Plant Leaf Disease prediction using Machine Learning

Ms. Jyoti Pal[1], Mr. Ankit Singhania[2], Ms. Ishita Mukherjee[3]

Prof. S.H. Darekar[4]

Student, Information Technology Engineering, Bharati Vidyapeeth College of Engineering, Navi Mumbai, Maharashtra, India

Assistant Professor, Dept. of Information Technology, Bharati Vidyapeeth College of Engineering, Navi Mumbai, Maharashtra, India

ABSTRACT

*Agriculture is critical to India's economic growth. Farmers struggle to choose the right fruit and vegetable crop. Disease management by hand is a difficult job. The majority of diseases affect the plant's leaves or stems. As a result, agriculturists must develop effective techniques. To predict the disease, we combined two techniques in this paper. The proposed work's aim is to diagnose disease using image processing and clustering techniques on a photo of diseased plant leaves. Preprocessing begins with the input image. The RGB to L*a*b conversion of the input image of leaves is then performed. Following that, hierarchical clustering is used to segment leaf disease. The predominantly green color pixels are masked after segmentation depending on specific threshold values. Finally, the proposed method's precision was compared to that of other approaches. The suggested procedure was shown to be 92% accurate.*

Keywords: Image Processing, Data mining, RGB, Clustering, Segmentation

INTRODUCTION

Plant blight Diagnostics is both an art and a science. The diagnosis process (i.e., the identification of symptoms and signs) is essentially visual and necessitates both intuitive judgement and analytical methods. [6]. Plant diseases decrease the quantity as well as the consistency of the goods produced by plants. Diseases are changes in a plant's natural state that affect or disrupt essential functions like photosynthesis, transpiration, pollination, fertilisation, germination, and so on. Pathogens, such as parasites, microbes, and viruses, as well as harmful environmental factors, cause these diseases. [7]. Farmers need professional supervision on a constant basis, which can be prohibitively costly and time consuming.

In all times of the year, the leaf has some benefits over flowers and fruits. Farmers are very nervous about the high costs of disease prevention operations, which results in significant losses. The cost-effectiveness, automated correct diagnosis, and classification of diseases based on their specific symptoms have become important and extremely useful to farmers and agriculture scientists alike. [8]. In horticulture/agriculture research, early disease identification is a big challenge. Many diseases cause signs, which are the primary methods for diagnosing diseases that produce external symptoms as a result of a sequence of reactions between the host and pathogen. As a result, some significant decisions on safe practices, plant production, and processing have recently been taken. [9].

EXISTING RESEARCH

Ajayi Adebowale, Idowu S.A, Anyaehie Amarachi.A[1] in 2013 presented their research in **"Comparative Study of Selected Data Mining Algorithms Used For Intrusion Detection"** Various academic groups have suggested a variety of approaches in the comparatively recent field of data mining and intrusion detection. Researchers are also looking for ways to improve the reliability of data mining tools for intrusion attack classification. They compare the results of well-known attack classification algorithms**.**

P.Revathi, M.Hemalatha [2] in 2014 proposed their research in **"Classification Of Cotton Diseases Using Cross Information Gain_Minimal Resource Allocation Network Classifier With Particle Swarm Optimization"** . To detect cotton leaf spot diseases, they created a framework focused on machine vision and data mining techniques. In the leaf spot regions, fungus, infectious, and bacterial diseases are most common, and they play an important role in the crop situation. and explains six different kinds of cotton plant diseases.

Kshitij Fulsoundar, Tushar Kadlag, Sanman Bhadale, Pratik Bharvirkar S.P.Godse[3] in 2014 introduced their research in **"Detection And Classification Of Plant Leaf Diseases"** Herbalists research medicinal plants extensively and learn about their medicinal properties. Near-Infrared Spectroscopy (NIRS) has a wide range of uses in agriculture, plants, and other areas, but it is still uncommon to use it to identify plant varieties. This project details the development of an Android system that would allow users to distinguish plant species through pictures of the plant's leaves captured with a camera.

Bhushan R. Adsule, Jaya M. Bhattad[4] in 2015 presented their research in "**Leaves Classification Using SVM and Neural Network for Disease Detection"** The use of neural network processing and help vector machines to identify cotton, citrus, and lemon leaf diseases. A few contaminated leaf samples were gathered and photographed in a sterile environment using a digital camera with a precise calibration protocol. The colour function extraction from RGB colour model, where the RGB pixel colour indices have been extracted from the specified Regions of Interest, is used to classify leaf diseases (ROI).

Ozge Aksehirli, Duygu Aydin, Handan Ankarali and Melek Sezgin [5]in 2015 proposed their research in **"Knee Osteoarthritis Diagnosis Using Support Vector Machine and Probabilistic Neural Network"** Artificial intelligence systems such as Support Vector Machines (SVM) and Probabilistic Neural Networks (PNN) have gained widespread acceptance. These methods provide generalizable results due to their solid statistical basis.

PROPOSED METHOD

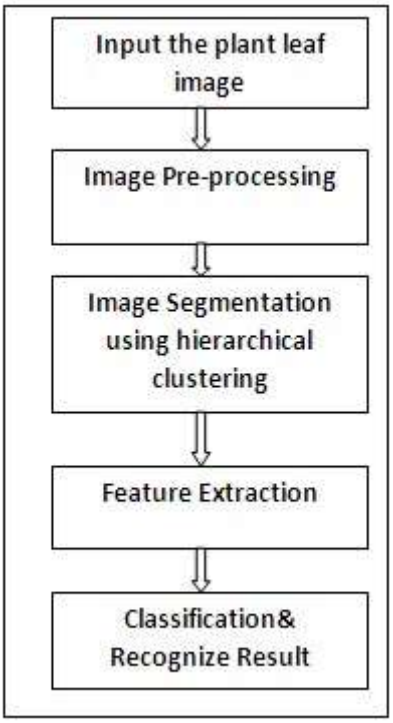

Figure 1. Flow of new method

The suggested system's flow is depicted in Figure 1. Different types of commercial crops, food grains, fruits, and cereals samples, both balanced and unchanged agriculture/horticulture produce used in the current study, are gathered and a dataset is generated in this scheme. As an input, these images are used. The picture is then pre - processed to eliminate the noise.

The suggested method is used to detect diseases in different types of leaves. For segmentation, the suggested scheme employs a hierarchical clustering algorithm, which divides the entity into three clusters. To test the disease affected area in the leaf, the proposed approach considers few attributes for feature extraction.

The features are:

- ✓ Correlation
- ✓ Contrast
- ✓ Energy
- ✓ Mean
- ✓ Standard Deviation
- ✓ Variance
- ✓ Homogeneity
- ✓ RMS
- ✓ Entropy
- ✓ Smoothness
- ✓ Skewness
- ✓ IDM
- ✓ Kurtosis

Finally, an SVM classifier is used to perform identification.

The input picture of anthranose and block spot impacted leaf as seen in figures 2 and 3. This image for the input is taken from the dataset archive. From RGB to CIELAB colour space, a colour transformation structure is developed. The contaminated area of interest is then identified using color-based segmentation and k-means clustering. The number of clusters chosen is entirely up to the customer. In most cases, 3 or 4 provide the best outcomes. The following Matlab figures demonstrate sample clustering findings on Septoria Leaf Spot of Tomato. The stable and disease-affected portions of the leaf can also be seen. This is the area that needs to be processed more.

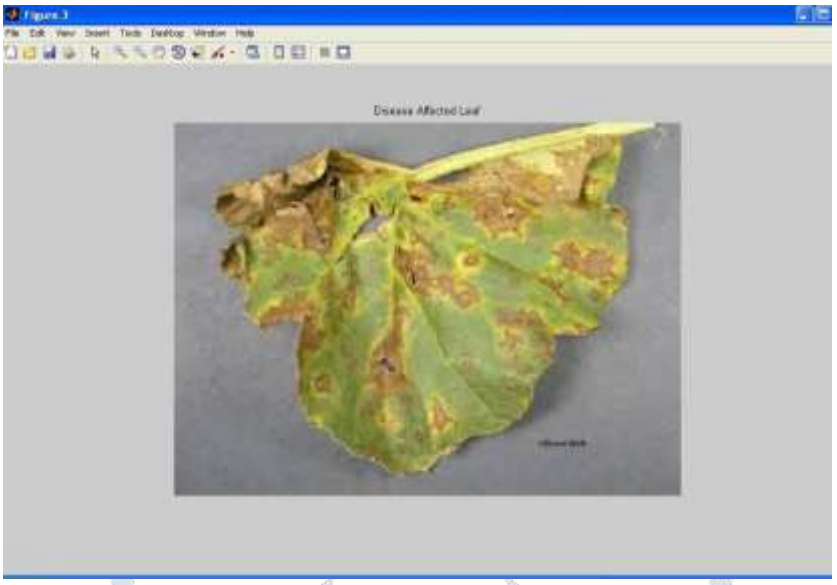

Figure 2. Anthronose affected leaf

The next contaminated cluster is chosen. Green pixels are concealed using the threshold value that has been set. Pixels on the borders are often omitted because they do not add to the disease detection process.
The hierarchical approach has a better detection accuracy than K–mean clustering.[10] The distinction between K – mean and Hierarchical clustering is shown in the diagram below. Segmentation using K mean clustering results in a single cluster with no clear boundaries. The differences are shown in Figures 4 and 5.

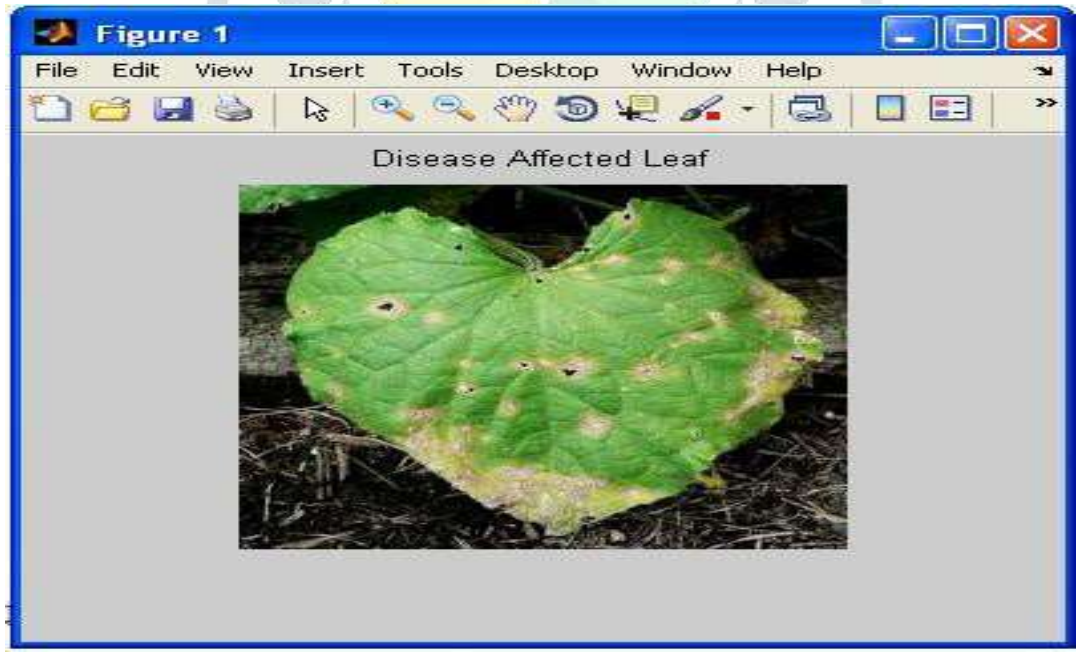

Figure 3. Black spot affected leaf

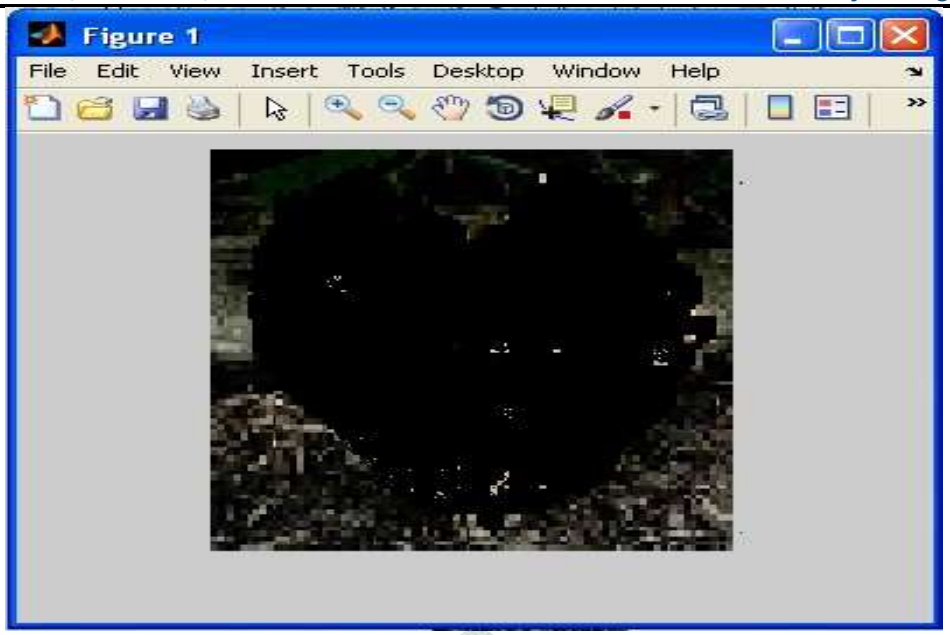

Figure 4. Result of segmentation using K Mean clustering

By using k mean clustering to slice, it is difficult to detect spots clearly. Figure 4 depicts the situation. However, in order to diagnose the disorder, hierarchical clustering produces three clusters. Figure 5 depicts the situation.

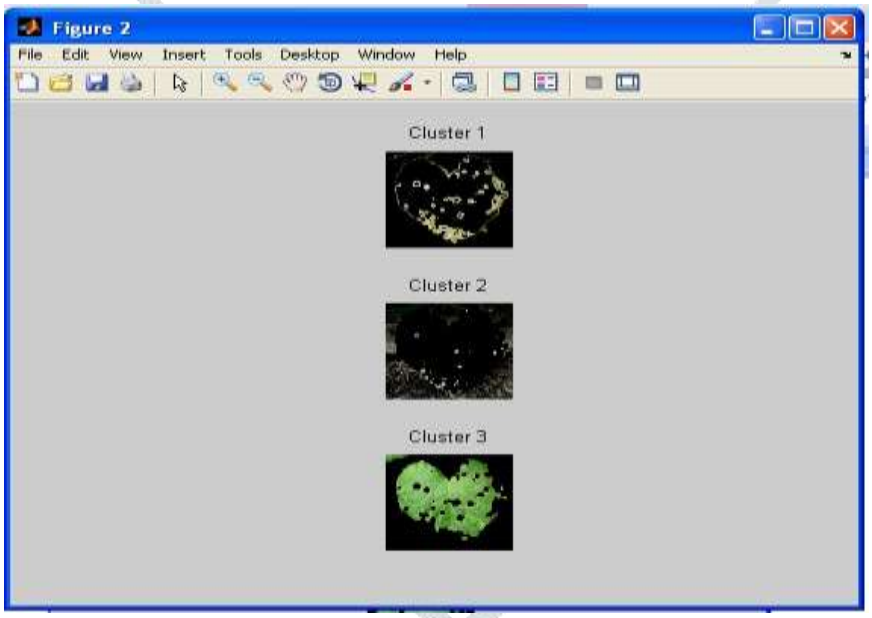

Figure 5: Segmentation with hierarchical clustering

The polluted area is shown in cluster 1 in figure 5, while the green area is shown in cluster 2. As a result, the disease-affected region is indexed using hierarchical clustering.

RESULTS AND DISCUSSION

Table 1. Recognition Accuracy

No. of dataset	Segmentation methods	Accuracy
10	k-mean clustering	75.86
20	Fuzzy c Mean	80.05
40	Hierarchical clustering	92.72

The identification precision for detecting diseases on leaves is depicted in Table 1. The hierarchical Clustering algorithm has the best precision, with a score of 92.72 percent. Figure 6 shows a graphical depiction of the largest.

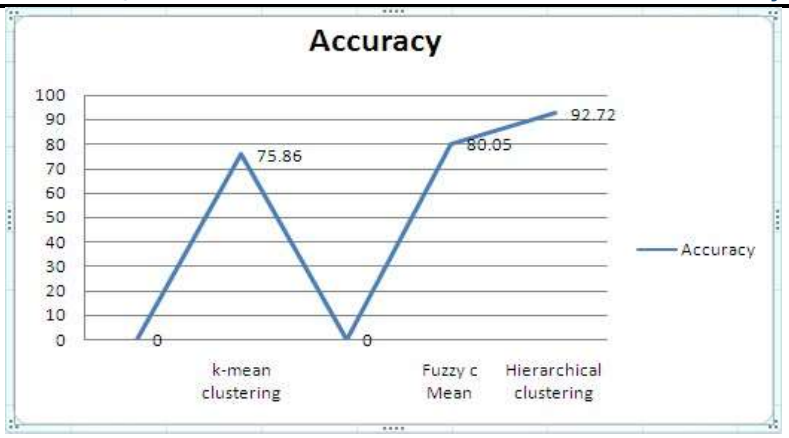

Figure 6. Comparison of accuracy between K mean and hierarchical clustering

Finally, the graph indicates that hierarchical has excellent disease detection accuracy.

CONCLUSION

The thesis examines and summarises image processing strategies for identifying plant diseases in a variety of plant species. Hierarchical clustering, GLCM, and SVM are the main strategies used. Optimization of the technique for a given plant, the effect of background noise in the acquired picture, and automation strategies for continuous automatic monitoring of plant leaf diseases under real-world field conditions are only a few of the difficulties that these techniques face. The proposed method is a useful one that can greatly aid in the effective diagnosis of leaf diseases with minimal computational effort.

REFERENCES

[1] Ajayi Adebowale, Idowu S.A, Anyaehie Amarachi.A "Comparative Study of Selected Data Mining Algorithms Used For Intrusion Detection" International Journal of Soft Computing and Engineering (IJSCE) Volume-3, July 2013 Issue-3 PP:2231-2307

[2] Revathi and Hemalatha "Identification of Cotton Diseases Based on Cross Information Gain_Deep Forward Neural Network Classifier with PSO Feature Selection" ISSN : 0975-4024 Vol 5 ,6 Dec 2013-Jan 2014, PP: 4637

[3] S. R. Dubey, P. Dixit, Nishant Singh and J. P. Gupta, "Infected fruit part detection using K-means clustering segmentation technique," International Journal of Artificial Intelligence and Interactive Multimedia, vol. 2, no. 2, 2013.

[4] Ozge Aksehirli, Duygu Aydin, Handan Ankarali1 and Melek Sezgin "Knee Osteoarthritis Diagnosis Using Support Vector Machine and Probabilistic Neural Network" IJCSI International Journal of Computer Science(IJCS) Vol. 10, May 2013 Issue 3, No1, PP: 1694-0814 Location: University of Düzce,Turkey

[5] Hrushikesh Marathe, Prerna Kothe "Leaf disease detection using image processing technique", International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 3, March – 2013 ISSN: 2278-0181.

[6] A. A. Gurjar and V. A. Gulhane, "Disease Detection On Cotton Leaves by Eigen feature Regularization and Extraction Technique," International Journal of Electronics, Communication & Soft Computing Science and Engineering (IJECSCSE), vol. 1, no. 1, pp. 1-4, 15 June 2012.

[7] Bhushan R. Adsule, Jaya M. Bhattad "Leaves Classification Using SVM and Neural Network for Disease Detection" International Journal of Innovative Research in Computer and Communication Engineering IJIRCCE Vol. 3, Issue 6, June 2015

[8] J. D. Pujari, R. Yakkundimath and S. B. Abdulmunaf, "Automatic Fungal Disease Detection Based on Wavelet Feature Extraction and PCA Analysis in Commercial Crops," , vol. 6, no. 1, 2013.

[9].Kshitij Fulsoundar, Tushar Kadlag, Sanman Bhadale, Pratik Bharvirkar S.P.Godse(2014) "Detection And Classification Of Plant Leaf Diseases" Volume: 03,2005 Publisher:IEEE

[10] Patil, V., Ingle, D.R. An association between fingerprint patterns with blood group and lifestyle based diseases: a review. Artif Intell Rev 54, 1803–1839 (2021). https://doi.org/10.1007/s10462-020-09891-w

[11] Sachin D., Shaikh M.Z., Kondke H.B. (2020) Network Behaviour of Open vSwitch as Per the Anticipated Functionality of Network Application Programmed Over SDN Using Pox Controller. In: Smys S., Senjyu T., Lafata P. (eds) Second International Conference on Computer Networks and Communication Technologies. ICCNCT 2019. Lecture Notes on Data Engineering and Communications Technologies, vol 44. Springer, Cham. https://doi.org/10.1007/978-3-030-37051-0_79