

Article

Plant Leaf Disease Detection using Data Mining

Manthan Mirgal

Research Scholar, MCA, Thakur Institute of Management Studies, Career Development & Research (TIMSCDR), Mumbai, India.

I N F O

E-mail Id:

mirgalmanthan98@gmail.com

Orcid Id:**How to cite this article:**

Mirgal M. Plant Leaf Disease Detection using Data Mining. *J Adv Res Cloud Comp Virtu Web Appl* 2020; 3(1): 25-28.

Date of Submission: 2020-05-14

Date of Acceptance: 2020-05-27

A B S T R A C T

Often farming can create great losses due to spread of disease in the crops yield. Many times farmer is unaware of the disease growing in the crops. This research aims on detecting diseases by scanning leaf of the crops and classifying diseases, that crop may have by identifying the patterns by Data mining techniques like K-mean, SVM, image recognition and KNN. These techniques can surely help grow economy of agriculture in some or other way.

Keywords: K-mean, SVM, image recognition, KNN

Introduction

Data mining methods collect large data and use it to recognize patterns in it. Many techniques have been discovered in the field of agriculture to improve quality of crops and avoid uncertainties cause through diseases in crops. Most common techniques include image recognition and combining it with classification and clustering with neural networks. In this research, I focus on identification of various diseases in crops through early detection. Because of the high cost of chemical control, the discovery and difference of numerous illnesses at early stage allow more efficiency in the field of agriculture. Data mining, the technique of dispensation and removal of useful data from large data set can help machine to learn on its own to develop knowledge from previously detected diseases and patterns. The following methods of data mining, neural networks, image recognition can be applied either with traditional image processing by taking in images in jpeg, Portable Network Graphic formats.¹ However, this approach requires high compression of images taken. Rather than these tradition methods, one can instrument this in smart phones using its camera. This will be a dynamic approach and increase complexity but the results will be very useful and appealing. The next section will show you the survey about what is people's view on Plant diseases and will show their experience regarding the same.

Literature Survey

In the existing environment, methods have been used to classify data. Machine learning methods such as ANN, Decision Trees are tried in agricultural field.

One big flaw in these methods is that it needs high compression of image of the leaf. This compression can compromise the quality of image and can make the image blur.² Due to this classification cannot be applied efficiently and the result may not be accurate.

The main objective of the research is to detect image dynamically with the help of smart phone's camera, which has huge potential. To use advance techniques of Data mining to make result more accurate and to implement Machine learning so that the system can develop the knowledge of plant diseases by itself and to improve ability to identify it.

Survey

The following surveys give the present condition in agriculture and views of people regarding plant disease.

- Familiarity of people with know-hows used in current agriculture or for nurturing a plant

The pie chart shows how many percentages of people in survey are familiar with current technologies present in agricultural environment.

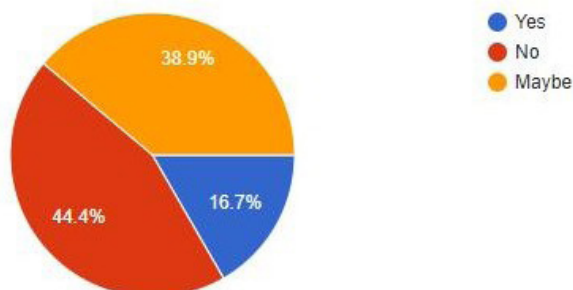


Figure 1. Reply of people when asked about familiarity with technologies in agriculture

The data above shows that 44.4% of the people are not even familiar or even heard of any skills used in agriculture. So there's a great need for promotion of skill for agriculture in order to increase quality of crops or increase a plant's life expectancy.

- Usage of technology in agriculture

When people were asked whether they have used technology with respect to agriculture then their responses were quite unexpected. Though technology has progressed drastically in today's world, it has not touched the field of agriculture in a great way. The subsequent chart shows the responses of the people when asked whether they have used any modern technology for agriculture.

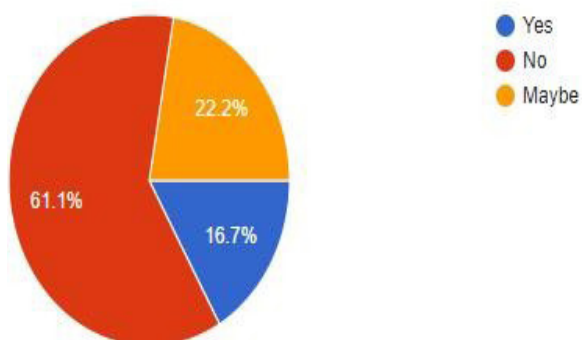


Figure 2. People response to usage of modern technology with plants

This might have happened because there is a lack of awareness about modern technologies and how it can be applied to agriculture. The result in favor of 'YES' may increase if campaigns and information are spread among groups which are unaware of it.

- Traditional remedies taken to avoid plant diseases

Some traditional remedies which can be carried on a local scale like using pesticides, soil conditioner, warming the soil before planting, shredding the affected leaf so that disease does not spread further. Sometimes, these remedies can also be successful to avoid disease in plants but they cannot assure that they work every time. This research tries to avoid such factors by using up-to-date skills to do the same.

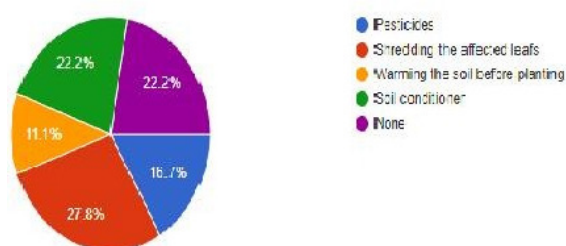


Figure 3. Usage of Traditional Remedies

- Encounter with Plant diseases

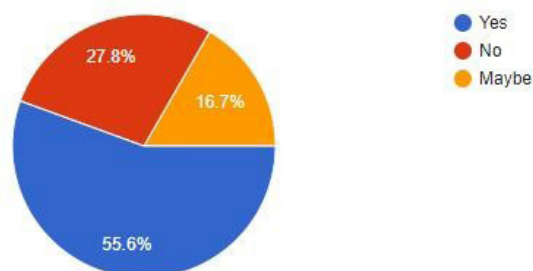


Figure 4. Percentage of People Encountered Disease in Plant

The data shows that more than 55.6%, which is more than half of the population in the survey, have encountered disease in their plants. However, due to a lack of knowledge, they are unable to classify them.

- Most common plant diseases

In this survey, pictures of leaves affected by diseases were shown to people and they were asked whether they have come across them or not. The diseases decided were Anthracnose, Downy Mildew, Leaf Spot, Late Blight. The bar chart shows the distribution of these diseases.

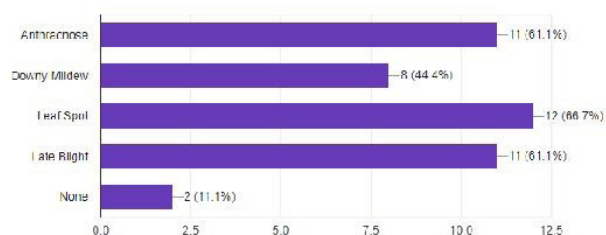


Figure 5. Distribution of diseases Encountered.

From the above graph, the most common disease found was leaf spot, Anthracnose, Late Blight, and Leaf.

Data Mining Techniques

Data Mining features are classified into two aspects: Organization and Clustering. Classification refers to classifying unknown values using information provided in a training set. In a training set, there is no previous knowledge about data, which is to be classified. Clustering is an unsupervised learning problem. It is used to split unknown data samples. The most common clustering technique is K-mean.

K-mean

It is the most usually used clustering technique. The Pie chart created from survey shows the percentage of people which have encountered any diseases in their plants. This data can also be used to show how percentage of plants is affected by disease but this will only be an assumption.

The K in K-means specifies the number of clusters formed. Given the set of unknown classification, I have to group the data with the help of similarities. This group forms a cluster.³

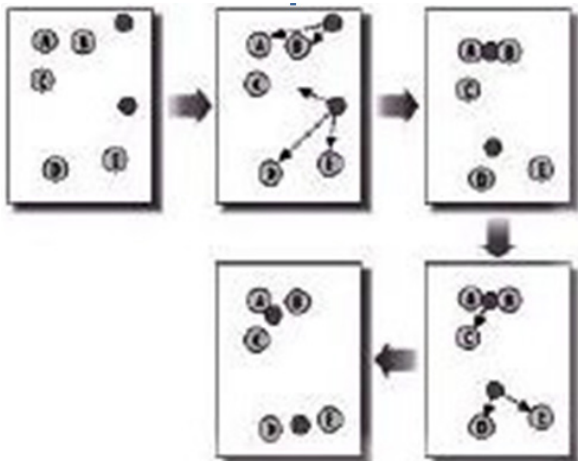


Figure 6.K-Mean

When the leaf of a plant is to be scanned using a smart phone or an image in jpeg, png format are provided for first time or new disease. K-means can be used to cluster the data of the leaf with similarities from large scanned data from similarities found.

Neural Network

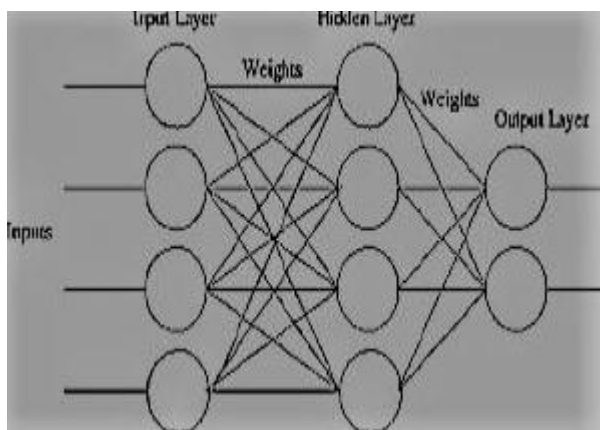


Figure 7.Neural Network

Neural network is a mathematical model. It is also called as artificial neural network. Neural network contains interconnected neurons to form a network. There is an input layer which will take the input of the leaf image by camera. There is another layer called the output layer which holds the correct output i.e. plant diseases. The middle layers

are hidden layers which process and compute the data in forward propagation.⁴ At early stages the output may not be correct but the network compares the output with the actual output disease and calculates error. This figure is transferred back in the network through backward propagation. Hence the system learns and can produce correct output disease by learning from previous knowledge gained from previous outputs.

SVM

Support vector machine, is used to arrange samples of data in to different clusters. It is a supervised learning method. SVM is mainly used to yield prediction.

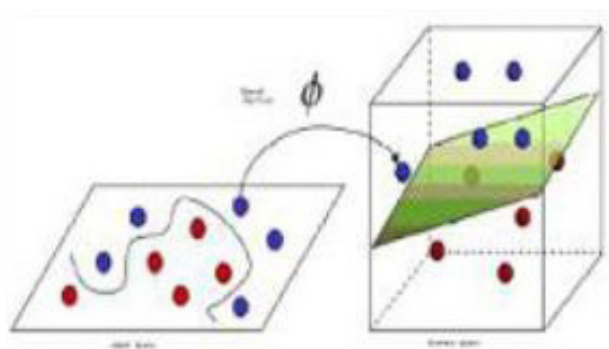


Figure 8.Support Vector Machine

SVM can be used for both organization and regression examination. Here I am taking a sample set. The sample set is founded on the "intimacy" of the elements.⁵ Support vector machine (SVM) is a powerful machine learning technique to classify data. SVM has been used in a range of problems and they have already been successful in pattern recognition in bioinformatics, cancer diagnosis, early detection of plant diseases and more. Figure shows SVM topology in hyperspace.

Problem Identification

Agricultural image data like jpeg, png and gif consume maximum storage and use maximum bandwidth for transmission that frequently results in degradation of image quality. The background data affects the resulting image. In the practical approach, optimization method used for a particular crop diseases and continuous computerized intensive care of plant can be done by automation technique. Leaf color, size and texture are changes with climate and environment conditions. The field expert and regular explanations are required in execution. The research suggests that disease detection methods show good results with an ability to find crop diseases and some limits.

Conclusion

This research is expected to recover in future. The discovery of plant leaf at early stage is essential to prevent losses in agriculture. It tries to eliminate the element of delay to save the plant before disease spreads. In modern stage, data

mining applications such as arrangement and clustering approaches along with image processing. In the final stage, smart phones can provide high accuracy and quick results. It arrange for high resolution images which will be helpful for detection leaf color and shape of any plant in agriculture. Furthermore, in future these methods could be merged and improved with skills such as Augmented Reality and Computer Vision.⁶ The survey concludes that I can compare all the features and find the effective prototype model that detects any leaf diseases. It will be helpful for noticing the color and shape of any type of plants in agricultural field. Threshold and neural network algorithms be responsible for high accuracy.

References

1. Vijai Singh "Detection of plant disease using image segmentation and soft computing techniques 4(1).
2. Subhadra k, Kavitha N. Plant leaf disease detection using data mining-A survey 5(1).
3. <https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a>
4. <https://towardsdatascience.com/understanding-neural-networks-19020b758230>
5. <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
6. http://www.research.ibm.com/haifa/dept/imt/ist_dm.shtml