IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

DISEASE DETECTION AND ANALYSIS OF VARIOUS LEAVES USING IMAGE PREPROCESSING TECHNIQUES

Annepu subhash[1], Appanna Sunil[2], Gogada Prem Kumar[3], Kona Vineela[4], Doddi Deepak[5]
Department of Computer Science And Engineering, Visakha institute Of Engineering & Technology, Narva, Visakhapatnam, Andhra Pradesh, India

Abstract. Agriculture plays a vital role in Indian economy, but owing to changing climatic conditions, crops often get affected a result agricultural yield is decreases drastically. If the condition gets worse, crops may become vulnerable to infections caused by agents. The method that can be adopted to prevent plant loss is real-time identification of plant diseases. Our proposed model provides an automatic method to determine leaf disease in a plant using a trained data set of pomegranate leaf images. The test set is used to check whether an image entered into the system contains disease or not. If not, it is considered to be healthy; otherwise, the disease of that leaf is predicted, and the prevention of plant disease is automatically proposed. Further, the rodent causing the disease is also identified with image analysis performed as certified by biologists and scientists. the accuracy of disease detection of leaf using image processing techniques can vary depending on the complexity of the disease, the data set used for training, and the environmental conditions

Keywords: Digital image processing, Foreground detection, Machine learning, Plant disease detection.

1 Introduction:

The research paper referenced in tells about the various techniques involved in detecting disease in a plant leaf through image processing techniques which involves image acquisition, per-processing of image and segmentation steps. The paper discussed methods to determine the health status of each plant by considering the requirement of that plant. This method was used so that the chemicals are applied only to those plants which require this treatment. This method will result in reducing the expense on plant health to a larger scale. aims at using real time applications to change captured RGB image to Grey scale images since it increases the clarity of the images and the disease is detected more efficiently. However, the paper discusses how diseases in a pomegranate plant can be detected by applying proper segmentation methods on the extracted images and identify the condition of a Pomegranate plant

Visual symptoms in the leaves are used to detect diseases in leaves. There is need for an automatic system used for leaf disease detection because the disease cannot be detected by naked eye. Our technology has grown to such an extent that a machine is capable enough to predict disease by looking at a high-definition image of that leaves at its early stage itself. The objective of our research is to find out the disease in a leaf. This process of detection

can be performed using image processing techniques which is method of forming a signal processing for an inputted image by a scientist. We have used the R programming language for implementing image processing of the diseased leaf and predicting the disease. The data set used in our experiment contains images of infected and healthy pomegranate leaves. Model Setup In order to conduct our experimentation, the data set of plant leaf diseases is taken from which is composed of 5358420 images of a single leaf disease. All the photographs present on this site are clicked by professional photographers to provide easy access to educational applications

2 Literature Review

1. Jayashri Patil Sachin Naik et.al, In this paper the proposed model purposes preprocessing, division, Feature extraction and characterization methods for pomegranate organic product illnesses. During feature extraction colour, texture and morphology features are used to identify and classify pomegranate fruit diseases. SVM, ANN, KNN, PNN classifier used to detected fungal and viral diseases. For image segmentation K-means clustering is used Fuzzy c means gives highest accuracy. Future scope is developing fully automated system with collaboration of agriculture universities and research centre for upgrading the system with new diseases.

2. Shivaputra S.Panchal, Rutuja Sonar et.al, In this paper the proposed model is disease detection using image acquisition, image preprocessing, image segmentation, statistical feature extraction and Classification K-mean clustering algorithm is used for segmentation. Support vector machine is used for classification of disease Image processing is a form of signal processing the methodology of the proposed work contains the five stages.
    - Image acquisition to get data set of disease and non-diseased images.
    - Enhancement enhances the contrast of images
    - K-mean is used for classification of object based on a set of features into k number of classes.
    - Feature extraction is utilized.

3. M. Bhange et.al, In this paper the proposed model gives an online apparatus has been created to recognize natural product infections by transferring organic product picture to the framework. Highlight extraction has been finished utilizing boundaries like tone, morphology and CCV (colour cognizance vector). Bunching has been finished utilizing the k-implies calculation SVM is utilized for characterization as contaminated or non-talented.

4. J.D. Pujari et.al, In this paper the proposed model has taken number of harvest types in particular, natural product, vegetable, oat and business yields to distinguish contagious sickness on plant leaves. Various strategies have been for natural product crop, k- means grouping is the division technique utilized, surface elements have been centred around and ordered utilizing ANN and nearest neighbour algorithms.

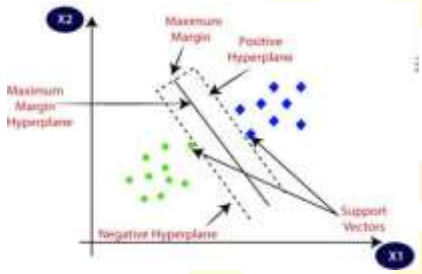
METHODOLODY (ALOGORITHM):

SUPPORT VECTOR MACHINE(SVM) 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.

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 hyper plane.

SVM chooses the extreme points/vectors that help in creating the hyper plane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyper plane:

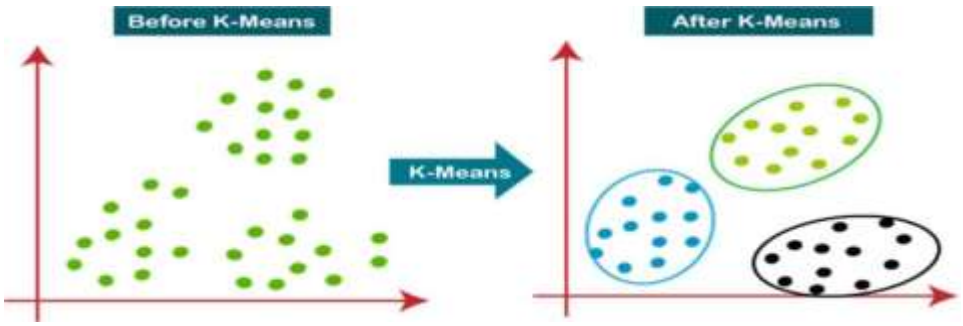


Support Vector Machine

K-MEANS CLUSTERING ALGORITHM:

K-Means clustering is an unaided learning calculation that is utilized to tackle the grouping issues in AI or information science. In this theme, we will realize what is K-implies clustering algorithm, how the calculation works, alongside the Python execution of k-implies grouping. K-means algorithm K-Means clustering is a Solo Learning calculation, which bunches the unlabelled data set into various groups Here K characterizes the quantity of per-characterize

bunches that should be made all the while, as though K=2, there will be two groups, and for K=3, there will be three groups, etc.



Before and After k-means

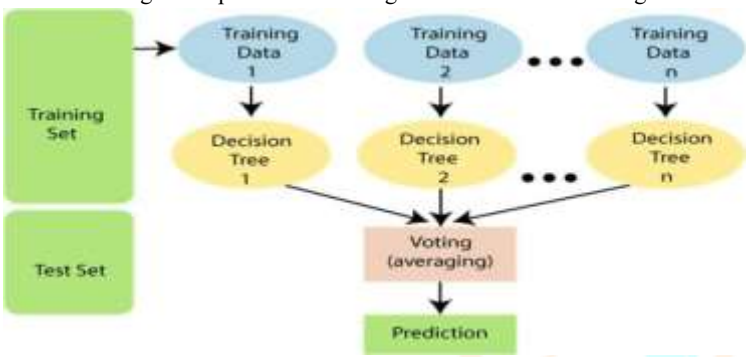
RANDOM FOREST ALGORITHM:

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 to improve the performance of the model.

Random Forest is a classifier that contains a number of decision trees on various subsets of the given data set and takes the average to improve the predictive accuracy of that data set. "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 it predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of over fitting.

The below diagram explains the working of the Random Forest algorithm:



RANDOM FOREST ALGORITHM:

Data Set

For this project we have used public data set for plant leaf disease detection called Plant Village curated by Sharada P. Mohanty et Al. [6]. The data set consists of 87000 RGB images of healthy and unhealthy plant leaves having 38 classes out of which We have selected only 25 classes for experimentation of our algorithm These classes are shown in Table 1.

Table 1. Dataset Specifications.

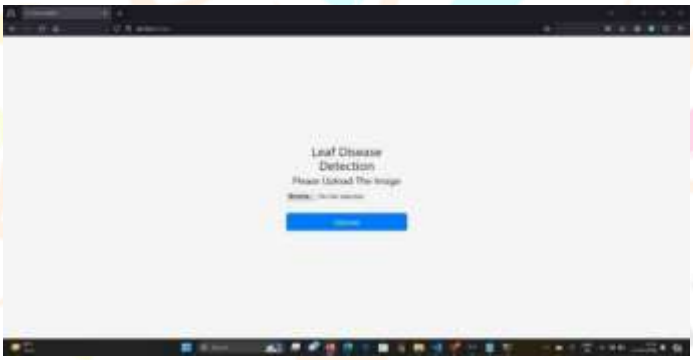
Fig-1: Sample images in the data set.

Data preprocessing and Feature Extraction:

Data preprocessing and feature extraction are fundamental stages within the machine learning pipeline that groom raw data for model consumption. Data preprocessing entails meticulously cleaning and manipulating the data to guarantee its integrity. This encompasses addressing missing values, outliers, and inconsistencies. It can also involve transforming the data into a format that facilitates analysis, such as scaling numerical features or encoding categorical features. Feature extraction, the subsequent stage, focuses on crafting new features from the available data. These novel features are meticulously designed to be more informative and relevant to the specific machine learning task at hand. There are numerous techniques for feature extraction, including dimensionality reduction and feature selection. By meticulously following these crucial steps, data scientists can significantly enhance the quality of their data, fostering the development of a more powerful and effective machine learning model.

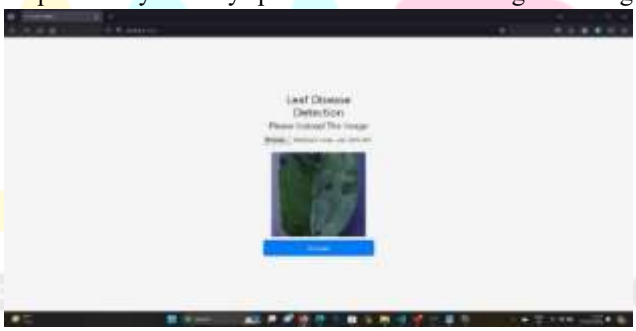
Fig-2: Steps for data preprocessing and feature extraction.

Results:



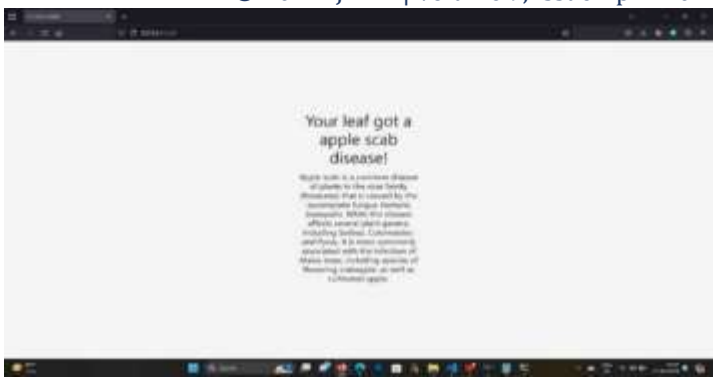
The image depicts a user interface for a web-based application likely used in the agricultural domain. The central message on the screen is "Leaf Disease Detection" which is accompanied by an instruction to "Please Upload The Image". A file selection menu is presented along with an "Upload" button, indicating the functionality to initiate an image processing task.

It can be inferred that this application is designed for the analysis of digital images containing leaves. The purpose is likely to automatically detect the presence and potentially identify specific diseases affecting the foliage. This type of technology can be



beneficial in precision agriculture practices for early detection and treatment of plant illnesses, potentially improving crop yields and overall farm efficiency.

1. The visual interface presented depicts a web-based application designed for the examination of plant leaves.
2. The central focus is on "Leaf Disease Detection,"
3. with clear instructions to "Please Upload the Image."
4. A file selection menu is available, allowing the user to designate the image containing the leaf sample for analysis.
5. An "Upload" button initiates the image processing task.



outputs generated by system.

The webpage provides confirmation of the analysis outcome by displaying a message such as "Your leaf has apple scab disease!" Alongside this result, there is additional information regarding apple scab disease.

shows Table 1 the performance matrices for each model developed for each of the plant. We can observe that the accuracy scores are nearly equal to f1 scores. This is because of balanced number of false negative and false positive predictions. This is considered as best case for any machine learning algorithm. The average accuracy was 98.92%.

Table 1. Performance matric for all models.

Plant	Accuracy	F1 Score
Apple Corn Grapes Potato	0.91 0.94 0.95 0.98	0.91 0.94 0.95 0.98
Tomato	0.87	0.87

Conclusion:

We have successfully developed a computer vision-based system for plant disease detection with average 98.92% accuracy and 0.98 F1 score. also the proposed system is computationally efficient because of the use of statistical image processing and machine learning models.

References:

[1] Nikita Goel, Dhruv Jain and Adwitiya Sinha, “Prediction Model for Automated Leaf Disease Detection & Analysis”, IEEE 2018.

[2] Shivaputra S.Panchal, Rutuja Sonar, “Pomegranate Leaf Disease Detection Using Support Vector Machine”, International Journal Of Engineering And Computer Science ISSN: 2319- 7242 Volume 5 Issues 6 June 2016, Page No. 16815-16818.

[3] J Shirahatti, R Patil and P Akulwar, “A survey paper on plant disease identification using machine learning approach”, IEEE 2018.

[4] Jayashri Patil, Sachin Naik, “Pomegranate fruit diseases detection using image processing techniques:” A Review", IT in Industry, Vol. 9, No.2, 2021.

[5] SD Khirade and AB Patil, “Plant disease detection using image processing”, IEEE 2015.

[6] Pujari, J.D., Yakkundimath, R., Byadgi, A.S., “Image Processing Based Detection of Fungal Diseases In Plants”, International Conference on Information and Communication Technologies, Volume 46, pp. 1802-1808, 2015.

[7] K. Sakthidasan Sankaran,N. Vasudevan and V. Nagarajan, “Plant Disease Detection and Recognition using K means Clustering”, IEEE 2020.

[8] G. Shrestha, Deepsikha, M. Das and N. Dey, “Plant Disease Detection Using CNN”, 2020 IEEE Applied Signal Processing Conference (ASPCON), 2020.