International Journal of Research in Agricultural Sciences
Volume 10, Issue 1, ISSN (Online): 2348 – 3997

An Overview on Detection and Classification of Plant Diseases through CBIR for Mobile Application

Saniya Santosh Puchalwar[1*], Om Vivek Mangrulkar[2], Prajit Himdev Bhoyar[3] and Milind V. Lande[4]

[1,2,3,4]Government Polytechnic, Department of Computer Engineering, Gadchiroli, Maharashtra, India.

*Corresponding author email id: saniyapuchalwar123@gmail.com

Date of publication (dd/mm/yyyy): 28/02/2023

***Abstract* – An average farmer is unaware of the numerous illnesses that might infect his agricultural plantation. This might result in a decrease in agricultural output. Agriculture is one of the most important sources of income in developing countries. Soil quality, humidity, temperature, and disease all play an important part in crop production. Farmers are typically uninformed and unaware of environmental conditions that may impact crop cultivation. To assist farmers, we may use machine learning and image analysis tools to deliver useful information. This may be accomplished using a web portal and mobile phone We suggested in this work to construct a system that offers real-time feedback based on plant input photos. The use of technology has become critical in assisting farmers in gathering significant and up-to-date information and knowledge, which are key resources on which farming depends. The goal is to treat plant illnesses and so manage them by meticulously identifying plant leaves. The aim is to create a prototype that is timely, relevant, accurate, and simple to use update to account for fast technological change. Based on a query image, the Content Based Image Retrieval (CBIR) approach is used to recover photos of damaged plants from a training dataset. The obtained photos are segmented using Hierarchical Clustering, resulting in clusters of sick plant images. Support Vector Machine is then used to classify the clusters (SVM) Classifier based on cluster characteristics that checks the proper type of illness impacting the plant set.**

***Keywords* – CBIR, Hierarchical Clustering, Segmentation, Feature Extraction, SVM, Plant Diseases, Android Application.**

I. Introduction

Agriculture is our country’s backbone. Many farmers are still having difficulty diagnosing illnesses that damage plants, vegetative crops, and grains. Though professionals are available to diagnose illnesses, prediction by naked vision may not always be accurate. As a result, having an automated expert system will be beneficial. Even though research is conducted on a regular basis to forecast plant disease, accuracy in identifying the correct illness and diagnosing it is still not ideal. In this research, the major performance measure to be focused on in recognizing the proper sort of illness in plants is accuracy [2].

Food is one of the most fundamental human needs. As the population grows, so does the availability of food. This is made more difficult by crop many illnesses cause yield loss, which can be avoided if recognized early. The global per capita food supply has increased, as has investment on agricultural research and development, and the aim is that the increase will occur in emerging nations. Advances in information and communication technologies (ICTs) have opened up new paths in knowledge management, which will be critical in tackling the current issues related with sharing, exchanging, and distributing agricultural knowledge. India is predominantly an agricultural country. Agriculture is the mainstay of its economy [3].

The dataset used for this experiment was obtained from the Online Data Repository. The damaged leaves of a tomato plant are used as the dataset in this case. Tomato plant diseases caused by fungi, bacteria, and viruses are being researched. This study focuses on Septoria Leaf Spot, Fusarium, Verticillium Wilt, Mosaic Virus, and Bacterial Spot. Experts use naked eye inspection to diagnose these disorders, which takes time. There are very few expert level consultants accessible nowadays, and their services are highly costly. Because plant diseases reduce plant quality and quantity, it is essential to employ technology for automated disease detection and diagnosis [2].

The goal of this work is to go through several picture retrieval, segmentation, and classification algorithms. The paper is divided into the sections listed below. This section discusses each phase of the disease detection process in plants. In a detailed examination of the methods utilized in the available literature is addressed. And discusses numerous methods employed by researchers for disease detection in plants, as well as a comparison analysis of their performance. This comparison research aims to demonstrate the benefits of each method. Finally, discusses the comparison study's findings [2].

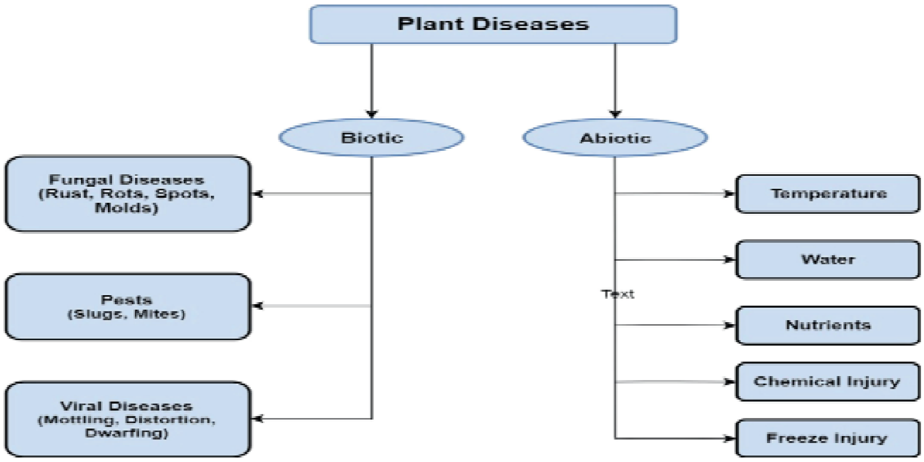

Fig. 1. Various causes of disease in plants.

II. Literature Survey

The authors of discussed a mobile phone strategy. A cell phone-taken image is utilized as the query image, and its properties are contrasted with those of a database image using the CBIR approach. The three elements of color, texture, and form are the key criteria used to retrieve images. K-means clustering, and Euclidean segmentation are used to separate the resulting pictures. With the use of android phones, a distance calculation is used to locate the image that matches the query image the closest [2].

The prediction of leaf diseases based on hierarchical clustering, k-means clustering, and fuzzy C- Mean (FCM) clustering was given by the authors in. These clustering methods created a lot of sick and healthy clusters. In this case, performance-wise, hierarchical clustering generated superior accuracy. Then, using an SVM classifier, the characteristics are retrieved for disease categorization [2].

The segmentation of plant leaf pictures using the k-means clustering approach to identify the damaged regions was covered by the authors in. Once the damaged areas have been located, color and texture characteristics are extracted from them utilizing a customized grey level dependency matrix and color co-occurrence matrix after the RGB values have been converted to HIS pixel values to serve as a discriminating feature for classification. Then, for illness categorization, ANN and SVM classifiers are employed. Both SVM and ANN are trained using the retrieved features. The picture samples are then recognized and classified using the test images. The SVM outperforms ANN in classification accuracy and has been shown to be a potent tool for the automated categorization of symptoms caused by fungi [2].

The authors of employed the CBIR Technique, which is a method for content-based image retrieval. The query picture and database image are both extracted based on color, shape, and texture, and they are both saved in a different database that has undergone Euclidean distance similarity assessment. The threshold value determines how to get the output photos [2].

II. Methodology for Disease Detection

The fundamental concept of plant disease detection and diagnosis is covered in this section. Figure 1 depicts the intended.

The approach for detecting and classifying plant diseases is provided [2].

2.1. *Image Capture*

The Image Dataset is developed by collecting images of normal and sick plant leaves from the internet repository and neighboring nurseries [2].

2.2. *Image Preparation*

There are several approaches for removing noise from a picture or other undesired object. The Median Filter has been proven to offer superior noise reduction outcomes in plant leaves than other filtering approaches [2].

2.3. *CBIR for Image Retrieval*

Color, shape, and texture are the primary aspects of Content Based Image Retrieval (CBIR) systems. The inquiry picture is an image with these fundamental qualities. The query picture is then compared to the image in the dataset, and only the photos whose features perfectly match the query image are returned. The collected photos are then saved in a separate database for future use. These photographs serve as the training dataset, which contains the finest diseased image collection for illness prediction and diagnosis. The following is an explanation of the image retrieval process using the CBIR technology [2].

2.3.1. *Retrieval of Colors*

When a comparison image is utilized, it is first preprocessed, and a color histogram is generated from it.

The histogram displays the distribution of colors in each picture pixel. Images are taken into consideration for retrieval if their colors closely like those of the query image [2].

2.3.2. *Texture Retrieval*

By removing the texture defining values from the test picture, texture similarity is calculated. Periodically, values are calculated for the texture analysis based on size, degree of contrast, and directionality [2].

2.3.3. *Shape Retrieval*

From the search picture, the shape's local features, such as sets of succeeding boundary segments, and global properties, such as aspect ratio, circularity, and moment invariants, are retrieved. The features that match the query image are retrieved after these features are matched to the features extracted from the dataset photos. The pictures that are subsequently acquired using CBIR methods are kept separately in a database [2].

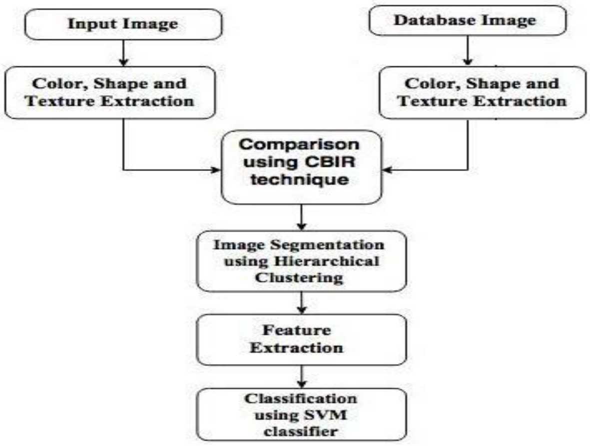

Fig. 2. Disease detection and diagnosis methodology.

IV. Summary of the Problem

Despite the government initiatives, the issue of young farmers lacking timely, accurate, and relevant knowledge-especially those without any prior experience in the field-remains pertinent and needs our immediate attention. The outdated system of teaching that was intended to transfer global research outcomes to African farmers has proven ineffective, and most institutions lack the equipment and expertise to share and distribute outputs to small-scale farmers and other players. The field of computer science that focuses on machine learning and image analysis has great potential for the agricultural community, particularly when it comes to the use of mobile devices that can execute some computer science applications both online and offline. Particularly in the diagnosis of plant illnesses by employing leaves and stems and in the detection of insects long before damage is done to the plantation, deep neural networks and image analysis show promise. Plant food security is seriously threatened by illnesses. Despite this, it is still challenging to quickly identify them in many regions of the world. The major cause of this is a lack of the required infrastructure. Young farmers frequently depend on the employees to recognize and treat a variety of common plant illnesses. Even with farmers' cooperation, it is difficult since they must inefficiently monitor such a big area [3].

Even on days when the service is provided on schedule, the diagnosis is made by sight, which increases the likelihood of a false diagnosis. Verbal information is frequently inaccurate and unscientific, which can lead to treating the wrong sickness with the incorrect chemicals and eventually hurting the environment [3].

V. Important Concerns and Problems

On [3], diverse plants and the potential illnesses that may damage them, several theoretical studies have been carried out. This has led to a wide range of potential approaches to finding a solution to our dilemma. Data from several sources must be gathered as input for the automation of illness identification. The following are some of the major problems that we have encountered:

- There aren't many images of plant leaves of high quality.
- To get correct findings, a substantial percentage of the data set must be considered.
- To enhance the outcome’s, newly gathered data must be periodically added to or updated.

- The size, shape, and color of plant leaves change as the environment changes.
- Reviews indicate that machine learning and image processing approaches have a greater ability to detect illnesses.
- Therefore, the current body of research has to be improved.
- Large dataset training necessitates high-end hardware that is not readily available.
- For specific plants, regular inspection is required.
- Various plant leaves have different illnesses, which makes diagnosis difficult [3].

VI. Purpose and Goals

Our [3], major goal was to create an accessible system that is simple to use even for those who know very little about computers and mobile devices. This means that our goal was to create a system that would be extremely beneficial to all types of farmers. Following are our primary goals:

- To create a system that any farmer could use simply.
- To compare the output of the system with professional judgement.
- To put the best machine learning approaches into practice to get reliable results.
- To create a system that can identify plant diseases based on visual input [3].

VII. Justification for Proposed Solution

The [3], expansion of mobile communication technology is largely to blame for the progressive increase in the accessibility of communication resources, which now cover a huge portion of India.

Due to their low cost, smartphones are now available to everyone, and because to recent developments in the availability of mobile data, the internet is now more broadly and cheaply accessible. According to early polls, as illustrated in Figure 1, farmers are more reliant on conventional information sources than on contemporary technologies.

Thus, there is no doubt that the system suggested in this study is what the world needs right now.

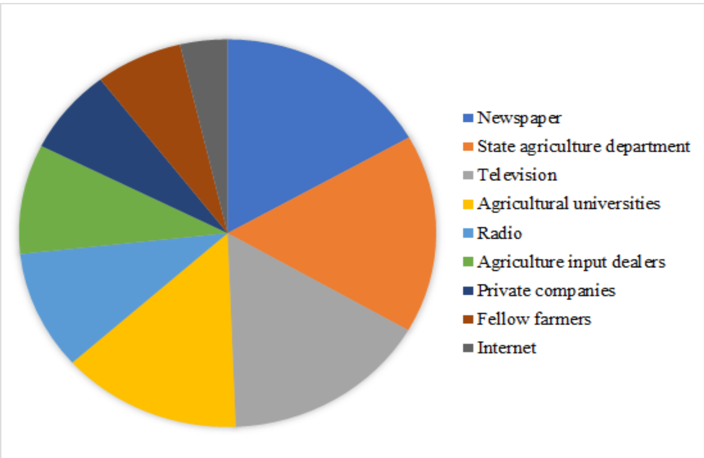

Fig. 3. Sources of information.

VIII. COMPARISON RESEARCH

Table 1 [2], provides a summary of all methodologies utilised in various articles, along with the distinct techniques they employed for the identification and categorization of plant diseases, and the degree of accuracy attained using each technique. Additionally, a summary of the benefits of adopting these techniques is provided, as mentioned in the cited studies.

Figure 2's graph compares the outcomes of the most popular algorithms for spotting plant diseases in terms of the level of precision attained in related research. The graph makes it very evident that hierarchical clustering and SVM classifiers offer superior accuracy for detecting and diagnosing plant diseases.

Table 1. A comparative study of algorithms and their advantages.

Algorithm Used	Accuracy Rate (%)	Advantages
K-Means Fuzzy C Means Hierarchical	75.86 80.05 92.72	Valuable Less Effort Accuracy
CBIR Thresholding	-	Faster Output
SVM ANN	83.83 77.75	Comparison results shows SVM is better
Feature Selection with PSO CIF-DFNN	95	Accurate Diminishes Error Rate High Performance
SVM Neural Network Pattern Recognition	70.21 96.27	Larger datasets are used More Features are extracted
K-Means	90.50	Complexity Decreases
K-Means SVM	88.89	Normal & abnormal leaves are studied Confusion matrix is plotted

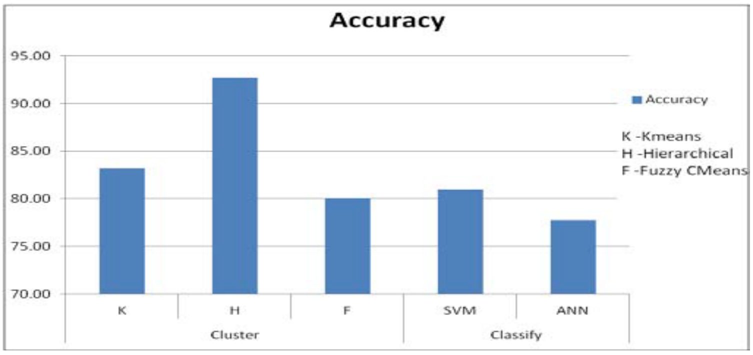

Fig. 4. Comparison of most frequently used algorithm's overall accuracy rate in percentage.

IX. SELECTION OF THE DATASET & AGUMENTATION

We [3], chose the freely available Plant Village dataset for our initial prototype. We have chosen nine of these species, including Apple, Cherry, Corn, Grape, Peach, Pepper bell, Potato, Strawberry, and Tomato, from this group. Table I lists the 33 classes that are accessible for these species. The datasets of choice are the Leaf Segmented and Colour datasets, as seen in Figure 2. We have chosen the colour and the size for the data augmentation variations with segmented leaves that will be cropped, zoomed, and rotated at random. As a result, the data will be fully used feasible, and there will be the greatest number of characteristics accessible for the neural network to recognise after training. Crops with only one class of labels have been left out since they don't help us identify plant diseases.

Table.2. List of crops and class labels.

Crops	Class Labels [3]	Images
Apple	Apple_scab, Black_rot, Cedar_apple_rust, Healthy	6342
Cherry	Powdery_mildew,Healthy	3812
Corn	Cercospora_leaf_spotGray_leaf_spot,Common_rust, Northern_Leaf_Blight, Healthy	7704
Grape	Black_rot,Esca_(Black_Measles), Leaf_blight_(Isariopsis_Leaf_Spot), Healthy	8125
Peach	Bacterial_spot, Healthy	5314
Pepperbell	Bacterial_spot, Healthy	4950
Strawberry	Leaf_scorch, Healthy	3130
Tomato	Bacterial_spot, Early_blight, Late_blight, Leaf_Mold, Septoria_leaf_spot, Spider_mites Two-spotted_spider_mite, Target_Spot,Tomato_Yellow_Leaf_Curl_Virus, Tomato_mosaic_virus, Healthy	36320
Potato	Early_blight, Late_blight, Healthy	430477

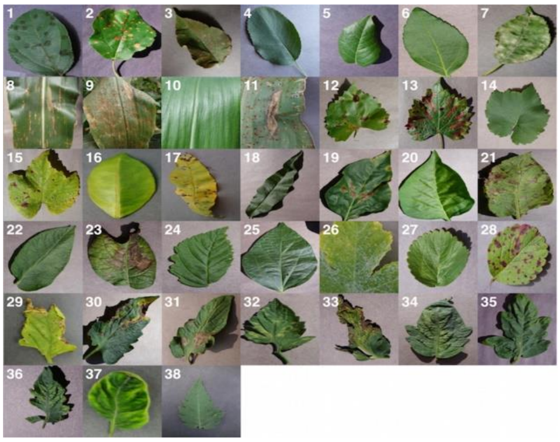

Fig. 5. Sample Images of Tomato_Bacterial_Spot Colour and Leaf Segmented.

X. K-MEANS CLUSTERING ALGORITHM

Clustering [4], is defined as the division or partitioning of input data points into clusters so that data points within the same group have comparable attributes to each other and data points in other groups have distinct features. For image clustering, several clustering approaches have been used to categorise, display, and scan pictures, as well as to improve the performance of clustering-related applications such as CBIR Techniques, Image annotations, and so on. Clustering techniques are often classified into two types: hierarchical and partitional. K-means is an unsupervised learning method that is used to solve the clustering algorithm. K-means employs a straightforward approach to divide a given dataset into a defined number of groups (k clusters). The first step is to locate k centre. Centers should be established in such a way that a change in cluster placement can affect the final output. The best approach is to keep them as far apart as possible. The data points are then assigned to the nearest centre. When all of the data points have been collected, the clusters have been formed. At this point, we must reassign k centroids as the centre points of the clusters produced in the previous phase. Following the acquisition of k centroids, the identical data points are assigned to their nearest new data centres. A loop has been established. The loop results in the k centres changing their present location after each step until no further modifications are made or the centres cannot be relocated.

K-Means Clustering Algorithm:

1. Determine the beginning coordinate of the centroid.
2. Determine the distance between each data point and the centroid.
3. Sort the data points by their shortest distance. (Find the centroid closest to you [4].)

XI. Approach

The [3], goal of this study is to build a Transfer Learning method by using the DenseNet121 architecture. Compared we suggest one easy but beneficial modification to the training strategy from the prior approaches. We have divided the dataset of the different unique plants, which will be utilised to train a new instance of neural network for each unique plant species, in contrast to earlier techniques, as shown in Figure 3, where both the plant species and illnesses are detected. This means that, as seen in Figure 4, each plant will have a unique instance of its own neural network. When necessary, this can be improved individually in the future disrupting other plant species' neural networks. Due to the greatly reduced quantity of images that must be processed at any one moment, this strategy will also significantly reduce the amount of training time needed. As previously noted, we used DenseNet121 architecture, which is substantially more accurate than its forerunners, for the training of the neural network itself. The training set and test set were each given an 80:20 split of the dataset. Table II displays the level of accuracy attained for each plant species throughout training. We individually trained the models for each distinct plant for 30 and 50 epochs, as shown in Table II. 98.87 percent accuracy was achieved on average while training for 30 epochs, and 50 epochs. Epochs provided an accuracy of 99.25 on average. For the two Epochs, the accuracy did not considerably alter. As a result, we resolved to prepare for 30 Epochs in our future undertakings.

Table 3. Training accuracy achieved with dense net 121.

Crops	Accuracy For 30 Epochs (%)	Accuracy For 50 Epochs(%)
Apple	99.4	99.6
Cherry	100	100
Corn	96.3	97.2
Grape	99.4	99.5
Peach	99.4	99.7
Pepperbell	99.5	100

Crops	Accuracy For 30 Epochs (%)	Accuracy For 50 Epochs(%)
Potato	99.3	99.6
Strawberry	100	100
Tomato	96.6	97.7
AVERAGE	98.87	99.25

XII. Recommended Solution

Based [3], on photographs of the plant leaves as shown in Figure 5, this study suggests a web-based tool that will work in conjunction with a mobile application to diagnose the various plant diseases. The suggested system will be known as "PotBot." With the help of PotBot, farmers will be able to photograph plant leaves using their smartphones. The image will be uploaded through the internet to the application server, where it will be categorised based on the data supplied. After then, the outcome will be seen on the mobile device. Then, using this knowledge, farmers may determine whether any illnesses have affected their farms and take appropriate action. Due to the high computational demands of the categorization process, we have chosen to assign that task to the web server that will host the web application.

Like how they use the mobile application, farmers may access PotBot's services through the web-portal. Moving on to the technical part, we suggest using Python-based Flask/Django to create the online application that will also function as the server for the mobile platform. The mobile software will be initially developed for portable devices using the Android operating system. The actual Android application will be created using a combination of Java and Kotlin. TensorFlow will be utilised on the back-end server to process the models [3].

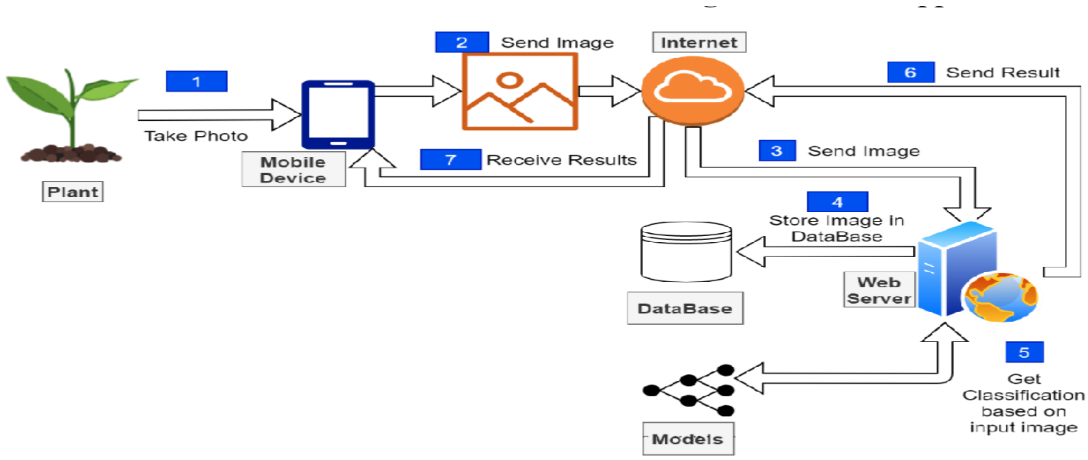

Fig. 6. System Architecture.

XIII. Relevance and Future Possibilities

The [3], potential for the suggested approach to help small-scale farmers is enormous. However, steps must be taken to guarantee that the application is appropriate and feasible. Farmers have a low level of literacy in a developing nation like India.

Making this programme broadly useable will be difficult. To make the application realistically useable, proper instruction and direction are required. To increase crop yields, PotBot can support the government's ongoing agricultural programmes. Focused research must be done on regionally particular flora. These research' findings must be included to the system. The usage of this application should be aimed at the younger generation of farmers since they will be better able to comprehend and learn how to utilise it. Making the programme totally offline will be a huge issue that can be solved in the future [3]. This [5], will be a significant accomplishment thanks to the development of Deep Learning and Mobile Technologies. Farmers won't need an internet connection to use the programme any longer. To better farmers' existing situation, information and communication technologies must be combined with agriculture. ICTs may assist in gathering and managing enormous amounts of data that can be saved and accessed as needed by the user. Farmers will benefit from having instant access to pertinent information [5].

XIV. Conclusion

The CBIR approach used to recover pictures of damaged leaves produces improved image extraction outcomes. It is evident from the comparative study that a variety of segmentation and classification methods are employed by researchers for the detection and diagnosis of plant diseases. In order to improve accuracy, this research suggests combining clustering and classification algorithms for plant disease detection and diagnosis. It is suggested to recover diseased pictures using the CBIR approach, segment the returned diseased leaf image using hierarchical clustering to anticipate the illness present in the leaf, then diagnose the proper type of disease using SVM classifier to generate more accurate results.

CBIR approaches successfully extract low-level image characteristics (colour, shape, and texture), and the k-mean clustering method improves system accuracy by grouping pictures with comparable features together. As a consequence, correct findings may be supplied to users via Android smartphones.

Acknowledgements

Our profound gratitude to all our professors who have assisted us in our attempt, whether directly or indirectly. Without Milind V. Lande (HOD of Computer Department in Government Polytechnic Gadchiroli, Var x Technology, this initiative would not have been feasible, hence we would want to convey our appreciation to him. Finally, we would want to express our gratitude to our parents for their unwavering support.

References

[1] Singh G, Yogi KK. A review on recognition of plant disease using Intelligent Image Retrieval Techniques. Asian J Biol Life Sci. 2020; 9(3): 274-85.
[2] Roshni C.R, Dr. M. Safish Mary, "A comparative study of algorithms used for detection and classification of Plant Diseases", International Journal of Science and Research (IJSR), Volume 6 Issue 2, February 2017, pp. 2147-2150, https://www.ijsr.net/get_abstract.php?paper_id=ART20171105
[3] Shrestha, U., Peter, J., Manandhar, S., Rajbanshi, S., & Padal, S. (2020). System Design for Identification of Plant Leaf Diseases using Deep Learning for Web and Mobile Platform.
[4] Deokar, A.S., Pophale, A., Patil, S., Nazarkar, P., & Mungase, S. (2016). Plant Disease Identification using Content Based Image Retrieval Techniques Based on Android System.
[5] Jayamala Kumar Patil, Raj Kumar, Analysis of content based image retrieval for plant leaf diseases using color, shape and texture features, Engineering in Agriculture, Environment and Food, Volume 10, Issue 2,2017,Pages 69-78,ISSN 1881-8366, https://doi.org/10.1016/j.eaef.2016.11.004.

Author's Profile

First Author
Saniya Santosh Puchalwar, Government Polytechnic, Department of Computer Engineering, Gadchiroli, Maharashtra, India.

Second Author
Om Vivek Mangrulkar, Government Polytechnic, Department of Computer Engineering, Gadchiroli, Maharashtra, India.
email id: ommangrulkar15@gmail.com

Third Author
Prajit Himdev Bhoyar, Government Polytechnic, Department of Computer Engineering, Gadchiroli, Maharashtra, India.
email id: rohitbhoyar111@gmail.com

Fourth Author
Milind V. Lande, Government Polytechnic, Department of Computer Engineering, Gadchiroli, Maharashtra, India.
email id: milind.jdiet@gmail.com