Journal of Theoretical and Applied Information Technology
30th April 2023. Vol.101. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

UNHEALTHY PLANT REGION DETECTION IN PLANT LEAVES USING ADAPTIVE GENETIC ALGORITHMS

[1]DR. SURYA PRASADA RAO BORRA, [2]DR.M.V.GANESWARA RAO, [3]DR P.RAVI KUMAR, . [4]DR.SRINU PYLA [5]DR.K. VIDYA SAGAR [6]DR A GEETHA DEVI [7]KOTESWARA RAO KODEPOGU

[16]Dept. of ECE, PVP Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, AP, India
[23]Department of Electronics and Communication Engineering, Shri Vishnu Engineering College for Women, Bhimavaram, AP, India.
[4]Dept. of ECE, Gayatri Vidya Parishad College of Engineering (Autonomous), Visakhapatnam, AP, India,
[5]Department of EIE, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India.
[7]Dept. of CSE, PVP Siddhartha Institute of Technology, Vijayawada, AP,
Email: suryaborra1679@pvpsiddhartha.ac.in , mgr_ganesh@svecw.edu.in ravikumar_tnk@svecw.edu.in, srinupyla@gvpce.ac.in, vidyasagar_k@vnrvjiet.in , koteswara2003@yahoo.co.in

ABSTRACT

The major issue faced by the farmers is crop diseases, which is causing a significant reduction in both the quality and quantity of the yield and hence, it needs to be addressed. The emergence of new and accurate techniques in the field of leaf-based image classification has shown impressive results. This paper shows the use of Random Forest in identifying between healthy and unhealthy leaves from the data sets created. Also, this paper consists of various phases of implementation namely dataset creation, feature extraction, training the classifier, masked cells, neural networks, k-means clustering, SGDM Matrix Generation. The created datasets of diseased and healthy leaves are collectively trained under Random Forest to classify the diseased and healthy images. Overall, using machine learning to train the large datasets available publicly gives us a clear way to detect the diseases present in plants.

Keywords: *Healthy leaf, Random forest, machine learning, k-means clustering, Feature extraction, Training, Classification*

1 INTRODUCTION

Earth is a phenomenal topic on which humans are trying different ways to study and understand it completely in the best possible way. On this Earth, plants are one important part of life. Plants play a huge role in the life of man-kind providing us with oxygen, food, maintaining climatic conditions, medicines and many more. India has agriculture as its backbone since ages wherein most of the population depends on agriculture. Research in agriculture is aimed at the increase of productivity and food quality at reduced expenditure, with increased profit. An agricultural production system is an outcome of a complicated interaction of soil, seed, climatic conditions and agrochemicals.

Vegetables and fruits are the most important agricultural products, which are consumed generally and are always high in demand. To obtain more demandable products, a product quality control is basically mandatory. Many types of research show that the quality and amount of production of agricultural products may be reduced due to plant diseases. [6]

If agricultural development is to be successful, then accurate knowledge of the identity, geographic distribution and uses of plants is essential. Unfortunately, such basic information can be accessed by professional stakeholders, scientists, teachers, and citizens, and often incomplete for the ecosystem that possesses the highest plant diversity. A noticeable consequence such as identifying plant species is usually impossible for the general public, and often a difficult task for farmers or wood exploiters and even for botanists themselves. [7] The only way to overcome this problem is to speed up the collection and integration of raw observed data, while simultaneously providing to potential users in easy and efficient access to this knowledge. So, they can implement it practically for the reduction of plant diseases. Hence, we would like to discuss about different methods used to identify plant diseases and add our perspective to this attention seeking topic.

2. LITERATURE KNOWLEDGE

[1] AAKANKSHA RASTOGI, RITIKA ARORA, and SHANU SHARMA," Leaf Disease Detection and Grading using Computer Vision Technology & Fuzzy Logic". K-means clustering used to segment the defected area; GLCM is used for the extraction of texture features, Fuzzy logic is used for disease grading. They used artificial neural network (ANN) as a classifier which mainly helps to check the severity of the diseased leaf.
[2] UAN TIAN, CHUNJIANG ZHAO, SHENGLIAN LU, and XINYU GUO, SVM-based Multiple Classifier System for Recognition of Wheat Leaf Diseases," Color features are represented in RGB to HIS, by using GLCM, seven invariant moment are taken as shape parameter. They used SVM classifier which has MCS, used for detecting disease in wheat plant offline.
[3] Presents the technique to classify and identify the different disease through which plants are affected. In Indian Economy, a Machine learning based recognition system will proves to be very useful as it saves efforts, money and time too. The approach given in this for feature set extraction is the Color Co-occurrence Method. For automatic detection of diseases in leaves, neural networks are used. The approach proposed can significantly support an accurate detection of leaf, and seems to be important approach, in case of steam, and root diseases, putting fewer efforts in computation.
[4] specifies disease identification process which include some steps out of which four main steps are as follows: first, for the input RGB image, a color transformation structure is taken, and then using a specific threshold value, the green pixels are masked and removed, which is further followed by segmentation process, and for getting useful segments the texture statistics are computed. At last, classifier is used for the features that are extracted to classify the disease. The proposed algorithm shows its efficiency with an accuracy of 94% in successful detection and classification of the examined diseases. The robustness of the proposed algorithm is proved by using experimental results of about 500 plant leaves in a database.
[5] Presents disease detection in malusdomestica through an effective method like K-mean clustering, texture and color analysis. To classify and recognize different agriculture, it uses the texture and color features those generally appear in normal and affected areas. In coming days, for classification purpose given classifier scan also be used, like K-means clustering, Bayes classifier and principal component classifier.

3.METHODOLOGY:

The detailed procedure in the methodology can be listed as

1. RGB image acquisition
2. Create the color transformation structure
3. Convert the color values in RGB to the space specified in the color transformation structure
4. Apply K-means clustering
5. Masking green-pixels
6. Remove the masked cells inside the boundaries of the infected clusters
7. Convert the infected (cluster / clusters) from RGB to HSI Translation
8. SGDM Matrix Generation for H and S
9. Calling the GLCM function to calculate the features
10. Texture Statistics Computation
11. Configuring Neural Networks for Recognition
12. Grading and percentage uses Naïve Bayes system[8]

1. *RGB Image Acquisition*:

The image is taken as input. The image of various plant diseases are considered.[9]

Figure1.Example Input Images:

2. *Colour Transformation Structure*:

First, the RGB images of leaves are converted into Hue Saturation Intensity (HSI) colour space representation. The purpose of the colour space is to facilitate the specification of colours in some standard. HSI (hue, saturation, intensity) colour model is a popular colour model because it is based on human perception. Electromagnetic radiation in the range of wavelengths of about 400 to 700 nanometers is called 'visible light' because the human visual system is sensitive to this range. Hue is generally related to the wavelength of a lightHue, is a colour attribute that refers to the dominant colour as perceived by an observer. Saturation refers to the relative purity or the amount of white light added to hue and intensity refers to the amplitude of the light. Colour spaces can be converted from one space to another easily. After the transformation process, the H component is taken into account for further analysis. S and I are dropped since it does not give extra information.[9]

3. *K-Means*:

All the algorithms we discussed so far are supervised, that is, they assume that labeled training data is available. In many applications this is too much to hope for; labeling may be expensive, error prone, or sometimes impossible. For instance, it is very easy to crawl and collect every page within the www.purdue.edu domain, but rather time consuming to assign a topic to each page based on its contents. In such cases, one has to resort to unsupervised learning. A prototypical unsupervised learning algorithm is K-means, which is clustering algorithm. Given X = {x1, . . . , xm} the goal of K-means is to partition it into k clusters such that each point in a cluster is similar to points from its own cluster than with points from some other cluster.[10]

The algorithm stops when the cluster assignments do not change significantly.

K-MEANS CLUSTERING:

The leaves are partitioned into 4 different clusters .It is useful when number of classes are known prior. The K-means extract and places the pixel information in the clusters. One of the cluster contains the infected leaf.

ALGORITHM: K-means clustering
INPUT: Leaf images
OUTPUT: 4 clusters, with one as infected cluster.[9]

- Let features of image is denoted as $X=\{x_1,x_2,\ldots\ldots.x_n\}$ are data and among these some of them are centroids.
- Initialize cluster k=4
- Randomly select ‘c’ centers
- Calculate the center using Euclidean distance formula for each data point x1,x2……and so on.
- Euclidean distance: $\sqrt{(x_1-y_1)^2+(x_2-y_2)^2}$
- Assign the data points who are nearer to the centers.
- Calculate the new centers
- New centers$(x,y)=(x_1-y_1)/2, (x_2-y_2)/2$
- Repeat the steps until no centroid position are changes.
- Otherwise proceed in calculation on Euclidean distance.

4. *Masking green pixels*:

In this step, we identify the mostly green colored pixels. After that, based on specified and varying threshold value that is computed for these pixels using Otsu's method , these mostly green pixels are masked as follows: if the green component of pixel intensities is less than the pre-computed threshold value, the red, green and blue components of the this pixel is assigned to a value of zero. This is done in sense that these pixels have no valuable weight to the disease identification and classification steps, and most probably those pixels represent healthy areas in the leave. Furthermore, the image processing time should become significantly reduced. In next step the pixels with zero red, green and blue values were completely removed. This phase is helpful as it gives more accurate disease classification and identification results with satisfied performance and the overall computation time should become significantly less.[11]

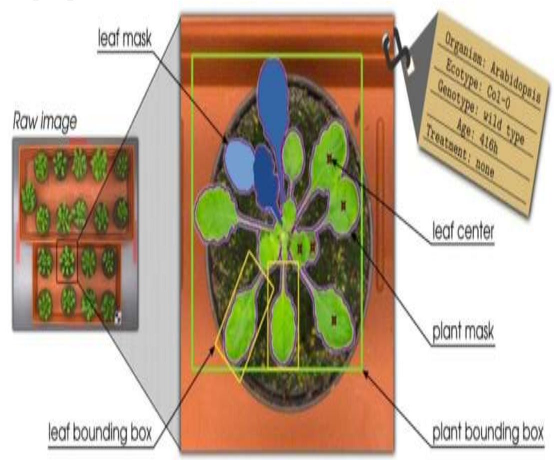

Figure2. Raw Images And Transformation

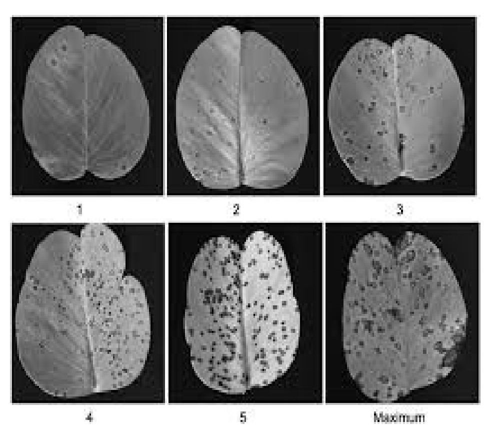

Figure3. Gray Scale Image

5. *Colour Co-occurrence Method*:

In the proposed approach, the method adopted for extracting the feature set is called the Colour Co-occurrence Method or CCM method in short. It is a method, in which both the colour and texture of an image are taken into account, to arrive at unique features, which represent that image. The image analysis technique selected for this study was the CCM method. The use of colour image features in the visible light spectrum provides additional image characteristic features over the traditional gray-scale representation. The colour co-occurrence texture analysis method was developed through the use of Spatial Gray-level Dependence Matrices [SGDM]. The gray level co-occurrence methodology is a statistical way to describe shape by statistically sampling the way certain grey-levels occur in relation to other grey-levels. These matrices measure the probability that a pixel at one particular gray level will occur at distinct distance and orientation from any pixel given that pixel has a second particular gray level. For a position operator 'p', we can define a matrix 'P_{ij}' that counts the number of times a pixel with grey level 'I' occurs at position 'p' from a pixel with grey-level 'j'. The SGDMs are represented by the function P(i, j, d, Ѳ), where 'i' represents the gray level of the location (x, y) in the image I(x, y), and 'j' represents the gray level of the pixel at a distance 'd' from location (x, y) at an orientation angle of Ѳ. Here we calculate feature set for H component only. However, we use GLCM function in Matlab to create gray-level co-occurrence matrix.[11]

6. *Neural Network Recognition:*

The neural network is used to find the infected leaves and its type. ANN model can effectively process based on input parameters and produce output.[8]

LIMITATIONS:

1. Genetic algorithm for segmentation: Efficiency and time of the process depends upon the initial generated population of chromosomes.
2. K-means clustering techniques: Guaranteed to converge, to reduce the number of false edges.
3. Support Vector Machine (SVM): Lack of transparency in the result for high dimension data.
4. Decision Tree Classifier (DTC): Instability, over fitting, unstable in small variations, cannot guarantee to achieve the globally optimal decision tree.

4.CONCLUSION:

The main approach of this system is to recognize the diseases on the different plant in agriculture environment where speed and accuracy are the main characteristics of disease detection. Hence, the extension of this work will focus on developing better segmentation technique, selecting better feature extraction, and developing hybrid algorithms such as genetic algorithms; in order to increase the recognition rate of the final classification process underscoring the advantages of hybrid algorithms. Also, we would like dedicate our future works on automatically estimating the severity of the plant disease.

After reviewing all above-mentioned techniques and methods, we can conclude that there is the number of ways by which we can detect disease of plants. Each has some advantages as well as limitations. Therefore, there is the scope of improvement in the existing

REFERENCES:

[1] AAKANKSHA RASTOGI, RITIKA ARORA, and SHANU SHARMA," Leaf Disease Detection and Grading using Computer Vision Technology &Fuzzy Logic" 2nd International Conference on Signal Processing and Integrated Networks (SPIN)2015

[2]UAN TIAN, CHUNJIANG ZHAO, SHENGLIAN LU, and XINYU GUO," SVM-based Multiple Classifier System for Recognition of Wheat Leaf Diseases," Proceedings of 2010 Conference on Dependable Computing (CDC'2010), November 20-22, 2010.

[3] MRUNALINI R.BADNAKHE and PRASHANT R.DESHMUKH," An Application of K-Means Clustering and Artificial Intelligence in Pattern Recognition for Crop Diseases", International Conference on Advancements in Information Technology 2011 IPCSIT vol.20 (2011).

[4] S. ARIVAZHAGAN, R. NEWLINE SHEBIAH, S. ANANTHI, S. VISHNU VARTHINI "Detection of the unhealthy region of plant leaves and classification of plant leaf diseases using texture features" Agric Eng : CIGRJournal,15(1): 2110217 2013.

[5] SABAH BASHIR, NAVDEEP SHARMA , " Remote Area Plant Disease Detection Using Image Processing", IOSR Journal of Electronics

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and Communication Engineering, ISSN: 2278-2834 Volume 2, Issue 62012, PP 31-34.

[6] MS.KIRAN R. GAVHALE, PROF. UJWALLA GAWANDE. "An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques", IOSR Journal of Computer Engineering (IOSR-JCE) Volume 16, Issue 1, Ver. V (Jan. 2014), PP 10-16.

[7] Introduction to Machine Learning Alex SMOLA and S.V.N. VISHWANATHAM Yahoo! Labs Santa Clara – and– Departments of Statistics and Computer Science Purdue University –and– College of Engineering and Computer Science Australian National University.

[8]https://www.researchgate.net/publication/287577015_Detection_of_unhealthy_region_of_plant_leaves_and_classification_of_plant_leaf_diseases_using_texture_features