is vital for efficient disease control. In summary, early detection and proactive measures are imperative for ensuring plant health and fostering sustainable agriculture.

II. LITERATURE SURVEY

The literature evaluation outlines specific techniques and technologies used for plant sickness detection and category. Mondalet al. (2015) emphasize function extraction and binary photograph conversion for disease detection. Padol et al. (2016) cognizance on photo amassing and segmentation the use of Gaussian filtering and KMeans Clustering. Ranjan et al. (2015) make use of HSV coloration space and Artificial Neural Networks for function extraction and class. Tejoindhi et al. (2016) recommend a technique evaluating wholesome and diseased samples, while Revathi et al. (2017) use edge detection and segmentation for disorder identification. Tanvinmehera et al. (undated) address tomato disorder detection with thresholding and K-Means clustering. Poojapawer et al. (2018) deal with cucumber disorder detection employing GLCM for function extraction and unsupervised/supervised classification. Meunkaewjinda et al. (2014) introduce a Hybrid Intelligent System for ailment detection. Tripathi et al. (2016) discover system mastering strategies for ailment identity. Arivazhagan et al. (2013) observe shade transformation, thresholding, and SVM for category. Singh et al. (2020) talk scalable disease detection using pc vision andgift a dataset and fashions for category. Li et al. (2021) review deep mastering's position in plant disease reputation, emphasizing characteristic extraction and performance. Sandhu and Kaur (2019) underscore the need for automated disease detection systems

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR444

in agriculture, reviewing picture processing techniques. These research collectively illustrate the methodologies and technologies used in plant sickness detection. It uses various methodologies for plant ailment detection and class, along with feature extraction, segmentation, and machine mastering strategies. Researchers emphasize image preprocessing, binary conversion, Gaussian filtering, and clustering algorithms for segmentation, highlighting the importance of technological improvements in agricultural practices.

➢ *System Overview :*

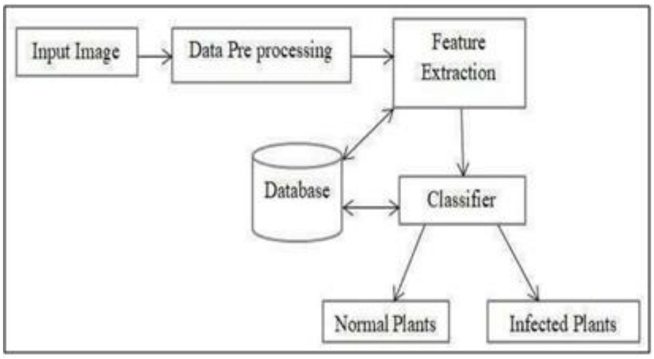

Fig 1 The above Figure Contains the Phases of Supervised Learning

The architecture diagram illustrates the essence of Supervised learning, a cornerstone in machine learning methodologies. In Supervised learning, algorithms are trained on classified datasets, where input data is matched with corresponding output data. The aim is to establish a mapping between input and output data, facilitating accurate predictions on unseen data. Typically, the classified dataset is divided into training and test sets, allowing algorithm parameter adjustment to minimize the disparity between predicted and actual outputs. The process begins with input provision, where images of leaves serve as input data, followed by data preprocessing, including tasks such as image format conversion and standard scaling. Subsequently, features are extracted from the images to assist in model training. Once trained, the model is tested on the generated test data, with images classified as healthy or diseased, and accuracy computed to assess model performance. This systematic approach underscores the effectiveness of Supervised learning in plant disease detection. detection.

III. METHODOLOGY

A. The Proposed Methodology has been Divided into the Followingsteps. Each Phase is Explained in Detail.

- *Extracting Data*
- *Preprocessing Data*
- *Splitting Data*
- *Processing Data*
- *Model Building*
- *Calculating Accuracy*
- *Disease Prediction*

➢ *Extracting Data:*

Dataset is collected from the Kaggle. Dataset consists of images of various plants leaves which are again categorized into healthy and diseased. There are two folders named healthy and diseased which contains respective category of images.

➢ *Preprocessing Data:*

Data preprocessing involves transformation of raw data to a format that data can easily be furtherprocessed. This can be easily understood by the following steps.

➢ *Splitting Data:*

The dataset is divided into 3 segments, namely, training, testing, and validation datasets. Training data is used to fit the model. Testing data is used to find the final accuracy. Training data contains both the categories healthy and diseased. The healthy folder contains around 800 images, and the diseased folder contains around 800 images.

➢ *Processing Data:*

In this step, one loop is run over the entire images in the two folders so that each image can be preprocessed individually. This step occurs after resizing of the images is done. This step occurs after data splitting. In this, the data is first convertedinto the RGB format from their normal form. Then the images are converted into BGR format and then finally brown and green color is extracted from the leaves. Now the global features are extracted fromthe images.

➢ *Model Building:*

After the splitting of the images and the pre-processing of the data is done, now the data from the preprocessing are used for training and the testing processes. But before that the model has to be created. We use SVC (Support Vector Classifier) to build the Support Vector Machine model to achieve forward goals. After building the SVM classifier and now it is trained with the traindata and it is tested against the test data and then the accuracy is calculated.

➢ *Calculating Accuracy:*

After all the computations done on the data, now the model is created and is tested with the test data. The accuracy is calculated as the number of predictions done correctly based on all the test data given to it. This gives the accuracy

➢ *Disease Prediction:*

After all the testing and the training is done, then the model is ready can be used to predict the disease in the plant leaf.

IV. RANDOM FOREST ALGORITHM

Random forest is a cluster learning algorithm used in machine learning for classification and regression tasks. It works by creating a number of decision trees during the training phase and resulting in different methods for classes (classification) or for prediction of individual trees (regression).

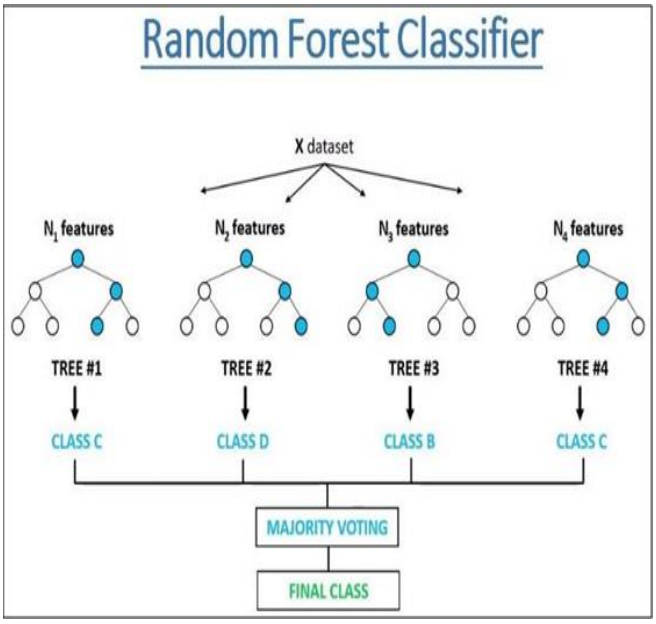

Fig 1 The above Figure is Random Forest Algorithm.

- *Random Subset Selection:*

Random Forest randomly selects subsets of the schooling records and builds choice trees independently on each subset. This process is referred to as bootstrap aggregating or bagging.

- *Random Feature Selection:*

At each node of the choice tree, as opposed to considering all capabilities to make a break up, Random Forest selects a random subset of functions. This enables in decorrelating the trees and guarantees that each tree makes its personal unique choices.

- *Voting or Averaging:*

Once all trees are constructed, Random Forestcombines their outputs to make a final prediction. For classification tasks, it takes a majority vote among the individual trees' predictions, while for regression tasks, it averages their predictions.

Random Forest is understood for its robustness, scalability, and capacity to deal with excessive-dimensional statistics with complicated relationships. It's much less liable to overfitting compared to man or woman choice bushes and often yields excessive accuracy. Additionally, it provides estimates of characteristic significance, which can be precious for knowledge the underlying records patterns. Due to its versatility and effectiveness, Random Forest is broadly used throughout various domains, including healthcare, finance, and ecology.

V. CONCLUSION

In this study, we performed experiments on the leaves of different plants to determine their health status, and differentiated between diseased and healthy samples. To this end, we developed seven different models and carefully evaluated their performance, finally selecting the most efficient model. Throughout the project, we carefully performed several data preprocessing steps, as outlined in the methodology. Notably, the random forest model emerged as the best alternative, providing excellent accuracy and precision compared to its counterparts. Looking ahead, there are many avenues for future exploration. Given the focus of our work on plant health classification, expanding it to include plant species and associated diseases could increase its utility. In this extension, multiple classifiers, plant species diversity, and associated disease data can be useful for training and thorough testing by using models optimized for such classification tasks is aimed at continuously improving the diagnostic accuracy and applicability of our method.

FUTURE WORK

Expanding to Multiclass Classification: While the cutting-edge recognition of our assignment is on predicting whether or not a plant is diseased or healthful, there exists ability for extension to embody a couple of diseases across numerous plant species. This expansion could involve compiling datasets comprising numerous plants and their corresponding illnesses for comprehensive schooling and testing. Utilizing a version optimized for multiclass type, we goal todecorate accuracy and expand the applicability of our classification system.

REFERENCES

[1]. Dhiman Mondal, Dipak Kumar Kole, Aruna Chakraborty, D. Dutta Majumder" Detection and Classification Technique of Yellow Vein Mosaic Virus Disease in Okra Leaf Imagesusing Leaf Vein Extraction and Naive Bayesian Classifier., 2015, International Conference on Soft Computing Techniques and Implementations- (ICSCTI) Department of ECE, FET, MRIU, Faridabad, India, Oct 8-10, 2015.

[2]. Pranjali B. Padol, Prof. AnjilA.Yadav, "SVM Classifier Based Grape Leaf Disease Detection" 2016 Conference on Advances in Signal Processing(CAPS) Cummins college of Engineering for Women, Pune. June 9-11, 2016.

[3]. Malvika Ranjan, Manasi Rajiv Weginwar, NehaJoshi, Prof.A.B. Ingole, “detection and classification of leaf disease using artificial neural network”, International Journal of Technical Research and Applications, 2015.

[4]. Tejoindhi M.R, Nanjesh B.R, JagadeeshGujanuru Math, AshwinGeetD'sa "Plant Disease Analysis Using Histogram Matching Based On Bhattacharya's Distance Calculation" International Conference on Electrical, Electroniocs and Optimization Techniques(ICEEOT)- 2016

[5]. P.Revathi, M.Hemalatha, “Advance Computing Enrichment Evaluation of Cotton Leaf Spot Disease Detection U sing Image Edgedetection”, ICCCNT'12.

[6]. Tanvimehera, vinaykumar,pragyagupta "Maturity and disease detection in tomato using computer vision" 2016 Fourth international conference on parallel, distributed and grid computing(PDGC)

[7]. Ms.Poojapawer ,Dr.varshaTukar, prof.parvinpatil "Cucumber Disease detection using artificial neural network"

[8]. A. Meunkaewjinda, P. Kumasawat, K. Attakitmong col and A. Srikew on Grape Leaf Disease Detection (2014) from Color Imagery using Hybrid Intelligent System Vegetable and fruits are the most important export agricultural products of Thailand.

[9]. Mukesh Kumar Tripathi, Dr.Dhananjay, D.Maktedar" Recent Machine Learning Based Approaches for Disease Detection and Classification of Agricultural Products" International Conference on Electrical, Electronics and Optimization Techniques (ICEEOT)- 2016.

[10]. S.Arivazhagan, R. Newlin Shebiah, S.Ananthi, S.Vishnu Varthini. (2013) Detection of Unhealthy Region of Plant Leaves and Classification of Plant Leaf Diseases using Texture Features. ‖ AgricEng Int:CIGR Journal, 15(1): 211－217.

[11]. D Singh, N Jain, P Jain, P Kayal, S Kumawat… - Proceedings of the 7th …, 2020 - dl.acm.org PlantDoc: A Dataset for Visual Plant Disease Detection. PlantDoc | Proceedings of the 7th ACM IKDD CoDS and 25th COMAD.

[12]. L Li, S Zhang, B Wang - IEEE Access, 2021 - ieeexplore.ieee.org Plant Disease Detection and Classification by Deep Learning—A Review DOI: 10.1109/ACCESS.2021.3069646