Disease detection in plant leaves using segmentation and autoencoder techniques

Vijayabharathi Durairaj¹* and Chellammal Surianarayanan²

Abstract
Efficient detection of plant diseases in agriculture is an important topic of research as the diseases in plants will directly influence crop production, quality of crops and agricultural economy. Though image processing techniques and classification algorithms are useful in automatic detection of diseases, still detection with adequate or enhanced accuracy is an open issue. Recently in contrast to traditional machine learning algorithms deep neural networks are being used for more accurate prediction tasks owing to their capability in solving complex problems. In this work an approach based on autoencoder technique is proposed for detection of diseases in plant leaves with an intention of improving the accuracy of detection. After preprocessing the images of plant leaves, the images are segmented into the normal and affected portions of the images using FCM algorithm. Features are extracted from the segmented clusters by constructing Discrete wavelet transform (DWT) based Gray-level Co-occurrence Matrix (GLCM). The extracted features have been detected for the presence of diseases using deep autoencoder technique. With typical datasets, the accuracy obtained using autoencoder technique is found to be higher than that obtained using conventional approach.

Keywords
Hybrid Laplacian of Gaussian (HLoG) filter, Fuzzy C-Means, Discrete Wavelet Transform (DWT), Gray-Level Co-occurrence Matrix (GLCM) and Deep Auto-encoder classification.

1,2 Department of Computer Science, Bharathidasan University Constituent Arts and Science College [Affiliated to Bharathidasan University], Navalurkuttapattu, Tiruchirappalli-620027, Tamil Nadu, India.

*Corresponding author: ¹vijayabharathiphd@gmail.com; ²drschellammal@gmail.com

Article History: Received 04 June 2020; Accepted 23 August 2020 ©2020 MJM.

Contents

| Section | Title | Page |
|---------|-------|------|
| 1 | Introduction | 711 |
| 2 | Related Work | 712 |
| 3 | Proposed Approach | 712 |
| 3.1 | Dataset | 712 |
| 3.2 | Processing | 712 |
| 3.3 | Segmentation | 713 |
| 3.4 | Feature Extraction | 713 |
| 3.5 | Classification | 713 |
| 4 | Results | 713 |
| 5 | Conclusion | 713 |

1. Introduction
Detection of diseases has to be performed efficiently in order to improve production and quality of crops. In general, the leaves are a major source of disease in most plants. Plant diseases vary in color, shape, and size. Some diseases are similar in size but different in color and some are similar in color but different in size. Each disease has specific characteristics. Manual detection of plant diseases in a large field is time consuming and expensive. In addition, errors are likely to occur in manual determination of diseases [1]. Image processing techniques are being extensively used for detecting plant diseases [2]. In this work, an approach is proposed to detect diseases in plant leaves using Fuzzy C-Means clustering based segmentation and Discrete Wavelet Transform based feature extraction. It is proposed to use deep learning based autoencoder technique for detection of diseases in plant leaves and to compare the performance with that of Support Vector Machine.

The paper is organized as follows. Section 2 presents research works that have the same theme as that of the proposed work. Section 3 describes the proposed approach in detail. Section 4 presents the results. Section 5 concludes the work.
2. Related Work

In this section we provide some recent researches studies associated to the recognition and classification of apple and Sugarcane plant diseases are specified as follows. In research works [3] clustering based segmentation and SVM based classification are used to detect diseases such as yellow mosaic and grasshopper leaf diseases in plant leaves. In another work [4] multi SVM classifier is used to classify leaf minor, leaf spot and mosaic disease on the cucumber leaves. In [5], K-Means clustering and threshold are used to segment the disease region along with GLCM features for classification using SVM. In [6], a combination of green pixel masking and threshold levels for segmentation to find the diseases on pepper leaves is proposed. In addition, this work uses GLCM features for classification using neural network. In [7], an approach based on k-means clustering for segmentation, GLCM technique for feature extraction and SVM algorithm is described for finding diseases in plant leaves and fruits. The drawback of this model, huge dataset is not processed in this model. In another research work [8] also SVM algorithm is used to detect disease in sugarcane. In research works [9-10] deep learning methods are used for detecting diseases.

3. Proposed Approach

The schematic view of the of the proposed approach is given in Fig. 1.

3.1 Dataset

Two datasets are used in this work. At first, the proposed approach is tested for its performance using benchmark dataset consisting of 100 apple leaf images collected from the URL (https://www.kaggle.com/) are used for analysis. The data set contains both healthy and affected leaves. Some sample apple images are shown in Fig. 2. In addition to benchmark dataset, the approach has been tested using the dataset consists of 100 sugarcane leaf images captured from agricultural land near Orathanadu, Thanjavur District, TamilNadu. The images are collected with the help of a mobile camera. The images contain $3456\times 5184$ pixels. The images are in the form of RGB images. Some of the sugarcane leaf images are shown in Fig. 3.

3.2 Processing

The size of input images are converted into $256\times256$ pixels. The images are converted as grey scale images. Laplacian of Gaussian (LOG) filters are used to detect the edges of an image with the required noise removal and smoothening of images using Gaussian filtering. The Log filtering and Gaussian filtering are given through (3.1) and (3.2).

\[
\nabla^2 G(x,y) = \frac{(x^2 + y^2 - 2\sigma^2)}{\pi\sigma^4} \exp\left(\frac{-x^2 + y^2}{2\sigma^2}\right) \quad (3.1)
\]

\[
G(x,y,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (3.2)
\]
3.3 Segmentation

After preprocessing the leaf images are clustered into two partitions, background and leaf portion using Fuzzy C-Means (FCM) algorithm as discussed in [11]. In contrast to K-Means clustering which clusters a given data point into some cluster in a deterministic manner. This means that, the algorithm assigns values of any data point lying in a particular cluster to be either as 0 or 1, whereas the FCM algorithm allocates a membership value for any data point belonging to a particular cluster.

3.4 Feature Extraction

For any classification problem determining appropriate features is essential as it influences the accuracy of detection. In this work, texture based features are used for classification. Texture based features represent the structural arrangement of surfaces in an image. In this work, DWT based GLCM is used to extract the required features, namely, (i) gray level (ii) colour and (iii) wavelet features. This is described in [12]. The resulting matrix is used for detection using classifiers. To calculate the input coefficient of the input images, Discrete Wavelet Transform (DWT) is used, which takes into account the rectangular function. The images are broken down into multiple sub-images at various resolution stages to keep the information low and high frequency. The DWT property is used to determine textured based content from images which is described in [12]. The results of the wavelet decomposition of an image is 4-orthogonal bands such as Low-Low (LL), High-Low (HL), Low-High (LH), and High-High (HH) bands [12]. From LL wavelet, various texture-based features such as autocorrelation, contrast, correlation, correlation, cluster output, cluster hue, difference, energy entropy, uniformity, uniformity, extreme likelihood, number of cubes, variance, normal sum, change, entropy of a sum, difference fluctuation, and entropy have been extracted. Significant features are used for detection.

3.5 Classification

The classification process is done by using the deep learning-based Auto Encoder deep neural network (DNN) classifier as discussed in [13]. Auto Encoder usually operates as a DNN based classification task. Basically in an autoencoder technique, the input data is compressed into an encoded format data in the hidden layer with help of encoding. Basically, the transformed model is used for training the classifier. The output of encoder is the compressed representation of input. After compression, the neurons are activated through the hidden network using equation (3.3). The neuron activity measured using equation (3.4). In DNN, the data flow is obtained from the input layer to the output layer without any looping function. The main benefit of the Auto Encoder Classifier is that the likelihood of error value is too low than another classifier. As discussed in [13], input-encoding neurons represent the true input value.

\[ a^{(i)} = x^{(i)}, l = 1 \]
\[ a^{(i)} = \sigma (W^l a + b), l = 1 \]  

(3.3)

\[ h(W, b, x) = a^{(N)} \]  

(3.4)

here \( x \) represents unlabeled data \( \{x^{(i)}\}^m_{i=1} \), \( w = 1 \), represents weight matrix which pedals the activation, \( b \) represents bias term, \( \sigma \) represents activation function, it can be set to hyperbolic target function to deliver non-linearity for the network to model difficult relationship, and \( h(W, b, x) \) represents input data as well as activation output layer. To train the unsupervised scheme, the loss of production has been employed as an objective function as given in (3.5).

\[ L(W, b, x, z) = \min_{W, b} E(W, b, x, z) + \gamma ||W||^2 + \beta K(W, b, x) \]  

(3.5)

Where, \( E(W, b, x, z) + \gamma ||W||^2 \) signifies the demonstration loss with the squared error. At output, the decoder is used to decode the compressed data into original representation. In addition, after training and construction of the model, the decoder part can be even be eliminated. Precisely speaking, the autoencoder works as a method for training the model in an unguided manner. An output layer has been added to the top of the trained self-coding stack which uses another activation function, as defined in

\[ h_i^l = \frac{e^{w_i^l} b_i^l}{\sum_j w_i^l h_j^{l-1} b_j^l} \]  

(3.6)

where \( w_i^l \) is \( i^{th} \) row of \( W^l \) and \( b_i^l \) is \( i^{th} \) ending layer bias term.

4. Results

The extracted features are given as input for classification. Two classifiers namely Support Vector Machine and deep learning based autoencoder techniques are used. The performance of the classification techniques are studied with different evaluation measures, namely, accuracy, sensitivity, specificity and precision. The performance obtained using SVM and autoencoder techniques with FCM segmentation for two different datasets are given in Table 1 and Table 2.

5. Conclusion

In this work, a method based on Fuzzy C-Means clustering segmentation and DWT based feature extraction have employed to two different datasets to analyze the performance of SVM classification and deep learning based autoencoder technique. From experimentation (From Table 1 and Table 2) it is found that the deep learning based autoencoder technique outperforms the SVM.
Table 1. Values of evaluation measures obtained for apple dataset

| Technique                                      | Accuracy (%) | Sensitivity (%) | Specificity (%) | Precision (%) |
|------------------------------------------------|--------------|-----------------|-----------------|---------------|
| FCM Segmentation and SVM Classification        | 70           | 70              | 70              | 72            |
| FCM Segmentation and Autoencoder technique     | 80           | 80              | 80              | 80            |

Table 2. Values of evaluation measures obtained for sugarcane leaf dataset

| Technique                                       | Accuracy (%) | Sensitivity (%) | Specificity (%) | Precision (%) |
|------------------------------------------------|--------------|-----------------|-----------------|---------------|
| FCM Segmentation with SVM Classification        | 75           | 75              | 75              | 75.5          |
| FCM Segmentation with Autoencoder Classification| 79           | 79              | 79              | 79.5          |

References

[1] Yuanyuan Shao, Guantao Xuan, Yangyan Zhu, Yanling Zhang, Hongxing Peng, Zhongzheng Liu and Jialin Hou, Research on automatic identification system of tobacco diseases, The Imaging Science Journal, 65(4)(2017), 252-259.

[2] Vijai Singh, A. K. Misra, Detection of plant leaf diseases using image segmentation and soft computing Techniques, Information Processing In Agriculture, 4(2017), 41-49.

[3] C. G. Dhaware and K. Wanjale, A modern approach for plant leaf disease classification which depends on leaf image processing, Computer Communication and Informatics (ICCCI), 2017 International Conference on. IEEE, (2017), 1-4.

[4] P. Krithika and S. Veni, Leaf disease detection on cucumber leaves using multi class support vector machine, Wireless Communications, Signal Processing and Networking (WiSPNET), 2017 International Conference on. IEEE, (2017), 1276-1281.

[5] V. Pooja, R. Das, and V. Kanchana, Identification of plant leaf diseases using image processing techniques, Technological Innovations in ICT for Agriculture and Rural Development (TIAR), 2017 IEEE. IEEE, (2017), 130-133.

[6] J. Francis, B. Anoop et al, Identification of leaf diseases in pepper plants using soft computing techniques, Emerging Devices and Smart Systems (ICEDSS), Conference on. IEEE, (2016), 168-173.

[7] G. Kshirsagar, A. N. Thakre, Plant Disease Detection in Image Processing Using MATLAB, International Journal on Recent and Innovation Trends in Computing and Communication, 6(4)(2018), 113-116.

[8] G. L. Anoop, C. Nandini, Detection and Classification of Ring, Rust and Yellow Sugarcane Leaf Diseases, International Journal of Engineering and Advanced Technology (IJEAT), 8(6)(2019).

[9] H. S. Malik, M. Dwivedi, S. N. Omkar, T. Javed, A. Bakey, M. R. Pala, A. Chakravarthy, Disease Recognition in Sugarcane Crop Using Deep Learning, In Advances in Artificial Intelligence and Data Engineering, (2020), 189-206.

[10] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, D. Stefanovic, Deep neural network based recognition of plant diseases by leaf image classification, Computational intelligence and neuroscience, (2016).

[11] Rehna Kalam, Ciza Thomas and M Abdul Rahman, Gaussian Kernel Based Fuzzy CMeans Clustering Algorithm For Image Segmentation, Comput. Sci. Inf. Technol, (2016), 47-56.

[12] Mohanty, P. Sharada, David P. Hughes, and Marcel Salathé. Using deep learning for image-based plant disease detection, Frontiers in plant science, 7(2016), 14-19.

[13] H. F. Pardeed, E. Suryawati, R. Susika and V. Zilvan, Unsupervised Convolutional Autoencoder-Based Feature Learning for Automatic Detection of Plant Diseases, 2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA), Tangerang, Indonesia, (2018), 158-162.