# Machine Learning Techniques for Leaf Analysis in Plant Disease Detection

Dr. V Venkata Ramana[1,] Dr. V Lokeswara Reddy[2] Dr K Sreenivasa Rao[3], M Ramanjeneya Reddy[4]

[1,2,3, 4]Professor, Department of CSE, K.S.R.M College of Engineering(A), Kadapa

## *Abstract*

***A country's productive innovation is tied to its agriculture sector. Agriculture, the "mother" of all human societies, is the source of food and building materials. We rely heavily on it as a food source, thus its importance cannot be overstated. As a result, plant diseases pose a serious challenge to society. Pests and diseases affecting plants might occur at any moment. It's possible for this to occur between planting and harvesting. The market's economic worth has dropped significantly as a result. Consequently, the ability to identify leaf diseases is crucial in the agricultural industry. Thus, conventional approaches were utilized for illness detection. However, conventional leaf disease diagnosis relies only on the trained eyes of agriculturalists and plant pathologists. Using this approach for disease detection in plant leaves was labor intensive, time consuming, expensive, and needed a deep understanding of plant pathogens. Software solutions that have been evaluated experimentally can now automatically identify and categorize plant leaf diseases. The process of new evolution is aided by machine learning. Plant diseases are being detected using machine learning. Machine learning is a branch of AI used to train computers to carry out certain tasks without human intervention. Machine learning's primary objective is to decipher and incorporate the training data into models that will be of service to humans. Therefore, plant illness may be identified using machine learning.***

## INTRODUCTION

India as a progressing nation and farming is main things to develop our nation. Agriculture played a big role in developing human civilization. Plants were only used for humans and animals feeding. But now, agriculture has become very important than past. People need to research in production techniques. Increasing product activity, use fewer pests, shorten effect on environment. Some of the techniques are used. The goal is develop plane land and growing crops, increase foods and profitable systems. Opportunities of employment are also given. Agriculture is the main source of income and gives the raw materials to the food industry.

The importance of plant is that it gives us food, shield, materials for building, Ayurveda medicines, fuels, woods etc. Environment can be protected by the plants from problems the like soil erosion caused by floods, rain. The air pollution is also minimum. By balancing the intake of CO2 and release of O2. Plants give foods; also it provides the animal's home. Cultivation of plants is important as it gives us the different types of fruits, different types of vegetables, grain, and different types of nuts and medicines. The wood can also be used for the construction purpose, furniture and also in making of the paper. The production of bio-fuels can be done by the decaying of plant as it forms the fertilizer. It is also used to for generating the electricity. The agricultural fields deal with difficulty

which includes the big losses in cultivation of crops. Discretion in agriculture things, the full economy will be affected. Plant diseases became a dilemma because it became a big question in both the quality and quantity of agricultural product. In this environment there are lots of diseases. It controls the big amount of wastage. According to these reasons, it is good to detect these diseases affectively and timely to detect the losses caused by it. Planting, Cultivating, and preventing the plants is important for the effective development of any country or region. Diseases in plants can be checked through many different methods which includes many man base and technology base checking. Some of the problem in plants is visible on human eyes. There are diseases which are known at the later stage of the leaves which would have already caused the damage to the leaves to the great extent. The plant diseases like pathogen, microorganisms which are living, bacterial problem, fungi infected plant, nematodes, viruses affected problems in plants. All though recognizing problem in plant is important to first stage protection. Diseases are identifying by sign. We should use that process which is maintained by computerized systems process. It is an only option to detect. People can use some kind of algorithm and analyzed the different type's tools. Which can be done by the machines, powerful microscopes? There are some of the signs that recognized by electromagnetic ranges. As it can give various and many kind of pictures which cannot be seen through the naked eyes. Other technique is Remote Sensing Technique (RST). It inspect thoroughly using different spectral. Remote Sensing Technique, normally taken by different tool which is usually used by digital ways. Maximum plants are infected by fungal infection. Fungal infection comes as dots in leaves. The marks wrapped the total area of the leave. This type of diseases can decline different plants and later happens to end such kind of plant variety. Leaves diseases happened using defoliant. Maximum amount of insecticide are harmful. Also, it has bad effect of the environment and also to the mankind and human. Applying insect killer, different forms of diseases are affected the plants. Maximum apply of pest killer by human increases the costing. It can be big loss. So that reasons they have to use it. One another way to solving this problem are to calculate the affected area and the amount of pest we can use. The appropriate amount of quantity and concentration of pesticides can solve this problem. The way to solve the problem is naked-eye observation. GC can improve the accuracy. But it takes time. Other way to detection is processing an image in different steps. It can be developed the agricultural sector. Another plant leave diseases is fungal infection problem. It gives marks on leaves. We have to take care of it, it became big loss. Other parts of the world, farmer face lots of problems. It affects farming. Anthracnose in lesions, Alter aria in lesions on watermelon. Extreme using insect killer in plants increases the costing. It affected nature. To solve this problem we have to apply perfect amount of pesticides. So we can use machine learning. It will help to identify the diseases in early stages. So it will use for disease detection mechanisms.

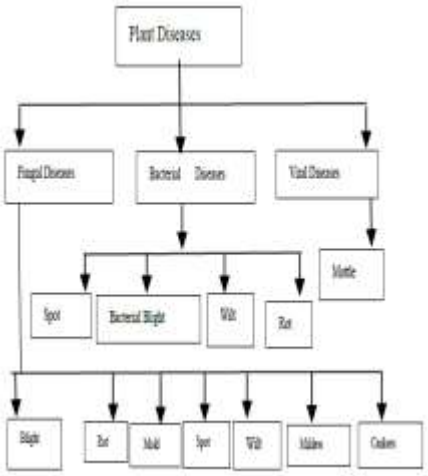


Fig. 1. Classification of Plant Diseases

Plants diseases can be like fungal disease, bacterial disease and viral disease as shown in fig. 1. Fungal diseases are of various types such as blight, rot, mold, spot, wilt, mildew, cankers. Bacterial diseases are of different kind like spot, bacterial blight, wilt, and rot. Viral diseases are such as mottle, etc.

## LITERATURE SURVEY

Sheryl et.al. (2019) proposed that much disease that is caused by the different types of bacteria or fungus in plant leaves can be detected by the prediction algorithm used in the machine learning. The classification algorithm is hard to classify the disease as it does not give the correct accuracy which is mainly reason of the different input data to algorithm. Different researcher tries different algorithms. And different processes give different results. And this way we get to know that plants can be affected by different viral infection problems, fungal signs and bacterial infection. Morphological feature is categorized and automated by the classification technique. We can work on mulberry plats to identify the diseases using CNN technique. Budiarianto et.al. (2018) proposes the different kind of the machine learning techniques. The human gets the carbohydrate from the corn. The corn plant also gets attack which affects the growth of the plant. The abnormal plant cell is increasing in the area of affected. The color is changed and the plant becomes wither and turn into stunted. The researcher focuses on improving plant disease using CNN techniques.

Researcher use different algorithms and compared the results in between. Researcher use support vector machines (SVM), use Decision Tree (DT), use Random Forest (RF), and use Naive Bays (NB). Normally empty eyes can mark the diseases using their color difference. If the leaves are change the color then human can recognized the problems. If the fruits are tested like rotten then it spotted as unhealthy material. If the leave have yellow or brown

marks then it is also a problem. Exanimation many different part of image clarification base on the feature of leaves color is also a processes. For dataset attribute use different parameters.

Parkas et.al. (2015) proposed that the plant disease also can be detected and classified by the chemical test. Speed in the detection and accuracy are characteristics of the detection using the machine learning. Human movement is very slow, so author use some faster technique. When the disease is spread over a large area the technique which is proved as the useful technique is farmers and segmentation of image is done by the k-means processes. Later researcher remove the unused part. Then count the texture attribute for effected parts. The parts are going through the next part to pre training processes. For this they use network of neurons. For recognizing the plant disease several image processing techniques are used on different plant species. They use K-means processes, GLCM technique, BPNN processes. But they face some problem like background noise. In the real world condition it is a very problematic situation.

## PROPOSED METHODOLOGY

### A. DATASET

In this paper plant leave disease recognition processes an appropriate and perfect dataset is required. First step of the proposed system is train the system. After data is given to the system, evaluate the performance. Machine learning algorithm dataset is required. Basically as a training data healthy and unhealthy leaves are required. For this training data we take commercial farming plants and specially we take tomato and potato leaves. A total of 1000 images are collected from different places for train system. And 2000 images are collected for test the system operation. For data set we search from internet. For more accurate data we go to some farming place and take some good and clear digital photo. For that we take digital camera. And for more data set search for Google website. Some of the websites also provide some farming plants. After collection, next processes are cleaning those data. Cleaning those data mean, take the important data image. Using python script compared the image. It is possible that data can be repeatable (Ganesh Babu Et.al. 2020). So using some attribute like name of the leave or size of the image

Data set can be erased (Parthia et.al. 2020). In the figure 2.1 sown some healthy leaves. And the other hand 2.2 figure shown diseases plant leave.

Fig.2.1: Healthy leaves

The different deceases are detected in this paper, like (1) Tomato Yellow Leaf Curl Virus, (2) Tomato Mosaic Virus, (3) Tomato leave Target marks-Corynispora casserole, (4) Tomato Two marks Spider Mite-Tetrannychus Utica, (5) Tomato Sartorial Leave marks-Sartorial lycopersici, (6) Tomato Leaf Mold- Passalora Fulda, (7) Tomato Bacteria marks -Xanthomoonas campestral pvVesicatoria, (8) Tomato Early Blights -Alter aria solaani, (9) Tomato leave Late Blight, Phytophthoa infections. (10) Potato Late Blight, Phytophthora infesting, (11) Potato Early Blight, Alter aria solace, etc.

Fig 2.2: Unhealthy leaves

**B. AUGMENTATION**

In this paper after dataset collection, next procedure is feature extraction. For fracture extraction on the above data convolutional neural network (CNN) used. Convolutional neural network inspired by biological inspect. It is one of the deep learning techniques. For feature extraction on the above data set done using a supervised learning technique convolutional neural network (CNN). For this we take fundamental considerable information. Large amount of data can give large and accurate amount of feature attribute in CNN. All the data set are divided into their different data category. Take a data image and rotate the image as 90,180,270 respectively in mirroring each rotated images. Expanding the dataset using CNN. The process of CNN is different in different images. It is divided into 3 different convocational layers. Each layer has in between max-pooling layers. First take an image as an input. Give the input to the first convocational layer and filters the input image with 32 kernels of size 3x3. And give the output of first convocation layer to the max pooling layer as an output. After first max pooling layer give the output to the second convocation layer as an input data. In the second convolutional layer image will be filter with 64 kernels of size 4x4.Give this output to the second max pooling layer as an input data. After that max pooling layer give the output to the last convolutional layer. It filters the data as 128 kernels of size 1x1. And give the output as fully connected 512 neurons layers. The output given to soft max function. The soft max function produces a probability distribution of the four output classes. The final layer is connected to MLP. All convocation layer output has the activated Rely function. And it fully connected to the layers. The system is trained using Adam. The batch size is size of 100 for 1000 epochs. In that way, we collect the features of the image dataset using the CNN algorithm.

**C. CLASSIFICATION PRINCIPAL**

In the data analysis and database management technique clustering is one of the data structure management techniques. Lots of data can be sub divided into subgroup. Same type of data can be placed in same group. Using this method we can define task of identification. Basically find homogeneous subpart inside the data point. Euclidean-based distance or correlation-based distance is use to identify this methods. It is an application-specification. Base on the features sub grouping clustering analysis is used. In this paper clustering is used in disease identification, where finding plant leave disease deferential is very important. Take different defected plant leave as a training data set. And try to find features in the training data images. Find features that are similar to each other in terms of image attributes, sample spot infected plant leaves, image data samples.

## EXPERIMENTAL RESULTS AND DISCUSSION:

In this section performance of plant leave diseases detection is calculated. The overall performance is calculated as how much time the system is detected the diseases as correct. In this way performance is calculated. For this experiment Anaconda navigator software system is used. Instanced of mat lab using Jupiter notebook is more efficient. Because the Jupiter notebook give more python access. In the Anaconda navigator software application is based on tens. Jupiter notebook 6.0.2. For evaluating the system, we determined that leaves (mainly tomato leaves) are affected from Bacterial, Fungal, Viral, and Fungus diseases. The proposed methodology evaluated the leaves base on their diseases. The proposed system asks for the particular leave image. Base on the image it will show that

the leave is affected any kind of diseases. If the leave is healthy then it shows healthy and if the leave is unhealthy then it shows the particular diseases. And after that it gives the remedies of the diseases. After the proposed system total loss will be 0.02303. Using multilevel k means and SVM algorithm efficiency of performance will be calculated. The efficiency tested on number of disease affected plant leaves and result will be concluded. After clustering if the data will go through the SVM multi-level classification will be better result. 3500 data image are used for train the proposed system. Test accuracy with best parameter set is 0.9940.

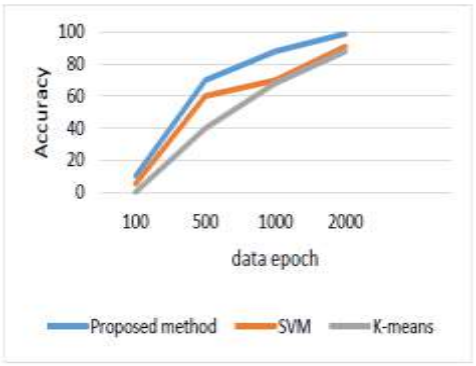


Fig: 4 Experimental Results and Discussion

In this fig 4 we derived that individual algorithm like k-means process give efficiency 88.6% and SVM give efficiency 91%. But in the proposed methodology it gives better result. The performance analysis is 99%. The accuracy is better than individual algorithm performance.

## CONCLUSION:

The recommended strategy is a valuable one that has the potential to improve performance. The K-mean technique performed poorly in both global clusters and clusters with varying data sizes and densities. As a result, if we use numerous SVM classes after clustering, we get more accurate results. This hybrid method outperforms both of the separate algorithms in the comparisons. This strategy has been shown to be effective in training and testing large datasets for illness prediction. This kind of thinking is quite helpful in modern life. Since the suggested technology allows for the application of just trace amounts of pesticides to crops, it will be of great service in the agricultural and medical industries. Improved segmentation methods for the algorithm might be a future project. Therefore, there is room for development in the methods.

**REFERENCES**

*1. Sheryl Pasha Annabel, T Annapoorani, and P Deepalakshmi. Machine learning for plant leaf disease detection and classification–a review. In 2019 (ICCSP), pages 0538– 0542. IEEE, 2019.*

*2. Shima Ramesh Maniyath, PV Vinod, M Niveditha, R Pooja, N Shashank, Ramachandra Hebbar. Plant disease detection using machine learning. In 2018 I (ICDI3C), pages 41–45. IEEE, 2018.*

*3. G. Prem Rishi Kranth, M. Hema Lalitha, LaharikaBasava, Anjali Mathur. Plant Disease prediction using Machine learning algorithms. In 2018 IJCA (09758887), volume 182-No. 25. IEEE, 2018.*

*4. Bin Liu, Yun Zhang, DongJian He, and Yuxiang Li. Identification of apple leaf diseases based on deep convolutional neural networks. Symmetry, 10(1):11, 2017.*

*5. Prakash M Mainkar, ShreekantGhorpade, and Mayur Adawadkar. Plant leaf disease detection and classification using image processing techniques. International Journal of Innovative and Emerging Research in Engineering, 2(4):139–144, 2015.*

*6. DS Gaikwad and KJ Karande. Image processing approach for grading and identification of diseases on pomegranate fruit: An overview. IJCSIT) International Journal of Computer Science and Information Technologies, 7(2):519–522, 2016.*