HYBRID METHOD FOR PLANT DISEASE DETECTION WITH TEXTURAL FEATURE ANALYSIS

Ketan Khullar, M.E Computer Science and Engineering , NITTTR, Chandigarh
Dr. Amit Doegar, Associate Prof, Dept of Computer science and Engineering, NITTTR, Chandigarh

ABSTRACT

The major focus of the image processing is on detecting the plant diseases. This technique is executed in 3 phases in which the picture is pre-processed, the image is segmented, attributes are extracted and the disorders are classified. The initial task is to de-noise and for improving the input image.To achieve this, GLCM model is adopted. The last task is to implement a classification method to predicting the diseasesaccording to the features that have been extricated. To detect diseases of plants, current techniques such as KNN, SVM, decision trees, and random forests are used. It is analysed that existing are unable to achieve high accuracy for the plant disease detection. In this research work, voting based process is proposed which achieve high accuracy for identifying the pathogens in plants. The new method is applied in python and compared with existing methods. It is analysed that proposed methods achieve high accuracy as compared to existing machine learning methods.
Keywords: Plant Disease, Machine learning, Voting Classification, SVM, Decision tree, Random forest

Introduction

Image processing is comprised of the multitude of tasks of digital pictures for altering the photometric properties of an image. DIP is a method wherein PCs are sent to control the pictures. Different sectors, like healthcare, amusement, instructive and guard, take on this cycle. Digital Image Processing based Multimedia system has become a basic component of information field. The technique, to sample and convert the image into a matrix of numbers, is put forward for handling the pictures [1]. The restricted accuracy helps to define the numbers for which the PC is utilized. Subsequently, to show such numbers digitally, it is essential to calculate these numbers. The restricted amount of accuracy values is controlled by means of DIP.Various classes are associated with DIP for upgrading, reinstating, breaking down and compressing the image. This cycle focuses on meta-heuristic techniques to control a picture with the goal of extracting the huge information from image under the period of upgrading an image. The corrupted images are processed after restoring the image in a measurable or numerical manner. This image is consisted of subtleties having degraded component. The processing of imagery is provided in the procedure of examining the image while determining the information automatically [2].The issue connected with safeguarding the plant straightforwardly alludes to the problem related to environmental variations and feasible horticulture. As indicated by researchers, the climate change prompts adjust the improvement stages and the sizes of microorganism development. The explanation which prompts cause this intricacy is a basic broad pace of transmission of infections in plant in existing situation when contrasted with the prior one. The districts, at which this sort of circumstance is happened and local mastery is inaccessible to manage these problems, are more inclined to new ailments. Detecting the infected plants in exact way and on time is challenging task to get exact horticulture. Preventing the excessively waste of monetary and different assets prompts gains the solid efficiency [3].The essential task is to prevent the infections in the varied climate so that the ailments are diagnosed ahead of time and precisely. The diagnosis of diseases happened on plants is carried out utilizing a few techniques. The side effects of certain diseases are not showed up anyway their influence should be visible later on. In this way, an upgraded analysis is put forward in handling these kinds of conditions. Some sort of show is acquired from different infections visually. The master experts can play out the analysis of image with unaided eye which is compelling procedure for diagnosing the leaves ailments. Be that as it may, the undeveloped designers tracked down intricacy for deciding the disorder. A CAD system is planned for diagnosing the diseases relying on the noticed and pictorial side effects of plants.An approach took on in horticultural fields is known as CVS. Such an approach is valuable to arrange the fruits and perceive the food items. This purpose is achieved by processing the image, characterizing the grains, diagnosing the weeds and a few other comparative errands are completed. The pictures are caught from digital cameras and the strategies are embraced so that these pictures are processed. The viable DIP systems like color analysis and thresholding are executed with the purpose of detecting the disorders. The viral, contagious and bacteriological diseases, the early and late scorch are regularly noticed messes on plants. The images are processed for diagnosing the ailments of plants in different stages like image acquisition, to pre-process and segment the images, separating the elements and arranging them. Figure 2 illustrates a general procedure to diagnose the plant disorders.

Literature

This work considered ConvNet (Traditional Neural Networks) along with certain conventional architectures like ResNet [4]. The accuracy of this research was remarkable, which was achieved by using an improved dataset with images of both healthy and defected leaves. The new architecture offered promise for tracking plant diseases. In addition, this architecture categorized photos as either normal or diseased. Finally, a new design was deployed in Anaconda 2019 for comparison. According to the testing, the new design appeared to be more effective than the previous ones at identifying plant diseases in regards to accuracy and recall.

This work noticed a requirement for creating a novel plant disease detection method [5]. By combining thresholding and morphological processes, the author has created a comprehensive segmentation technique. High processing power was not necessary for this method. The leaves were classified into different categories on the basis of infected and healthy by utilizing a DNN [6] model. When the plant village database was used to evaluate the techniques, the accuracy obtained was calculated to be approximately 99.25%.By reducing the feature vector size and increasing accuracy in this research, the designed architecture can be improved.

In order to identify and classify illness in tomato plant leaves, this research suggested the utilization of CNN algorithm and LVQ focused method [7]. After feature extraction, images were automatically categorized using CNN modelling. In order to classify the different types of diseases found in plant leaves, the colour information was used. According to RGB components, this system implemented filters to three types of channels. The LVQ algorithm employed output feature vector from the convolutional section to train the network. The experiments were run on an open-source dataset that included 500 photos of tomato plant leaves and 4 disorders of illness. The test findings demonstrated that the methods offered made it possible to precisely and quickly locate four different illnesses on tomato leaves.

This work developed the RFC-GLCM, an IoT based pathogen detection tool that is simple but effective. This technique was used to identify and categorize Sigatoka disease and gathered banana ends in mountain banana plants [8]. The developed architecture used image processing to pre-process and recover the leaf surface features. The GLCM features at the monitoring location were utilized to classify the photos using the RFC (Random Forest Classification) approach, and plant pathologists looked at these attributes to offer useful remedies. A new methodology was evaluated by using a number-criteria which includes, clarification, recall and accuracy on the mountainous banana dataset. In order to detect the diseases on plants the findings depicted that the introduced design had a 99% accuracy rate.

This research developed a system to detect infection on a kinds of plant species, that includes. maize, tomato, grape, potato, sugarcane, and apple[9]. To instruct DL model, a dataset containing 35000 images of normal and unhealthy plant leaves was used. It was used to identify and categorize the existence or lack of plant pathogens. The created framework produced 100% accuracy for identifying plant variety and plant diseases, as well as 96.5% accuracy for identifying plants.

This work utilized machine learning and image processing to autonomously recognize plant illnesses[10]. For a crop to provide a higher-quality harvest, plant diseases must be accurately and promptly detected. Early diagnosis decreases the need for ineffective medicines and lowers treatment expenses. The TL (Transfer Learning) technique was used to retrieve the crucial traits after amassing images of various plant species. The developed system's success rate was pegged at 94%.

For the identification of plants diseases this research developed a convolutional auto encoder-based unsupervised feature learning system based on the convolutional auto encoder [11]. The two steps in this algorithm were. Initially, as a result of the network's learning ability of producing differential features, manual characteristics were ignored. After that, a disease diagnosis procedure utilizing an unsupervised algorithm was started. Consequently, labelling the data was not required. To automate the process of identifying illness, the output of the auto encoder was fed into the SVM (Support Vector Machine) classification design[12]. The results showed that the planned algorithmic technique had an advantage over the standard approaches

Research Methodology

The PDD mechanism is designed to detect the affected portion of a leaf image. This research work introduces a comprehensive approach for plant disease detection, which encompasses several phases outlined below:

1. Pre-processing: - This phase focuses on preparing the image for disease detection from plant leaves. The process involves taking input images obtained from a reliable data source and incorporating them into a dataset[13]. A publicly available dataset called PlantVillage is utilized here. Plant Village is a website that provides information about plants and their associated diseases. The dataset used consists of images specifically depicting potato leaves, 5 categorized into three sections: good leaves, leaves affected by early blight sickness, and leaves showing late blight disorder. To facilitate further processing, the input image is converted into grayscale.

2. Segmentation: - This process entails segmenting an electronic image into different areas or parts. It makes the analysis process simpler by enabling the recognition and extraction of pertinent data from the segmented images[14]. The area of Interest (RoI) and picture boundary lines are established using the image segmentation technique. A label is given to each pixel, yet pixels with the same label can have different properties. In this study, the K-means clustering (KMC) technique segments the imagery of leaves. This algorithm clusters samples into different groups based on their distances from each other. By setting the value of k to 3, the image can be effectively segmented. Consequently, the afflicted area of input leaf is selected for further processing.

3. Feature Extraction-The output from the preceding phase is the required area.In this phase, the primary objective is to extract attributes from the designated area. These attributes, known as features, capture essential information about the image and are utilized for processing later. To identify plant illnesses, many factors like color, texture, morphology, and color coherence vector are taken into account. Utilizing feature extraction methods, these attributes are taken out of the picture. The GLCM, histogram-based algorithms, etc. are a few noteworthy methods. Particularly, the GLCM refers to a statistical method frequently employed for categorizing texture properties[15].

3. Classification of Data: - The final stage focuses on developing a classifier to determine plant pathogens leveraging the extricated attributes.There are two sections to the dataset, with a larger portion used for training and a smaller portion for testing. In this phase, the K-Nearest Neighbors (K-NN) algorithm is introduced. K-NN algorithm classifies unknown samples by comparing them to known samples using similarity functions[16]. Training and testing of the algorithm can be performed simultaneously. K-NN identifies the K nearest neighbors and assigns the majority vote of their classes to the unknown instance. Another effective and flexible machine learning algorithm used in this phase is Random Forest (RF), which combines multiple tree predictors. RF is able to manage enormous data volumes and achieves optimal results. It constructs an ensemble of trees and makes predictions with high accuracy. The RF algorithm incorporates hyper-parameters corresponding to decision trees or bagging classifiers. It creates random subsets of attributes in the dataset, resulting in overlapping random trees[17]. By integrating the outputs of RF and K-NN, the voting classification algorithm is implemented. This algorithm chooses one of the 2 classifiers at random and outputs the final prediction outcome.

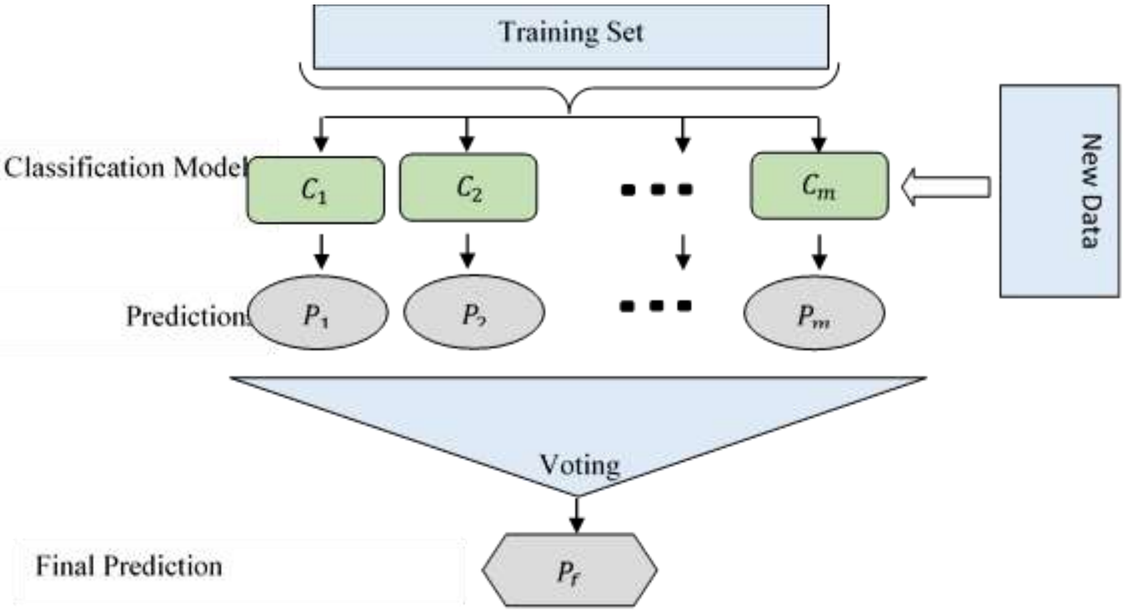

Figure 1: Proposed approach

Result and Discussion

Python is a sophisticated programming language with a significant emphasis on object-oriented features, dynamic semantics, and accessibility. It is frequently used as a scripting language to link various components and for quick application development. Python's simple syntax encourages code modularity and reuse by supporting modules and packages while lowering the cost of program maintenance.

4.1. Proposed Model and Dataset Description

The experiment conducted on the developed model involves using the Plant Village dataset. This dataset is publicly available and provides comprehensive information about various plants and their associated contagions[18]. The dataset's imagery is annotated with the appropriate disease kind that it depicts.


Figure 2: Sample Dataset Images

As shown in figure 2, the dataset is collected from the Plant village. The sample images of the dataset is shown which are used for training and test purpose.

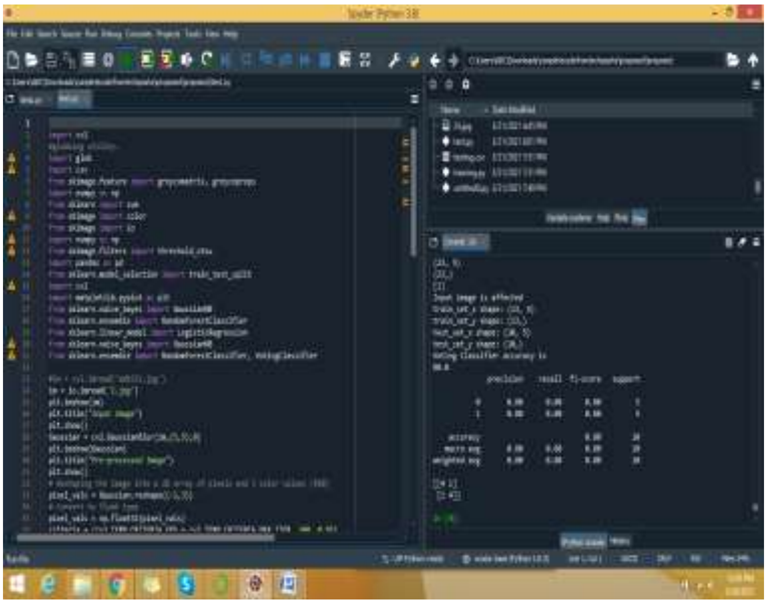

Figure 3 Implement Voting Classifier

The application of the vote classification system for forecasting plant diseases is shown in Figure 3.

4.2. Results

a. Accuracy: One popular statistic for assessing a program's effectiveness is accuracy. It calculates the percentage of correctly identified samples among every one of them. It has the following mathematical representation:

$$A_i = \frac{t}{n} \cdot 100$$

In this formula, n is the overall quantity of specimens, and t is the number of the specimens that have been identified correctly.

b. Precision: The ratio of correctly predicted positive outcomes to all projected positive cases is measured by an indicator of efficiency called precision.

Precision= TP/TP+FP

c. Recall: In psychological research, recall—also called sensitivity—is a performance metric that assesses the percentage of real positive situations that are correctly predicted to be positive. It shows how well the actual positive instances are covered by the positive prediction rule (+P)[19].

Recall= TP/TP+FN

Three universal efficiency metrics are used to compare the voting classifier to the KNN model in Table 1. The metrics are displayed as values in percentages[20].

Table 1: Performance Analysis

Parameters	KNN Classifier	Voting Classifier
Accuracy	60 percent	80 percent
Precision	62 percent	80 percent
Recall	60 percent	80 percent

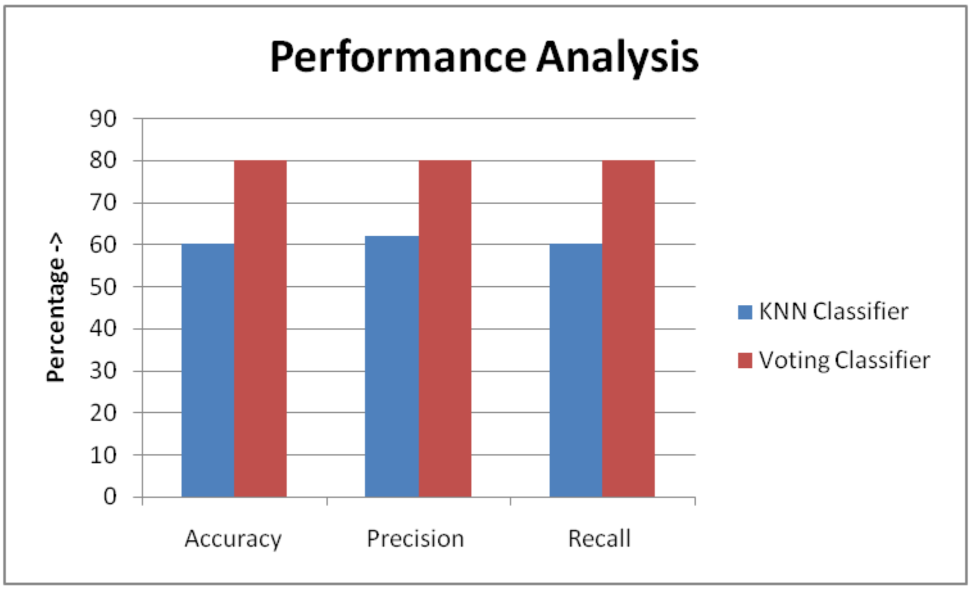

Figure 4 Performance Analysis

The results of the contrast among the proposed methodology, the voting classification algorithm, and the current method, KNN classification, is shown in Figure 4. According to the analysis, the proposed method performed better for forecasting plant illnesses than the current method with respect to of precision, recall, and accuracy.

Conclusion

Digital image processing techniques, coupled with machine learning algorithms, allow plant pathologists to detect diseases from digital photographs of plant leaves. The proposed approach in this work utilizes digital image processing methods and a voting-based architecture for disease detection. Digital cameras have been used to capture the photos, and images are processed to abstract the necessary attributes. In contrast to the currently used technique, which uses the GLCM algorithm for texture characteristics, this work uses k-means clustering. Additionally, the presented method for multi-class classification makes use of a voting-based architecture rather than a KNN classifier. Accuracy, precision, and recall are the three evaluation criteria that are used to gauge how well the given algorithm performs. Comparing with the previous methods, there has been a considerable gain in accuracy and a fall of up to 10% in the false positive rate (FPR).

References

[1] Muthuselvi, S. and Prabhu, P. (2016) ‘Digital Image Processing Techniques: A Survey’, International Multidisciplinary Research Journal, 12(12), pp. 297-306
[2] Kavita, Saroha, R., Bala, R. and Siwach, S. (2013) ‘Review paper on Overview of Image Processing and Image Segmentation’, International Journal of Research in Computer Applications and Robotics, 1(4), pp. 29731-29740
[3] Kulkarni, P.M., Naik, A.N. and Bhadvankar, A.P. (2015) ‘Review Paper on Image Processing Techniques’, International Journal for Scientific Research & Development, 3(10), pp. 113-118
[4] Basavaprasad, B. and Ravi, M. (2014) A Study on the Importance of Image Processing and Its Applications’, International Journal of Research in Engineering and Technology (IJERT), 3(3), pp. 178-188
[5] Mathur, S., Purohit, R. and Vyas, A. (2016) ‘A Review on basics of Digital Image Processing’, International Journal of Engineering Research & Technology (IJERT), 4(12), pp. 93-97
[6] Suresh, V., Gopinath, D., Hemavarthini, M., Jayanthan, K. and Krishnan, M. (2020) ‘Plant Disease Detection using Image Processing’, International Journal of Engineering Research & Technology (IJERT), 9(3), pp. 7166-7172

[7] Kumar, S., and Raghavendra, B. K. (2019) ‘Diseases Detection of Various Plant Leaf Using Image Processing Techniques: A Review’, 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 5(12), pp. 212-220
[8] Saradhambal, G., Dhivya, R., Latha. S. and Rajesh, R. (2018) ‘Plant Disease Detection and its Solution Using Image Classification’, International Conference on Advanced Computing & Communication Systems (ICACCS), 119(14), pp. 879-884
[9] Sujatha, R., Kumar, Y. S. and Akhil, G. U. (2017) ‘Leaf disease detection using image processing’, Journal of Chemical and Pharmaceutical Sciences, 10(1), pp. 230-237
[10] Deshpande, T., Sengupta, S. and Raghuvanshi, K.S. (2014) ‘Grading & Identification of Disease in Pomegranate Leaf and Fruit’, IJCSIT, 5 (3), pp. 4638-4645
[11] Revathi, P. and Hemalatha, M. (2012) ‘Classification of Cotton Leaf Spot Diseases Using Image Processing Edge Detection Techniques’, IEEE International Conference on Emerging Trends in Science, Engineering and Technology (INCOSET), 1(9), pp. 169-173
[12] Gavhale, K. R., Gawande, U. and Hajari, K. O. (2014) ‘Unhealthy Region of Citrus Leaf Detection using Image Processing Techniques’, IEEE International Conference on Convergence of Technology (I2CT), 1(19), pp. 1-6
[13] Jhuria, M., Kumar, A. and Borse, R. (2013) ‘Image processing for smart farming detection of disease and fruit grading’, IEEE 2nd International Conference on Image Information Processing (ICIIP), Shimla, 6(8), pp. 521-526
[14] Sridhathan, C. and Kumar, M. S. (2018) ‘Plant Infection Detection Using Image Processing’, International Journal of Modern Engineering Research (IJMER), 8(7), pp. 23-28
[15] Hashim, H., Haron, M.A., Osman, F.N.S. and Al Junid, A.M. (2010) ‘Classification of Rubber Tree Leaf Disease Using Spectrometer’, Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation, 1(2), pp. 467-471
[16] Pallavi and Marathe, S. (2017) ‘Plant Disease Detection using Digital Image Processing and GSM’, International Journal of Engineering Science and Computing, 8(4), pp. 3-15.
[17] Zhang, S. and Zhang, C. (2013) ‘Orthogonal Locally Discriminant Projection for Classification of Plant Leaf Diseases’, Ninth International Conference on Computational Intelligence and Security, 4(1), pp. 2307-2311
[18] Ramakrishnan M. and Anselin S. N. A. (2015) ‘Groundnut leaf disease detection and classification by using back probagation algorithm’, International Conference on Communications and Signal Processing (ICCSP), 15(3), pp. 6-11
[19] Halder, M., Sarkar, A. and Bahar, H. (2019) ‘Plant Disease Detection by Image Processing: A Literature Review’, SDRP Journal of Food Science & Technology, 3(6), pp. 811-819
[20] Padol, P. B. and Yadav, A. A. (2016) ‘SVM Classifier Based Grape Leaf Disease Detection’, Conference on Advances in Signal Processing (CAPS) Cummins college of Engineering for Women, 8(3), pp. 9-11