ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2021.0346

PLANT DISEASE RECOGNITION AND CLUSTERING USING FUZZY ALGORITHM ON DATA MINING

R. Sabitha[1] and G. Kiruthiga[2]

[1]*Department of Electronics and Communication Engineering, Hindustan College of Engineering and Technology, India*
[2]*Department of Computer Science and Engineering, IES College of Engineering, India*

Abstract

Due to large size and intensive processing needs, deep learning models are not suited for mobile and handheld devices. Our goal is to develop a process that begins with pre-processing, diagnoses diseased leaf areas, uses the GLCM to choose and classify features, and culminates in a conclusion. We developed fuzzy decision methods for assigning photos of common rust to various severity levels, using data on diseased leaf regions isolated by threshold segmentation. The outcomes of these experiments were determined by six different colour and texture attributes. In plant disease clustering, the Fuzzy Algorithm is utilised. The test results demonstrate that the new method is more efficient than the conventional approaches and ranks first for feature extraction techniques. This appears to say that plant disease diagnosis using leaves should be utilised. Additional disease classifications or crop/disease classifications can be added to define these capabilities.

Keywords:

Plant Disease, Plant Leaf, Recognition, Clustering

1. INTRODUCTION

As researchers who have so far utilised deep learning models to identify plant leaf disease severity have done, they employed training datasets that were organised using human decisions and observations, and in these datasets, researchers only observed classifications of leaf disease levels. In this procedure, biases and unreliability are possible because of factors such as eyesight impairment.

The VGG-16 is employed in identifying apple diseases [1]. In their research, they employed a botanist to determine how far the sick apple leaves had progressed. To assess the various stages of a disease outbreak, the researchers gathered specimens from leaves in healthy, early, middle, and late stages of disease progression. Following the instructions of a botanist, the network was trained on sets of labelled data that were utilised to train the VGG-16 model, which had an accuracy of 90.4 percent.

However, in this study, a novel approach that utilises computerised fuzzy decision rules was used for assigning images of the maize Common Rust to their severity classes of data sets that were used to train the VGG-16 network [2].

Using percentages of sick leaf area as fuzzy decision rules for assignment to respective severity classes, the suggested approach uses thresholding images of diseased maize and then derives classifications for various severity levels. Once the severity classes were created, we trained the VGG-16 network to differentiate among four disease severity classes and evaluated maize common rust images [3].

Due to advancements in both technology and AI, scientists and researchers are now able to use deep learning convolutional neural networks to perform plant disease identification. Many artificial intelligence applications have used the machine learning method known as deep learning. Convolutional neural networks were used to construct an abnormal breast identification model which included nine layers. Despite this, for the sake of this study, studies have been done on the use of deep learning for disease identification that has already been done after photos have been evaluated.

The updated Faster R-CNN architecture was built by modifying the parameters of a CNN model and an Automatic Detection of Leaf Spot Disease in Sugar Beet model, which were both altered to improve performance. In order to test and evaluate the imaging-based expert systems method for detecting disease severity, this method was trained and evaluated with 155 images. The results show that the overall correct classification rate was 95.48%.

A vision-based method was created to detect indications of Olea europaea infection by Xylella fastidiosa, which is how leaf scorch was discovered [4]. As part of this project, the algorithm was able to identify low-level traits that originated from raw data to determine the presence of veins and colours that would indicate the presence of symptomatic leaves.

The model was built utilising a convolutional neural network (which is a type of neural network that is often trained using stochastic gradient descent) that was created and trained using the stochastic gradient descent approach. According to literature research, model parameter tweaking and regularisation strategies have an impact on the accuracy of deep learning models that are employed for plant disease diagnosis.

The authors in [5] employed pre-trained image recognition networks to be used as training material for Image Net, a notable dataset with tagged images. With this strategy, results improved versus other state-of-the-art procedures, and validation accuracy improved to at least 91.83 percent. The team's proposal to process the rice plant class of photos with a complicated background brought about typical performance results with a precision of 92.00 percent.

Deep learning was applied to the diagnosis of potato tuber disease by Oppenheim et al. [6]. The VGG model used numerous new dropout layers in order to improve overfitting while working with a limited dataset. The accurate VGG network needs an input image size of 224×224. CNN layers contained eight layers of convolutional ones, with the first five of them using convolutional operations and the final three using fully connected operations, and finishing with a softmax layer. To effectively train CNNs, a huge amount of tagged data is required. The strategies utilised for data augmentation were as follows: Mirroring resulted in extra examples by randomly flipping images in training; while cropping helped to achieve data diversity by cropping images randomly to different sizes, with the minimum dimension remaining at 190.

2. METHODS

The model presented in [7] [8] helped resolve the current issues associated with using deep learning for disease detection and classification. To deal with the issue of data augmentation, they applied two ways. The first instance was the use of traditional ways to enhance the original material, while the second instance was about the artistic style of Generative Adversarial Networks (GANs).

A change in the colour, form, or functionality of the plant is one of the main symptoms of plant disease. The leaves' green pigment has been lost due to this change in colour. A feasible technique for treating this data set's Common Rust photos was employed using the Otsu threshold-segmentation method, which was utilised to identify the sick leaf area in percentages, and then two fuzzy decision rules were applied to assign images of Common Rust severity classes.

In order to apply the developed severity classifications to training a VGG16 neural network to categorise test photos of the Common Rust illness, the network was first trained on images with different degrees of severity. In basic terms, Image Analyzer is open-source software that provides above-average analysis, editing, and optimization capabilities for photos. Since it includes several online repositories, the PlantVillage dataset has been frequently used by machine learning researchers.

We utilised an open-source tool named "Image Analyzer" to do the Otsu thresholding on images. The Otsu thresholding threshold is based on the assumption that the image contains two distinct groups of pixels, and hence the Otsu global threshold value of 127 is applied across all situations. The Fig.1 gives a step-by-step explanation of our technique.

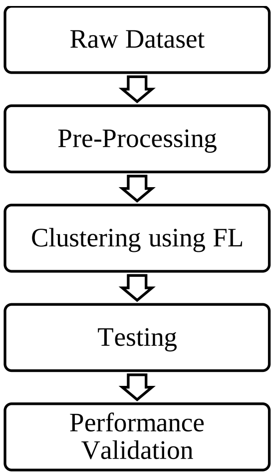


Fig.1. Proposed Methodology

2.1 FUZZY LOGIC CLUSTERING

Fuzzy logic helps when it comes to being flexible when it comes to reasoning. The method resembles human reasoning. Using the way people do decision-making, this technique applies intermediate values between "Yes" and "No." In many applications, fuzzy logic reasoning delivers acceptable reasoning and helps engineers confront the ambiguity that they face in their work. Every rule and if-then condition included in the fuzzy logic architecture was supplied by human experts to govern the decision-making mechanism.

The Otsu technique is a thresholding approach that employs the concept of minimising variation within each class. Adjusting the threshold alters the two portions of the distribution, but distributions are fixed because of their fundamental characteristics. For starters, choose a threshold that keeps the total dispersion to a minimum.

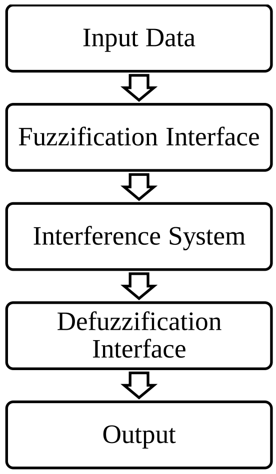


Fig.2. Fuzzy Logic Modelling

The variance within each class is equal to the weighted sum of each class variances.

$$\mu_w^2(T) = n_b(T)\mu_B^2(T) + n_F(T)\mu_F^2(T) \tag{1}$$

where

$$n_b(T) = \sum_{i=0}^{T-1} p(i)$$

$$n_F(T) = \sum_{i=T}^{T-1} p(i)$$

$\mu_B^2(T)$ = Background pixel variance and

$\mu_F^2(T)$ = Foreground pixel variance

The computation of the within-class variance and the various thresholding values are not inexpensive, and they must be avoided. In order to save computation time, it is beneficial to redefine the calculation of between-class variance, which is a less expensive step, as the total minus the between-class variance.

$$\mu_B^2(T) = \mu^2 - \mu_W^2(T) \tag{2}$$

where μ^2 is the combined variance and e is the combined mean. When taking into account all classes, the class-specific variance is the weighted sum of the cluster means, which in turn accounts for the overall mean. It is a simple process to substitute

$$e = n_b(T)e_B(T) + n_o(T)e_o(T)$$

and simplifying the result, we get

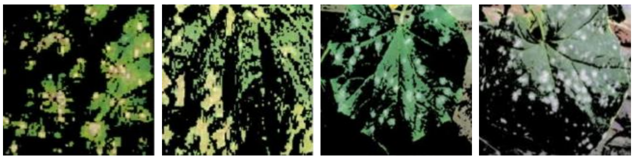
$$\mu_B^2(T) = n_b(T)n_o(T)\left[e_b(T) - e_o(T)\right]^2 \tag{3}$$

As each threshold is passed, the programme splits the pixels into two clusters, with each cluster containing pixels whose values meet or exceed the threshold.

3. PERFORMANCE EVALUATION

This section provides the results of performance evaluation between the proposed and existing methods. The simulation is conducted in core i7 processor with 16GB RAM and it is validated in Python simulator. A total of 800 images are clustered and validation is reported in Table.1 and the result shown in Fig 3.

(a) Original Regions



(b) Clustered Regions

Fig.3. Images used for clustering validation

Table.1. Simulation Results of Various Parameters on different image datasets for testing

Images	K-means			Fuzzy Logic		
	Precision	Recall	F1-score	Precision	Recall	f1-Score
100	0.8358	0.8598	0.8448	0.7877	0.8117	0.7977
200	0.6464	0.5702	0.6003	0.6975	0.6133	0.6504
300	0.7746	0.8047	0.7877	0.7275	0.7987	0.7606
400	0.8608	0.7957	0.8247	0.7806	0.7857	0.7806
500	0.8247	0.8648	0.8388	0.9129	0.6373	0.7446
600	0.8358	0.9139	0.8718	0.8057	0.8849	0.8398
700	0.9710	0.9510	0.9580	0.9981	0.9310	0.9620
800	0.8809	0.8448	0.8608	0.9320	0.9500	0.9390

Table.2. Simulation Results of Various Parameters on different image datasets for Validation

Images	K-means			Fuzzy Logic		
	Precision	Recall	F1-score	Precision	Recall	f1-Score
100	0.7786	0.8097	0.7927	0.5682	0.4570	0.4930
200	0.5782	0.5922	0.5822	0.5492	0.4580	0.4960
300	0.7696	0.7165	0.7376	0.6524	0.7546	0.6985
400	0.7245	0.7776	0.7476	0.6123	0.7396	0.6604
500	0.8809	0.6273	0.7295	0.3257	0.1834	0.2225
600	0.8618	0.8588	0.8588	0.6664	0.7987	0.7245
700	0.7335	0.9791	0.8368	0.8919	0.8899	0.8859
800	0.7867	0.5131	0.6053	0.6634	0.4810	0.5492

The Fuzzy algorithm and K-means clustering are both utilised in the trials presented in Table.1 and Table.2. The first step is to decide on the parameters used to start. But, because the distance measurements in the two algorithms are calculated in different modules, the results are guaranteed to be different. The superpixel-based method is far more effective than the K-means-based disease leaf image segmentation method. Using the squared Euclidean distance, K-means clustering uses the Euclidean distance, and picking a reference colour for comparison parameters is employed while implementing the Fuzzy algorithm. Instead, the Fuzzy method calculates the likelihood of each pixel in the observation falling into each cluster by utilising a mixture of two-Gaussian distribution, and the goal is to locate the clusters that optimise the total probability of the data. The Fuzzy algorithm's initial parameters, which are determined in compact superpixels, suppress noise to some extent, while segmentation methods that employ a random starting generally are not robust to colour picture segmentation problems.

Due to randomly determining the starting parameters of the Fuzzy algorithm for picture segmentation, Fuzzy segmentation is the slowest. The study indicates that each superpixel is comprised of pixels that retain a good boundary to some extent, according to their distinct directional and colour properties.

Superpixels have less of the inherent complications seen in Fuzzy image processing. Compared with the Fuzzy method, the image is divided into K superpixels, and the iteration number increases while the computing time lowers. Because the Fuzzy algorithm has the capability of swiftly converging into a superpixel, that is why. However, raising K will lead to over-segmentation. Because of this, computing the time for superpixels with Fuzzy has become less, but also faster.

4. CONCLUSION

To solve this problem, we've developed an automated approach for plants that includes feature detection, feature selection, disease separation, feature classification, and decision making. In order to extract the infected leaf portions from the maize plants, we utilised threshold segmentation, and then we created fuzzy decision rules using this information for assigning photos of Common Rust to their severity class. The data presented in this experiment utilised six colour and texture features. Fuzzy Plant Disease Clustering is used to execute these operations. As shown by the measurements, the conventional approaches extract more features. This is the best method for diagnosing plant diseases due to the inclusion of leaves in the analysis.

REFERENCES

[1] N.V. Kousik, M. Sivaram and R. Mahaveerakannan, "Improved Density-Based Learning to Cluster for User Web Log in Data Mining", *Proceedings of International Conference on Inventive Computation and Information Technologies*, pp. 813-830, 2021.

[2] H. Azath, M. Mohanapriya and S. Rajalakshmi, "Software Effort Estimation using Modified Fuzzy C Means Clustering and Hybrid ABC-MCS Optimization in Neural Network", *Journal of Intelligent Systems*, Vol. 29, No. 1, pp. 251-263, 2018.

[3] H. Azath and R.S.D. Wahidabanu, "Function Point: A Quality Loom for the Effort Assessment of Software Systems", *International Journal of Computer Science and Network Security*, Vol. 8, No. 12, pp. 321-328, 2008.

[4] N.V. Kousik, "Privacy Preservation between Privacy and Utility using ECC-based PSO Algorithm", *Proceedings of*

International Conference on Intelligent Computing and Applications, pp. 567-573, 2021.

[5] S. Karthick and P.A. Rajakumari, "Ensemble Similarity Clustering Frame work for Categorical Dataset Clustering Using Swarm Intelligence", *Proceedings of International Conference on Intelligent Computing and Applications*, pp. 549-557, 2021.

[6] G. Kiruthiga, G.U. Devi and N.V. Kousik, "Analysis of Hybrid Deep Neural Networks with Mobile Agents for Traffic Management in Vehicular Adhoc Networks", *Proceedings of International Conference on Distributed Artificial Intelligence*, pp. 277-290, 2020.

[7] K.M. Baalamurugan and S.V. Bhanu, "An Efficient Clustering Scheme for Cloud Computing Problems using Metaheuristic Algorithms", *Cluster Computing*, Vol. 22, No. 5, pp. 12917-12927, 2019.

[8] K.M. Baalamurugan and S.V. Bhanu, "Analysis of Cloud Storage Issues in Distributed Cloud Data Centres by Parameter Improved Particle Swarm Optimization (PIPSO) Algorithm", *International Journal on Future Revolution in Computer Science and Communication Engineering*, Vol. 4, pp. 303-307, 2018.