PRINT ISSN 2319-1775 e-ISSN 2320-7876, www.ijfans.org
Volume.11, Issue13, Nov- 2022
Research Paper

UGC CARE Listed (Group -I) Journal

Plant Leaf Disease Classification Using SVM and CNN Algorithms

Asha Patil
Shrimad Rajchandra Institute of Management and Computer Application, UTU Bardoli

Abstract: -

The detection and diagnosis of plant leaf diseases is a major concern in agriculture. Farmers need to track crop fields and recognize signs of disease at early as possible. The images processing is an aid to the identification and classification of leaf diseases. For the identification of leaf disease, there are three image features, i.e., texture, color, and shape. Textures are the most important feature out of the three. Image features value entered in CNN. The k-fold methods are used for train the model greater accuracy evaluation. The image features provide input for the identification of a disease and are reliable using the GLCM algorithm. CNN provides 92.53 % accuracy.

Keywords: Texture Features, GLCM, CNN, k-fold

1. Introduction

India is an agrarian nation with approximately 70% of its population relying on agriculture for their livelihoods. For farmers, the process of selecting appropriate crops and finding effective herbicides and pesticides is diverse and extensive. The presence of diseases in plants has a significant detrimental impact on agricultural products, leading to a substantial reduction in both their quality and productivity[1]. Therefore, strategies and technical knowledge and the field became an important matter to be mastered. The systematic and structured should be developing so that they will use by operators to increase the overall production. In most cases, disease symptoms are seen on the leaves, stem, and fruit [2]. It needs to be detected that disease but very time appointing an expert would may be more costly. Continuous monitoring with naked-eye observation is not possible for a farmer. So, we use digital image processing techniques [3]. Particular medications can be custom fitted to battle particular pathogens if plant illnesses are accurately analyzed and recognized early and ecological increases[4].Initially, the input RGB image is converted into the HIS model. The k-mean segmentation is an efficient segment of the disease affected area of leaves of crops. The GLCM is a features extraction algorithm that measures the feature values from the ROI disease-affected image from the 'H' component [5]. Exact disease finding is a difficult task for farmers which results in loss of income to the farmers and the state[6]. In the classical method, expert people detected diseases in leaves by naked eyes which are very expensive for farmers [7]. The main goals of the research are:

1. To study the disease of plant leaf.
2. To calculate disease-affected area on a leaf.
3. To calculate essential features from leaf images
4. To choose the correct classifier with the help of result comparison.

The research purpose is to improve the farmer's economical life by increasing the production of crop. The disease identification and exact pesticide suggestion is a crucial part of this work. To make the system that is automatically run and identify the disease is reduce the effort and cost of farmers. Feature extract from the image is a crucial part for finding the accuracy of disease-affected leaf [8].

2. Literature Review

Leaf images affected by the disease are pre-processed by resizing the image or color conversion of an image or calculate the histogram of an image. To remove the noise from image filters is used and enhance the image quality. The color, shape, and texture features are calculated by different segmented methods, k-mean is the most popular method used for finding the region ROI of the image. In some work combination of texture, color, and shape features are used to classify the various diseases on a leaf. The texture features are mostly used in the disease detection of the leaf. These calculated values were given to the classifier and the leaf disease was classified by this method. Machine learning classifiers are used for leaf disease classification. After studying the previous work, found the research gap for our research work.

1. For the identification and classification of the correct disease, feature selections are more important. The strategy for image classification is critical and balanced.
2. Most researchers find the twenty-three features or some researchers find two features of the leaf disease. More image features make the classifier more time-consuming and cost-intensive, and few features make the classifier less efficient if the dataset differs. The model can go in underfitting.
3. The leaf disease data collection for most research work is not obtained from an open farm or nursery. Leaf disease samples were obtained from an online platform. Plant leaf diseases are collected from open farms and nurseries for this work.

4. To make the model more suitable, the size of the leaf sample data set is very limited.

We chose the suitable features of different plant leaf disease from the digital image in this work. Texture and shape features are derived from that images. The feature vector set for the classification and identification of plant leaf disease as an input to classifiers.

3. Diseases on Plant Leaf

Plant leaf samples are collected from open farms of Nandurbar District with a different session at the early stage, middle stage, and last stage of the disease. These samples are collected from July to February of the year. In this study, four diseases and one healthy leaf were observed in plant leaves. The affected diseases are harmful to crop production. There is a variety of disease spot which are same with different disease spot which makes confusion to recognize the disease[9]. The wrong prediction of the disease may go in the wrong direction regarding spraying pesticide or chemical treatment. The loss of money and time in this is very useless [10]. Leaf plants affected by fungal, bacterial, and viral diseases are captured by the digital camera[11].

Table 1. Sample collection of plant leaf diseases

Sr. No	Name of Disease	Total
1	Cercospora Leaf Spot	103
2	Anthracnose	101
3	Leaf Curl	102
4	Powdery Mildew	100
5	Healthy Leaf	100

The total sample size of a leaf is 506 collected. The samples are maintained by the disease with their intensity. These symptoms of all samples are confirmed by agriculture experts. The growth of the disease is divided into three states i.e., early-stage middle stage, and last stage. The farms are visited annually and, for each disease, samples are collected during the year. The external parameters of humidity, humidity, temperature and heavy rain influence the growth cycle of each disease.

4. Proposed Model

Image processing has some fundamental steps for the classification and identification of results. They are image preprocessing, image segmentation, image feature extraction, and Classification. The following figure shows the flow of research work that is applied in this work. Firstly, we take samples of a disease-affected leaf of crop. Then for preprocessing the image apply the resizing of images. An original image is in RGB form then converts into HSI form. Image is segmented by K-mean algorithm, and features of the image are extracted using the GLCM algorithm[12][13]. These extracted features a train and tested for classification. The classifier is applying for disease identification, and classification.

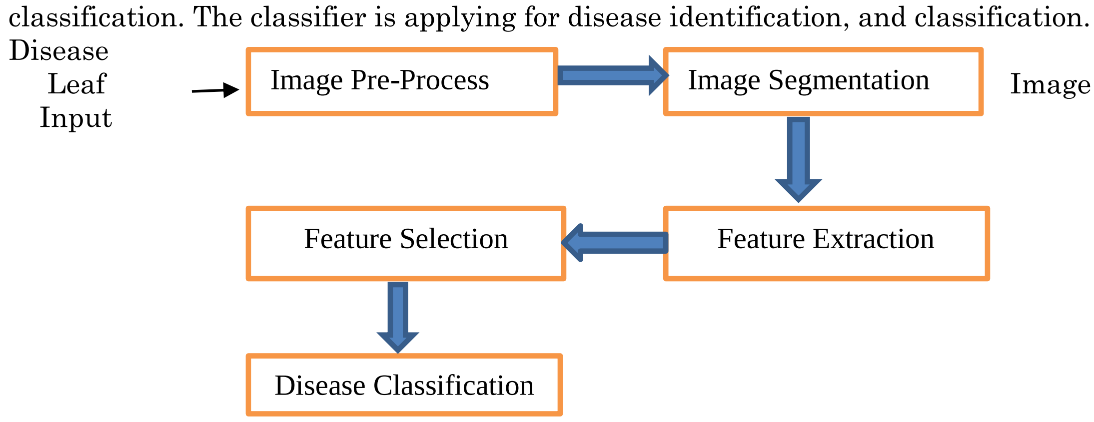

Figure.7 Proposed Model of Leaf Disease Detection, and Classification System (LDDCS)

In above figure show that, identification, and classification of plant leaf disease with help of image processing fundamental steps. The diagram indicates the flow of research work step by step. The classification of the disease will be getting by machine learning and deep learning algorithm classifiers.

4.1 Image Preprocessing

The plant leaf is acquired as an image that can be of different sizes. The size of the resized image becomes 255 × 255 diseases affected leaf of plants are captured by digital camera. A digital camera (Canon EOS REBEL T2i, 20 Megapixels) was used for image acquisition. All of the images are stored in the standard format of JPEG. In preprocessing procedure original image is an RGB image, and it was converted into HIS format to reduce the dimension, and complexity of the image.

4.2 Segmentation

The K-mean segmentation method is applied on images for disease-affected parts isolated from the 'H' component of HIS image. Using k-mean cluster segmentation, the proposed system calculates the region of interest with iterations. However, K-mean clustering is used to separate the leaf image into different clusters. In our case, k=3 is chosen, and the three clusters will have the background, diseased portion of the leaf, and normal green leaf. We have fixed the cluster value as two for the disease portion is maximum covered in this cluster.

4.3 Feature Extraction

The GLCM algorithm is used in the plant leaf disease image to measure features one by one, and extracted features will be tested. Contrast, Energy, Correlation, Entropy, Cluster Shadow, Cluster Prominence, Kurtosis, and skewness are the top priority features. This selection of features is a combination of the features of texture and color.

4.4 Classification

The machine learning algorithms is used for the classification purpose to classify the data that is Support Vector Machine (SVM) and one Deep Learning Algorithm Convolution Neural Network (CNN) gives the result with accuracy. On the same data set, the machine learning, and deep learning algorithms function differently on different hyperparameters for leaf disease detection [29]. Every outcome depends on the size of the data for training, and testing. The test data may change, and the resulting outcome will change. One way of strengthening the K-fold cross-validation over the holdout approach is used[14]. The data set is split into k subsets, and k times are repeated by the holdout process. One of the k subsets is used as the test set each time, and the other k-1 subsets are combined to form the training set [15].

4.4.2 Support Vector Machine (SVM)

SVM constructs a hyperplane in the space, which can be used for classification that has the highest distance to the closest training data point of any class [16]. The multisvm () function is used to classify the result of leaf disease detection, and classification[17]. SVM uses the RBF kernel to classify plant leaf disease. The fitcsvm and predict function are used for training the data set, and test data set respectively. The kernel functions linear, polynomial and RBF also help in giving decision boundaries for higher dimensions. In the following table, we have been calculating the accuracy of using RBF kernel with other performance measurement parameters like precision, recall, F1_Score.

Table 3: SVM Result with Kernels

Kernel	**Accuracy**	**Precision**	**Recall**	**F1_Score**
Polynomial	68.42	0.1	0.21	0.14
Linear	71.46	0.21	0.28	0.2
RFB	85.7	0.79	0.64	0.65

In the above table RBF provides more accuracy over the linear and polynomial kernel. The accuracy is 85.7 % in the RBF kernel at iteration 300. The accuracy of the linear and polynomial kernels is not greater than that of the RBF kernel.

4.4.3 Convolutional Neural Network (CNN)

A CNN consists of an input layer, a layer of output, as well as numerous hidden layers. Convolutional layers, pooling layers, completely linked layers, and normalization layers (ReLU) are usually the hidden layers of a CNN. Input from CNN was generated by the GLCM feature set. CNN presented image texture features as input to describe the identification of the plant leaf disease. For better performance, we have to use the epoch size performance tuning parameter. The varying effect is the increase in the epoch size of the accuracy and another output measure. The cross-fold method is used. The 2-fold gives optimum output after an experiment on the CNN model than other folds.

Table 4: CNN Result with epoch

Epoch	**Accuracy**	**Precision**	**Recall**	**F1_Score**
5	81.56	0.80	0.78	0.69
10	82.52	0.76	0.68	0.69
15	87.97	0.77	0.67	0.78
20	**95.33**	**0.89**	**0.98**	**0.93**
25	91.54	0.80	0.83	0.81

The epoch is a neural network that learns input data patterns by reading the input dataset and applying various calculations to it. At epoch size 20 we get 95.33% which is maximum accuracy.

5. Result and Discussion

In SVM, and CNN, leaf images of affected diseases are qualified for disease detection, and classification. The training is based in various ways on each classifier's hyperparameter. The kernel function is used for accuracy in SVM, and the RBF kernel gives maximum accuracy after iteration variation. The image texture features are inputted into CNN, an epoch used in it as a hyperparameter. The number of epochs is increased or decrease the accuracy is change.

Table 6: Comparison of KNN SVM, and CNN Results

Classifier	**Accuracy**	**Precision**	**Recall**	**F1_Score**
SVM	85.07	0.79	0.64	0.65
CNN	**95.33**	**0.89**	**0.98**	**0.93**

After comparing machine learning classifiers, and deep learning classifier, conclude that CNN provides better accuracy with less timing than SVM. In detecting plant leaf disease when used with texture features, deep learning architectures have achieved good efficiency. We provided a performance comparison of five image datasets for the disease class.

6. Conclusion

The data collection consists of five plant leaf diseases. These images are preprocessed by converting the HIS color to RGB images. Using the K-mean cluster algorithm, the 'H' elements of the affected portion of the leaf are segmented. Using the GLCM algorithm, feature extracted from cluster part of leaf image. The features are tested, and train then sorted based on maximum accuracy. With the best linear combination SVM and CNN get input of selected of image features. There are 70% images for training, and 30% images are being tested. To evaluate the data properly, the K-fold method is used. The CNN deep learning algorithm offers 95.33 % accuracy, and the SVM machine learning algorithms offer 85.07 % accuracy. Therefore, CNN is said to be better than SVM. In future other leaf disease detection is also performing using various dataset with different diseases.

References:

1. [1] Z. Husin, A. Shakaff, 2012, “Feasibility study on plant plant disease detection using image processing techniques,” *ieeexplore.ieee.org*, Accessed: Mar. 19, 2021. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6169716/.
2. [2] A. K. Mahlein, U. Steiner, C. Hillnhütter, H. W. Dehne, and E. C. Oerke, “Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases,” *Plant Methods*, vol. 8, no. 1, Jan. 2012, doi: 10.1186/1746-4811-8-3.
3. [3] J. D. Pujari, R. Yakkundimath, and A. Syedhusain. Byadgi, “Automatic Fungal Disease Detection based on Wavelet Feature Extraction and PCA Analysis in Commercial Crops,” *Int. J. Image, Graph. Signal Process.*, vol. 6, no. 1, pp. 24–31, 2013, doi: 10.5815/ijigsp.2014.01.04.
4. [4] J. D. Pujari, R. Yakkundimath, and A. S. Byadgi, “https://doi.org/10.1109/ICCIC.2014.7238283,” *2014 IEEE Int. Conf. Comput. Intell. Comput. Res. IEEE ICCIC 2014*, pp. 31–34, 2015, doi: 10.1109/ICCIC.2014.7238283.
5. [5] J. D. Pujari, R. Yakkundimath, and A. S. Byadgi, “Recognition and classification of Produce affected by identically looking Powdery Mildew disease,” *Acta Technol. Agric.*, vol. 17, no. 2, pp. 29–34, 2014, doi: 10.2478/ata-2014-0007.
6. [6] O. S. Android, “Detection of Plant Leaf Disease Using Image Processing Approach,” *Int. J. Sci. Res. Publ.*, vol. 6, no. 2, pp. 73–76, 2016.
7. [7] H. P. Kanjalkar, “Feature Extraction of Leaf Diseases,” *Int. J. Adv. Res. Comput. Eng. Technol.*, vol. 3, no. 1, pp. 153–155, 2014.
8. [8] G. Tripathi and J. Save, “an Image Processing and Neural Network Based Approach for Detection and Classification of Plant Leaf Diseases,” *Int. J. Comput. Eng. Technol.*, vol. 6, no. 4, pp. 976–6367, 2015.
9. [9] G. Bhogade, R. Mawal, A. Patil, and P. S. R. Desai, “Leaf Disease Identification using Image Processing,” vol. 7, no. 5, pp. 11489–11491, 2017.
10. [10] P. Rokita, “Application of image processing techniques in computer graphics algorithms,” *Comput. Networks*, vol. 29, no. 14, pp. 1705–1714, 1997, doi: 10.1016/s0169-7552(97)00085-8.
11. [11] J. D. Pujari, R. Yakkundimath, and A. S. Byadgi, “Image processing Based Detection of Fungal Diseases in Plants,” *Procedia Comput. Sci.*, vol. 46, no. Icict 2014, pp. 1802–1808, 2015, doi: 10.1016/j.procs.2015.02.137.
12. [12] J. Baghel and P. Jain, “K-Means Segmentation Method for Automatic Leaf Disease Detection,” vol. 6, no. 3, pp. 83–86, 2016.
13. [13] J. Baghel and P. Jain, “Disease Detection in Soya Bean using K-Means Clustering Segmentation Technique,” *Int. J. Comput. Appl.*, vol. 145, no. 9, pp. 15–18, 2016, doi: 10.5120/ijca2016910758.
14. [14] U. Mokhtar, M. A. S. Ali, A. E. Hassenian, and H. Hefny, “Tomato leaves diseases detection approach based on Support Vector Machines,” *2015 11th Int. Comput. Eng. Conf. Today Inf. Soc. What's Next?, ICENCO 2015*, pp. 246–250, 2016, doi: 10.1109/ICENCO.2015.7416356.
15. [15] E. Alehegn, “Maize Leaf Diseases Recognition and Classification Based on

Imaging and Machine Learning Techniques," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 5, no. 12, pp. 1–11, 2017.

16. [16] P. Krithika and S. Veni, "Leaf disease detection on cucumber leaves using multiclass Support Vector Machine," *Proc. 2017 Int. Conf. Wirel. Commun. Signal Process. Networking, WiSPNET 2017*, vol. 2018-Janua, pp. 1276–1281, 2018, doi: 10.1109/WiSPNET.2017.8299969.
17. [17] A. N. I. Masazhar and M. M. Kamal, "Digital image processing technique for palm oil leaf disease detection using multiclass SVM classifier," *2017 IEEE Int. Conf. Smart Instrumentation, Meas. Appl. ICSIMA 2017*, vol. 2017-Novem, no. November, pp. 1–6, 2018, doi: 10.1109/ICSIMA.2017.8311978.