RESEARCH ARTICLE OPEN ACCESS

PLANT LEAF SHAPE CLUSTERING AND ROT DETECTION

Ms. N. Zahira Jahan MCA., M.Phil.,*, P. Karuppusamy**
* Associate Professor, Department of MCA, Nandha Engineering College, (Autonomous), Erode, Tamilnadu, India.
Email: zahirajahan1977@gmail.com
**Final MCA, Department of MCA, Nandha Engineering College (Autonomous), Erode, Tamilnadu, India.
Email: karuppusamy58010@gmail.com

****

Abstract:

Leaf disease detection plays a vital role in agricultural field. However, it requires huge manpower, more processing time and extensive knowledge about plant diseases. Hence, machine learning is applied to detect diseases in plant leaves as it analyses the data from different aspects, and classifies it into one of the predefined set of classes. The first effort in learning about plants is observing plant features. This project developed a plant search system that allows users to do a search even when they do not know the plant name simply by observing plant characteristics. The system consists of a plant features, searches for the features according to the input features, and returns the leaves with selected clusters. At present, leaf classification uses machine vision to extract and analyse colour, size, shape, and surface texture. However, the proposed extraction margin method can only be carried out roughly and there is still a difference between the margin of the extracted shape, polygon, and the margin of the shape of the original image. This project clusters the leaves using image area size, pixel colour values similarity, based on brightness values of the image and leaf shapes. In addition, the project aims in finding the rots in the leaves. Based on the count of pixels of rot colours, the total rot percent in the leaf is calculated and displayed. This assists in evaluating the leaf quality. If future researchers were to expand to other features, leaf apex, etc., even those that are hard to quantify, can also be quantified. The project is designed using R Studio 1.0. The coding language used is R 3.4.4.

Keywords ——***Artificial Neural Network, Classification, Disease Detection, Support Vector Machine, Machine Learning.***

****

I.INTRODUCTION

Currently, most plant search systems use RFID or QR Code to obtain plant information. However, these two technologies have specific hardware requirements that must be setup in advance to carry out scanning and sensing through a mobile phone to find the name of the plant. Since the variety of plants is very diverse and covers a wide range, these two methods have regional limitations.

Some researchers have attempted to classify different tobacco leaves using the Fuzzy Function. Their auto-inspecting and grading system uses machine vision to extract and analyse colour, size, shape, and surface texture. However, the proposed extraction margin method can only be carried out roughly and there is still a difference between the margin of the extracted shape, polygon, and the margin of the shape of the original image. Therefore, to improve the method for capturing the plant outline, this study proposes a Centroid-Contour distance to capture the outline of the plant features and the distance from the center point to each margin point to more accurately quantify the plant features of the original image.

As a result, the captured image can also be consistent with the original plant image. Since image recognition technology for quantifying three dimensional features, such as outlines of flowers, is difficult, and the accuracy of the quantified value cannot be verified, the accuracy of the feature search query is definitely impacted and thus cannot be performed. Therefore, this project applied the Association Rule method to those plant features that cannot be accurately quantified.

The method of associating similarity with the association rule analysis can effectively improve the fault-tolerance and accuracy of the overall systematic plant search query. This study aims to develop a plant search system with high tolerance and accuracy, as well as prove that the Association Rule can effectually complement the shortcomings of the inability to quantify features and further improve the fault-tolerance and accuracy of the search query system.

The advantage of this system is that it allows users to easily search information of a plant without knowing the plant's name. Furthermore, the science curriculum can combine the teaching strategy of inquiry-based learning with this study's plant search system to improve elementary school students' plant observation ability. Students can thus pursue learning about plants through an approach based on self-observation. In this study, the plant search was implemented based on the plant features using the approach of Centroid-Contour distance in addition to the Fuzzy Function calculation.

However, the flowers of the plant are difficult to be quantified with various features which may lead to biased tolerances. That's why the study has incorporated Association Rule analysis by manually collecting feedback on similarities among the plant features, and applied the Association Rule to identify the similarity rules by species of plant as the supplemental calculation for the feature similarity and increase the overall tolerance and accuracy of the system search. Also, three different methods were applied to identify three different accuracy rates in the study.

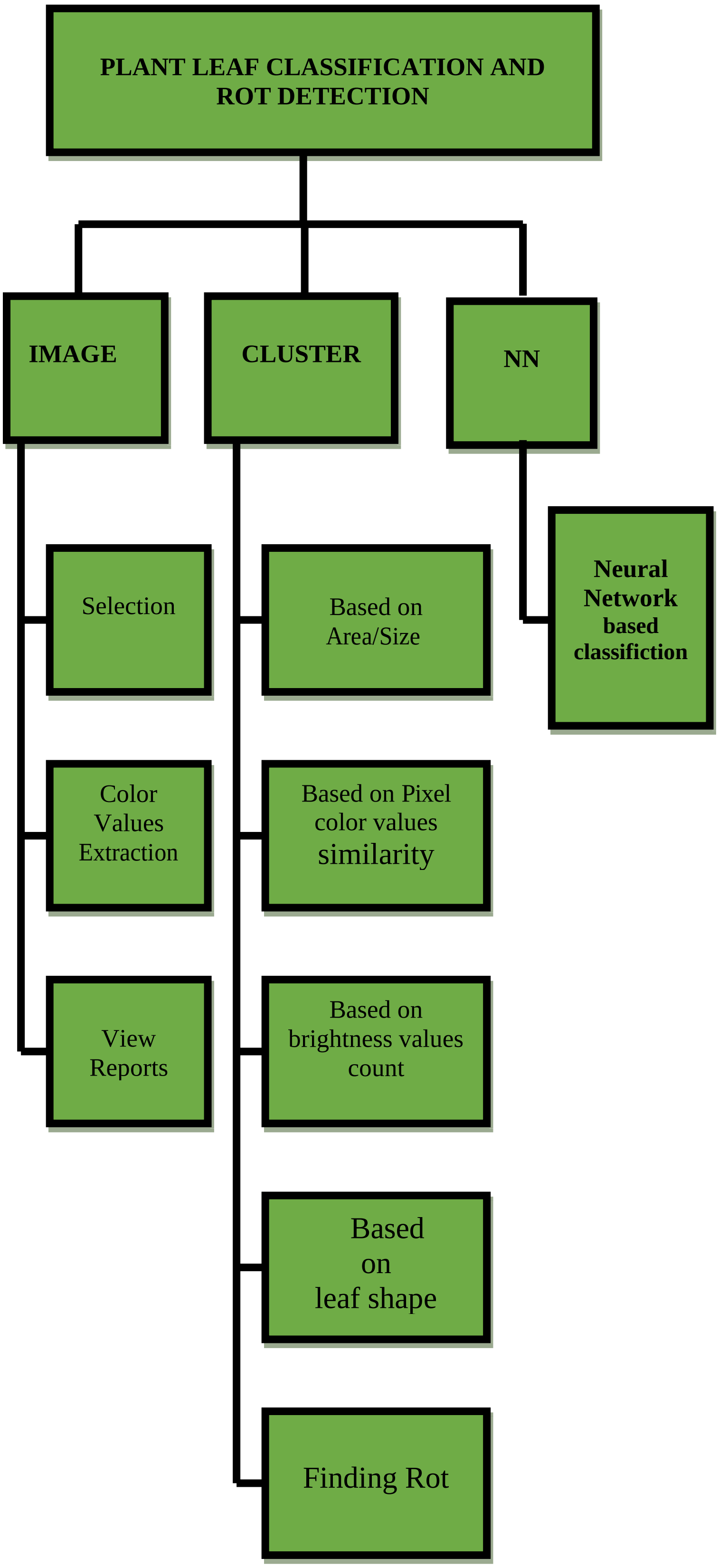
The Association Rule analysis would not only effectively increase the tolerance and accuracy of the system but also help users to search unknown plants based on the observed features of the targeted plants. After the features input, the system would calculate and screen out the Top 10 plants with the highest similarities to show their names and related information. In this study, a plant search query system was developed to perform three different analyses of similarity, namely the Fuzzy Function, the Association Rule, and the similarity from the combined Fuzzy Function and the Association Rule and performed for single feature error conditions and n feature error conditions.

II. CLASSIFICATION OF PLANT DISEASES

The plant leaves are full of bacterial, fungal and viral diseases which include leaf rust, mildew, bacterial blight, Downey mildew, brown spot etc. Fig. 1 illustrates the classification of the bacterial, fungal and viral diseases. Artificial Neural Network, Probabilistic Neural Network, and Support Vector Machine for vegetable crops, commercial crops and cereal crops respectively for disease detection proposed an algorithm for classifying diseases of plants including jackfruit, tomato, etc. by using

Support Vector Machine classifier .Table I demonstrates the comparative study of disease detection in different plant leaves.

III.SYSTEM FLOW DIAGRAM



IV. CLASSIFICATION ALGORITHMS

This section explains the classification algorithms in machine learning that are used for classifying diseases [1] in plant leaves. Its accuracy depends on the amount of samples taken and varies in line with the classification algorithms used. The classification algorithms are divided into supervised and unsupervised classification algorithms [2-4] Fig. 2 illustrates the assorted forms of classification algorithms for plant plant disease detection.

A. Unsupervised Classification Algorithms

Fuzzy C-means is an iterative algorithm which helps to seek out the cluster centres that minimize a dissimilarity function and to handle the overlapped data efficiently. It gives better ends up in cases where data is incomplete or uncertain, but computation time is longer and it's sensitivity to noise. Fuzzy C-means clustering Neural Network [5] consists of unsupervised fuzzy clustering and supervised artificial neural networks which help in achieving more optimal results with relatively few data sets. K-means is an iterative learning helps find the cluster centres for every group and has no guarantee for optimum solution. It is easy to implement and computationally faster. But the quantity of cluster prediction is difficult. Principal Component Analysis [6] is an unsupervised technique helps find the foremost accurate data representation and maximizes the variance. Linear Discriminant Analysis finds the projection to a line and maximizes the component axes for sophistication separation

B. Supervised Classification Algorithms

K Nearest Neighbour may be a used for statistical estimation and pattern recognition. It is easy, simple, flexible and robust to noisy training data but computation cost is higher. Artificial Neural Network uses forward propagation which is that the heart of a neural network. Probabilistic Neural Network may be a feed forward algorithm which is extremely faster and more accurate than multilayer perceptron network. Generalized Regression Neural Network could be a supervised algorithm used for classification. Convolution Neural Network could be a class of deep, feed forward Artificial Neural Network which consists of input, output further as multiple hidden layers, convolutional layers, pooling layers, fully connected layers and normalization layers. Pooling reduces the dimensionality of the features map by condensing [6] the output of small regions of neurons into one output. Fuzzy-Relevance Vector Machine is effective in handling unbalanced data and reducing the results of noise or outliers. Relevance Vector Machine may be a machine learning technique that uses Bayesian inference for regression and probabilistic classification. In Support Vector Machine, the information points with maximum margin are chosen and separated by hyper plane. Radial Basis Function has three layers namely, input, hidden and output layer. It's used for function approximation which depends only on the space from the origin. Random Forest performs both classification and regression which is additionally referred to as ensemble machine learning algorithm. It adapts divide and conquer approach accustomed improve the performance. It creates a forest with more decision trees which helps in strong prediction. Decision trees otherwise referred to as Classification and Regression Trees are used, when dependent variable is continuous. Table II demonstrates the performance of classification algorithms for plant leaf disease detection.

V. METHODOLOGY

A. Image Selection

In this module, the betel leafs images are submitted to the system as input. The leaf brown spot diseased leaves are taken in this module. Images are taken in controlled environment and are stored in the JPEG format. These are submitted to the clustering processes as well as rot identification.

B. Colour Values Extraction

In this module, the betel leaves images pixels colours are retrieved with their red, green and blue components values. They are used in pixel colour value similarity checking module.

C. Cluster Based On Area Size

In this module, the betel leaves images dimensions are retrieved using dim() from ‘bmp’ library and are saved in a vector. The minimum and maximum area size are found out and the difference is divided by three to get range size. Then range1 is calculated as minimum value and ‘minimum value’ plus ‘range size’. Then range2 is calculated as ‘minimum value’ plus ‘range size’ and ‘minimum value’ plus 2 * ‘range size’. Then range3 is calculated as minimum value and ‘minimum value’ plus 3 * ‘range size’. Then the sizes fall within these ranges are clustered in their respective clusters.

D. Cluster Based On Pixel Colour Values Similarity

In this module, the betel leaves images colours are retrieved and are saved in a vector. Then an image is compared with other image pixels' red, green and blue component values. If matched, then the variable is incremented. Then the total similarity count is stored. This logic repeated for all the images. Then the similarity percent fell within 33% are grouped in one cluster, within 66% are grouped as next cluster and the remaining as third cluster.

E. Cluster Based On Brightness Values Count

In this module, the betel leaves images brightness values are retrieved and are saved in a vector. Then an image is compared with other image pixels' brightness values. If matched, then the variable is incremented. Then the total similarity count is stored. This logic repeated for all the images. Then the similarity percent fell within 33% are grouped in one cluster, within 66% are grouped as next cluster and the remaining as third cluster.

F. Cluster Based On Leaf Shape

In this module, the betel leaves images distance total from midpoint to all the edge pixles are calculated and are saved in a vector. Then an image is compared with other image distance total values. The unique distance values are used for clustering the images with shape as feature.

G. Rot Area Calculation

In this module, the betel leaves images grayscale values are calculated and are saved in a vector. Then an image is compared with whiter area pixels and black area pixels. Then the percent of white area is found with total black area pixels to find the rot percent of the leaf image

H. Neural Network Based Classification

The plant feature component is used to extracts plants' hidden features based on plant's parameters. The parameters can be any of width, height, area, leaf shape, average brightness, number of rots, rot area, etc. There are two field types in our model. Linear fields are the fields with continuous numeric values. The values of the linear fields are scaled into [0–1]. One-Hot fields are the fields with enumeration values. The field values are encoded as one-hot vectors separately. We added the one-hot vectors together in case a leaf has multiple values, for example, a leaf may have several artists. The concatenation result of all the fields is the meta vector of the corresponding leaf. Assuming a plant's width is wid(s), it has ng height types, na shape types, nc rotcount and nl rotarea, the meta vector of this plant M(s) is calculated as Equation (1).

$$M(s) = [len(s); (g1(s) +...+ gng(s)); (a1(s) +...+ ana(s)); (c1(s) +...+cnc(s)); (l1(s) +...+lnl(s))] \quad (1).$$

Where M(s) is the metavector of s, g1 is the one hot vector of the first height type, is the concatenation operation.

A Deep Neural Network (DNN) is used to extract hidden features from the meta vector M(s). We calculate the output of layer i as Equation (2).

$$li = ReLU(Wi \cdot li-1 + bi) \quad (2).$$

where li is the output of the ith layer, i = 1,...N, l0 is the input of the DNN, Wi is the weights of the neurons in ith layer and bi is the bias, lN is the output of the DNN and ReLU is the activation function. In DTNMR, the input vector length of the DNN is about 1,000. The DNN consists of four layers; the first layer is the input layer, the second is a hidden layer with 512 nodes, the third layer contains 64 nodes, and the output layer contains 32 output nodes. The layer sizes are chosen empirically. The output item vector is denoted as S(s) = lN. After the neural network is trained the new weight and bias values are noted for future classification process.

Overall, all the available plant leaves' parameter details are feed into the neural network. Then if a new leaf's parameters is given, the leaf matching in the existing database is found out.

VI. CONCLUSION

This project developed a plant search system that applied some features and cluster them using adaptive approach. The first method used was the Centroid-Contour distance to quantify some features and combine the Fuzzy Function theory; the second method used was similarity based on area, brightness values, and pixel colour similarities. The method with high efficiency can be applied for other search query systems besides plants and can be easily applied to other systems with the concept of quantifying feature similarities for applications.

It is believed that almost all the system objectives that have been planned at the commencements of the software development have been net with and the implementation process of the project is completed. A trial run of the system has been made and is giving good results the procedures for processing is simple and regular order. The process of preparing plans been missed out which might be considered for further modification of the application. In future, this project may find the similarity using rots present in the leaves.

VII. REFERENCES

[1] Santanu Phadikar and Jaya Sil, "Rice disease identification using pattern recognition techniques," 11th IEEE International Conference on Computer and Information Technology, pp. 420-423, 2008.

[2] Sun, Yu, Yuan Liu, Guan Wang, and Haiyan Zhang, "Deep learning for plant identification in natural environment," Computational intelligence and neuroscience, vol. 2017, 2017.

[3] John William Orillo, Gideon Joseph Emperador, Mark Geocel Gasgonia, Marifel Parpan, and Jessica Yang, "Rice plant nitrogen level assessment through image processing using artificial neural network," IEEE International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), pp. 1-6. 2014.

[4] Mukherjee, Gunjan, Arpitam Chatterjee and Bipan Tudu, "Morphological feature based maturity level identification of Kalmegh and Tulsi leaves," Third IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), pp. 1-5, 2017 M. C. Ramya, V. Lokesh V, T. N. Manjunath and R. S. Hegadi, "A Predictive Model Construction for Mulberry Crop Productivity," International Conference on Advanced Computing Technologies and Applications (ICACTA- 2015).

[5] Siddharth Singh Chouhan, Ajay Kaul, Uday Pratap Singh and Sangeev Jain, "Bacterial Foraging Optimization Based Radial Basis Function Neural Network (BRBFNN) for identification and classification of plant leaf disease: An automatic approach towards plant pathology," IEEE Access, vol. 6, pp. 8852-8863, 2018.

[6] X. E. Pantazi, D.Moshou, and A. A. Tamouridou, "Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers," Computers and Electronics in Agriculture, vol. 156, pp. 96-104, 2019.