International Journal on Emerging Technologies **11**(2): 1082-1086(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

A Multiclass Plant Leaf Disease Detection using Image Processing and Machine Learning Techniques

Nilay Ganatra[1] and Atul Patel[2]

[1]Assistant Professor, Faculty of Computer Science and Applications,
Charotar University of Science and Technology (CHARUSAT), Changa-388421 (Gujarat), India.
[2]Dean & Professor, Faculty of Computer Science and Applications,
Charotar University of Science and Technology (CHARUSAT), Changa-388421 (Gujarat), India.

(Corresponding author: Nilay Ganatra)
(Received 12 March 2020, Revised 06 April 2020, Accepted 10 April 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: In Indian economy, Agriculture is considered as a one of the strongest pillars. Agriculture sector contributes significantly in GDP of the country and it provides employability to many people live in rural area. Plant disease interrupts normal growth of plant and it is one of major reason behind less production that turns to economic losses. Timely detection of disease helps to provide the remedies that can control the spreading of disease in plants. Leaf examination is deliberated as a one of the best technique in plant disease diagnosis. Computer vision and machine learning become evolving fields that makes computers able to recognize and understand information from digital images. The paper proposed a prediction model for plant leaf disease detection and classification using computer vision and machine learning methods. The raw image of a leaf is pre-processed, segmented and features like shape, color, texture, vein etc. are extracted. Different machine learning classifiers are applied to classify the leaf image. The experimental results are evaluated and compared with Random Forest, Support Vector Machine, K-Neighrest Neighbor and Artificial Neural Network. The proposed prediction model performs well with random forest as compare to other classifiers.

Keywords: Plant leaf disease, classification, image processing.

I. INTRODUCTION

In Indian economy, Agriculture is considered as a one of the strongest pillars. In worldwide farming, India ranks at second. In vast rural area, Agriculture sector is the largest employer and provided employment to many. It significantly contributes to the GDP (Gross Domestic Product) of India. Plant disease interrupts normal growth of plant and it is one of major reason behind less production that turns to economic losses. Different parts of a plant include leaf, stem, seed may infected by plant diseases. Timely detection of plant disease plants is an essential requirement as it will increase the yield more than 60% of the total productivity [1]. Crop loss causes harmful effects on plants including hunger and starvation. The limited disease control method is one of the major reasons behind losses in major crops. There are two type of plant disease including infectious and noninfectious. Fungus, virus, bacteria are some of the major causes of infectious plant diseases.

It is essential to diagnose the disease quickly and accurately to provide remedies that control diseases. It is mainly dependent on the characteristic symptoms presented in diseased plant. As stated earlier, disease can be affected any part of a leaf like flower, stem, root etc. however, leaf examination is deliberated as a best technique in plant diagnosis [2]. The infected leaf mostly suffers from distortion in shape, color, size etc. The tradition method including expert for diagnosing plant leaf disease is out dated, time consuming and not appropriate. Using ICT in agriculture is seen as an evolving arena that greatly changes the dimension of agricultural and rural development. In ICT, artificial intelligence is an imminent field which can be applied to the farming sector quite efficiently. Moreover, computer vision becomes an evolving field that makes computers able to recognize and understand information from digital images. In artificial intelligence, machine learning has emerged as an effective computing paradigm that helps to solve many complex tasks of computer vision problems [3]. Machine learning makes computer enable to learn without human intervention.

This paper represents the methodology for plant leaf disease detection and classification using machine learning integrated digital image processing techniques. We have proposed a prediction model for plant leaf disease detection and classification that apply different image processing techniques for pre-processing, segmentation and feature extraction. Different classification methods are taken into consideration that classifies disease into one of the classes. We have experimented our proposed model on two different dataset. The paper organizes as follows. Section II contains the literature review. Section III discussed the proposed model for leaf disease detection and classification. Section IV represents the results and discussion.

II. LITERATURE REVIEW

This section reviewed the different work carried out in area of leaf disease detection. Many researchers worked in the field of implementing computer enabled diagnosis of leaf disease in different plant. The following table 1 represents the type of leaves, algorithm used for classification, accuracy reported and the type of disease detected.

Table 1: Literature Review for Plant Leaf Disease Detection and Classification.

Author and Year of Publication	Types of leaves	Algorithm Used for Classification	Reported Accuracy	Types of disease detected
Biswas *et al.* 2014 [4]	Leafs of Potato	Fuzzy C-means Clustering Neural Network	93%	Potato Leaf Blight
Joshi & Jadhav 2016 [5]	Disease affected Rice leaf	Minimum Distance classifier	89.23%	Rice blast & Rice Sheath Rot
Orillo *et al.,* 2013 [6]	Rice leaves	Back Propagation ANN (Artificial Neural Network)	93.33%	Identification of Nitrogen Level in healthy leaves
Pawar *et al.,* 2016 [7]	Leaves of cucumber plant	Artificial Neural Network	80.45%	Downey Mildew & Powdery mildew
Kaur & Laxmi 2016 [8]	General Leaves	SVM with optimization technique	96.77% to 98.42%	Bacterial Blight, Anthracnose
Sun *et al.,* 2017 [9]	Tulip Tree & Chinese Buckeye	Deep Learning	91.78%	Classification of plants
de Luna *et al.,* 2017 [10]	Various herbal plants	Artificial Neural Network	98.61%	Various disease like amplaya, akapulko and many other.
Mukherjee *et al.,* 2017 [11]	Tulsi Leaves	Multi-Layer Back Propogation Perceptron	80%	Health of tulsi leaves
Renugambal & Senthilraja, 2015 [12]	Sugarcane leaf	SVM	91%	Sugarcane mosaic virus, Brown spot, Red rot, Leaf spot, Brown spot
Jothiaruna, *et al.,* 2019 [13]	General Plant disease images	Region Growing and Color feature Extraction	87%	Disease affected portion segmentation
Naikwadi & Amoda 2012 [14]	Diseases affecting pomegranate	Artificial Neural Network	91%	Alterneria, BBD and Anthractnose

In recent years, many researchers have been applied computer vision and machine learning for detecting plant leaf disease. Many of them considered these techniques as an effective architecture as it provides feature extraction without manual intervention.

III. A PREDICTION MODEL FOR PLANT LEAF DISEASE DETECTION AND CLASSIFICATION

The usual characteristic of computer vision methods are they used to acquire, process, analyze and understand patterns from digital images. The computer vision applications are face recognition, satellite imaging, medical imaging etc. From large dataset, machine learning algorithm tends to be used to identify hidden pattern from data. Some of the machine learning techniques specially developed to perform computer vision tasks. The following figure 1 narrates the phases of the prediction model.

Dataset Accumulation: From the above figure 1, we can see that dataset accumulation is a primary and basic step to follow. Raw images are collected that are related to plant leaf disease are collected in this phase. We have collected the plant village dataset available on Kaggle publically for experimental purpose. We have acquired 14956 images from this dataset that are categorized into 38 distinct classes.

Image Pre-processing: Pre-processing on images is necessary to remove the noise from an image. It is also needed to normalize the intensity values from an image. Image preprocessing is the process of enhancing the quality of image by removing background noise and normalizing the intensity of the various elements of the image prior to any computational process applied on it [15].

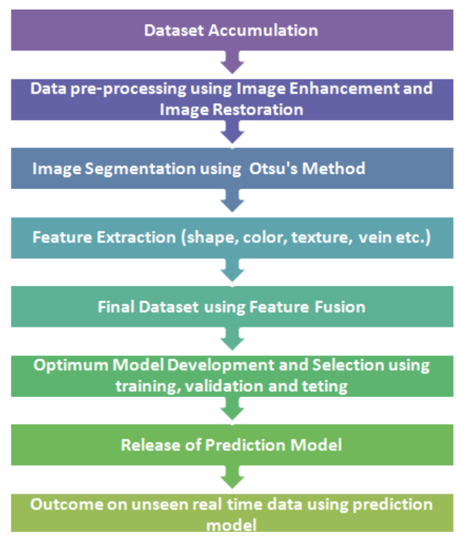

Fig 1. Phases of a Prediction Model for Plant Leaf Disease Dectection and Classification.

It is necessary to apply image preprocessing as it is useful for noise removal and color space transformation that helps to correctly focus the area of interest in an image. Image enhancement and restoration can be useful to pre-process an image. During our experiment, we have considered contrast stretching and histogram equalization.

Moreover, different filters like filters like Roberts, Prewitt and Sobel are applied on input image [16]. To convert the noisy image into clean image, image restoration techniques are useful. We have removed the unwanted noise and blur from leaf images using different filters like Average, Linear, Median and Adaptive.

The different types of noises considered are Poisson, Salt and Pepper, Gaussian and Speckle. For performance measurement, PSNR and RMSE values are considered.

Image Segmentation: After that, image segmentation is performed. Normally an image is surrounded by background or unwanted objects. For that, it is necessary to perform image segmentation techniques. In digital image processing, image segmentation is a technique of digital image processing that is used to partition the digital image into various meaningful or focused parts required for problem domain. It normally works on similar characteristics of the image pixels [17]. It is used to differentiate and isolated the foreground from the background of an image. It helps to extract the correct features when extraction is performed.

Based on the various experiment, machine-driven thresholding techniques known as Otsu's technique is implemented in this research for segmentation [18].

Feature Selection and Extraction: After applying appropriate pre-processing and segmentation on the disease affected leaf, various feature extraction techniques are applied to obtain the targeted features which can be used for the classification. Features can be viewed as quantifiable properties obtained from the various targeted portions of the image.

For any machine learning based application feature identification and feature extraction is considered as pioneer step. Overall accuracy of the machine learning algorithms is highly dependent on the obtained features. [19] Feature selection is considered as the most crucial step of any computer vision based machine learning task. For the plant disease classification 31 features are identified and extracted for the classification task. These features are broadly divided into six categories: Color Moments Features, Texture Features, Shape Features, Vein Features, Gobor Wavelet Transform Feature and Zernike Moment Features.

Color features are considered as the very basic features for image related problem. In image processing commonly used color models are RGB and HSV. Both are 3D coordinate representation and use three output channels for representing output [20]. Texture feature are the most common and important feature of an image. It is considered as the regional descriptor for any image and considered as the powerful feature for classification related problems. Here, GLCM features are used to obtain the texture properties of disease affected leaf [21]. Shape features are also fall under the basic features for any image.

Shape feature generally helps in identifying the various parameters of leaf as well as the disease affected region [22]. Vein features are considered as the derived features. Image vein structure is obtained and white pixel from the area of disease affected leaf is calculated [23]. Zernik Moment features are used to obtain information from the entire image rather than specific region of an image. Thus, some global properties are identified from images using this feature.

Table 2: Values of Features Extracted from Leaf Image.

Category of Feature	Name of the Feature	Sample Value
Color Moments Features	Mean	108.8120
	Standard deviation	31.3685
	Skewness	0.0722
	Kurtosis	2.0161
Texture Features	Contrast	0.2039
	Correlation	0.8941
	Homogeneity	0.9090
	Energy	0.2099
	Entropy	1.9018
	Variance	16.6503
Shape Features	Eccentricity	1
	Width Ration	324.3250
	Area	0.7921
	Equivdiameter	43
	Roundness or circularity	95.8857
	Solidity	1
	Centroid	43
	Convexarea	7221
	Convexhull	[0 0 … 256]
Vein Features	V1	1
	V2	0.9951
	V3	0.9904
	V4	0.9887
Gabor Wavelet Transform Feature	Total 256 features obtained	[82302, 868,538,…,21421]
Zernike Moment Features	AOH	0.1303
	PhiOH	-10.3223

Gabor Texture Analysis is used to obtain specific frequency content in appropriate direction in an image. Gabor feature with various orientation and frequency may be used to obtain valuable features out of the image. The table 2 represents the all feature used for the plant disease identification.

Model Development and Selection: In machine learning, classification refers to categorize the data into labeled classes. Predictive modeling builds a model that uses samples of data for which the class is known and it then after classifying new observations. To develop any prediction model, applying classification algorithm is necessary. Moreover, a model should be trained, validate and test appropriately. We have partitioned our data into train, validate and test. After the training, the model is tested against the performance parameters. Moreover, the final accuracy of the model is found using test dataset.

We have considered four classifiers namely random forest, support vector machine, K-nearest neighbour and artificial neural networks. Moreover, to measure the performance of these classifiers, different parameters including accuracy, recall precision and F-score are considered.

Artificial Neural Network (ANN): ANN is made up with input layer, hidden layer(s) and an output layer. Each layer is made up with neurons and contains weights that pass to the neurons residing next layer. It works similarly as biological neurons available in human brain. Training is mandatory for any ANN model. For that, feed forward back propagation is applied[24]. It is necessary to apply hyper parameters to control the underfit and overfit issues normally occur during training of an ANN.

Support Vector Machine (SVM): SVM is one of the widely used algorithms that perform well with classification and regression. It can be successfully applied for linear and non-linear problems. To have more accuracy, SVM normally uses concept of kernel [25]. There are many types of kernel are available like Linear Kernel, polynomial kernel and radial basis function kernel.

Random Forest: Random forest is one type of ensemble learning method. Like SVM, it can be also used for classification and regression tasks. Random forest uses the concept of decision tree as a basic building block. To classify a new object from an input vector, it sets the input vector down each of the trees in the forest, where each tree gives a classification [26]. Random forest needed more computational resources. However, it is also able to work with a large range of data items with high accuracy.

K-Nearest Neighbour (KNN): KNN is considered as a non-parametric method and for distribution of the data, it does not make any underlying assumptions. It work on the concept of lazy learning where separate training phase is not required. It uses the training data at the time of classification [27]. It normally uses a distance measure function and Euclidean distance is considered a one of the most used method for calculating the distance.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

We have carried out our implementation work on the Matlab2020 on windows operations system platform. We have created GUI based automated plant disease detection and classification system. Our system is trained on dataset with the 38 classes and 14956 images.

Disease affected images is taken after training out model to carry out various steps like pre-processing, segmentation, feature extraction, feature fusion to combine all features and classification of disease. For the classification task we have used state-of-the-art machine learning algorithms Random Forest, SVM, KNN and ANN. The confusion matrix is used as evaluation parameters for the experimentation carried out. Various calculations are performed based on confusion matrix like accuracy, recall, precision, specificity and F-Score a represented in following table 3.

From the obtained results, it can be identified easily that Random Forest classifier outperforms compared with other classifiers as per represented in following figure 2. However, with the more number of classes and images, SVM gives significantly good accuracy compared with Artificial Neural Network and K-Nearest Neighbour Algorithms.

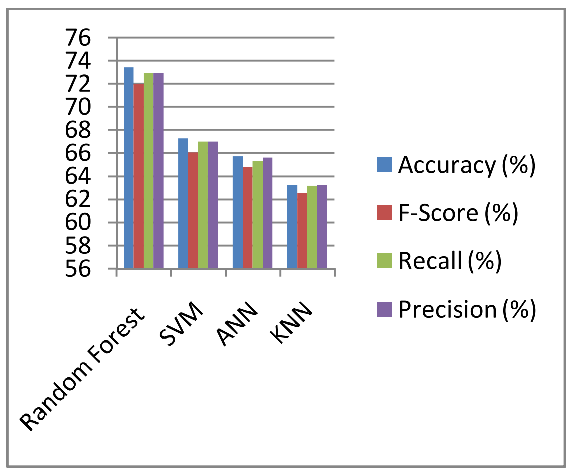

Fig. 2. Performance Evaluation of Different Classifiers.

Table 3. Performance Evaluation of Various Machine Learning Classifiers.

Classifier Name	Accuracy (%)	FScore (%)	Recall (%)	Precision (%)
Random Forest	73.38	71.98	72.90	72.88
SVM	67.27	66.02	66.97	66.99
ANN	65.68	64.76	65.30	65.60
KNN	63.20	62.55	63.15	63.18

Most of the experimentation carried out on the literature survey are focusing on the specific plant disease type. As the number of classes and images increases, performance of the machine learning algorithm declined significantly.

V. CONCLUSION

We have proposed an automated prediction model for leaf disease detection and classification using digital image processing and machine learning techniques. We have applied different pre-processing techniques and segmentation technique that helps to extract the different types of features adequately. We have accumulated 14956 images that are classified in 38 classes. The experimental results are evaluated and compared with Random Forest, Support Vector Machine, K-Neighrest Neighbor and Artificial Neural Network. The proposed prediction model performs well with random forest as compare to other classifiers as it gives 73.38% accuracy with 71.98% F-Score, 72.90% recall and 72.88% precision values. However, it has been also observed that with increasing number of classes and images, the performance of machine learning algorithms is significantly decreased. This limitation can be overcome by applying more optimum algorithms.

REFERENCES

[1]. Mattihalli, C., Gedefaye, E., Endalamaw, F., & Necho, A. (2018). Plant leaf diseases detection and auto-medicine. *Internet of Things*, *1*, 67-73.

[2]. Iqbal, Z., Khan, M. A., Sharif, M., Shah, J. H., ur Rehman, M. H., & Javed, K. (2018). An automated detection and classification of citrus plant diseases using image processing techniques: A review. *Computers and electronics in agriculture*, *153*, 12-32.

[3]. Saradhambal, G., Dhivya, R., Latha, S. & Rajesh, R. (2018). Plant disease detection and its solution using image classification. *International Journal of Pure and Applied Mathematics*. *119*, 879-883.

[4]. Biswas, S., Jagyasi, B., Singh, B. P., & Lal, M. (2014, June). Severity identification of Potato Late Blight disease from crop images captured under uncontrolled environment. In *2014 IEEE Canada International Humanitarian Technology Conference-(IHTC)* (pp. 1-5). IEEE.

[5]. Joshi, A. A., & Jadhav, B. D. (2016, December). Monitoring and controlling rice diseases using Image processing techniques. In *2016 International Conference on Computing, Analytics and Security Trends (CAST)* (pp. 471-476). IEEE.

[6]. Orillo, John William, Emperador, Gideon, Gasgonia, Mark, Parpan, Marifel & Yang, Jessica. (2014). Rice plant nitrogen level assessment through image processing using artificial neural network. 1-6.

[7]. Pawar, P., Turkar, V., & Patil, P. (2016, August). Cucumber disease detection using artificial neural network. In *2016 International Conference on Inventive Computation Technologies (ICICT)* (Vol. 3, pp. 1-5). IEEE.

[8]. Kaur, L., & Laxmi, V. (2016). Detection of unhealthy region of plant leaves using neural network. *Dis Manag*, *1*(05), 34-42.

[9]. Sun, Yu & Liu, Yuan & Wang, Guan & Zhang, Haiyan. (2017). Deep Learning for Plant Identification in Natural Environment. *Computational Intelligence and Neuroscience.* 2017. 1-6.

[10]. de Luna, Robert, Baldovino, Renann, Cotoco, Ezekiel, De Ocampo, Anton Louise, Valenzuela, Ira, Culaba, Alvin & Dadios, Elmer. (2017). Identification of philippine herbal medicine plant leaf using artificial neural network. 1-8.

[11]. Mukherjee, Gunjan, Chatterjee, Arpitam & Tudu, Bipan. (2017). Morphological feature based maturity level identification of Kalmegh and Tulsi leaves. 1-5.

[12]. Renugambal, K. & Senthilraja, B. (2015). Application of Image Processing Techniques in Plant Disease Recognition. *4*(3). 919-923.

[13]. Jothiaruna, N., Sundar, K. & Karthikeyan, B. (2019). A segmentation method for disease spot images incorporating chrominance in Comprehensive Color Feature and Region Growing. Computers and Electronics in Agriculture. 165.

[14]. Naikwadi, S., & Amoda, N. (2013). Advances in image processing for detection of plant diseases. *International journal of application or innovation in engineering & management (IJAIEM)*, *2*(11).

[15]. Singh, Vijai & Misra, A.K.. (2016). Detection of Plant Leaf Diseases Using Image Segmentation and Soft Computing Techniques. Information Processing in Agriculture. 4

[16]. Radha, Suja. (2017). Leaf Disease Detection using Image Processing. *Journal of Chemical and Pharmaceutical Sciences.*

[17]. Zhen Ma, J M R. S. Tavares, R. M. Natal Jorge.(2009). A review on the current segmentation algorithms for medical images. *1st International Conference on Imaging Theory and Applications (IMAGAPP).* 135-140.

[18]. Jamil A. M. Saif, Al-Kubati A. A. M, Hazaa A. S, Al-Moraish M. (2012). Image segmentation using Edge Detection and Thresholding. *The 13th International Arab Conference on Information Technology (ACIT).*

[19]. Yogamangalam, R. & Karthikeyan, B. (2013). Segmentation techniques comparison in image processing. *International Journal of Engineering and Technology*. *5*, 307-313.

[20]. Bora, Dibya, Gupta, Anil & Khan, Fayaz. (2015). Comparing the Performance of L*A*B* and HSV Color Spaces with Respect to Color Image Segmentation.

[21]. Guru, Devanur, Kumar, Y H & Shantharamu, Manjunath. (2010). Texture Features and KNN in Classification of Flower Images. *International Journal of Computer Applications, Special Issue on RTIPPR*. *1*. 21-29.

[22]. Salman, Adil, Semwal, Ashish, Bhatt, Upendra & Thakkar, V.. (2017). Leaf classification and identification using Canny Edge Detector and SVM classifier. 1-4.

[23]. SibiChakkaravarthy S., Sajeevan G., Kamalanaban E., Varun Kumar K. A. (2016) Automatic Leaf Vein Feature Extraction for First Degree Veins. In: Thampi S., Bandyopadhyay S., Krishnan S., Li KC., Mosin S., Ma M. (eds) Advances in Signal Processing and Intelligent Recognition Systems. *Advances in Intelligent Systems and Computing*, vol 425.

[24]. Vyawahare, Vishwesh, Espinosa-Paredes, Gilberto, Datkhile, Gaurav & Kadam, Pratik. (2018). Artificial neural network approximations of linear fractional neutron models. *Annals of Nuclear Energy*. *113*, 75-88.

[25]. Ebrahimi, Mohammad, Khoshtaghaza, Mohammad-Hadi, Minaee, Saeed & Jamshidi, Bahareh. (2017). Vision-based pest detection based on SVM classification method. *Computers and Electronics in Agriculture*. *137*, 52–58.

[26]. V.F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, J.P. Rigol-Sanchez. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *67*, 93-104.

[27]. K-Nearest Neighbor(KNN) Algorithm for Machine Learning, https://www.javatpoint.com/knearest-neighbor-algorithm-for-machine-learning

How to cite this article: Ganatra, N. and Patel, A. (2020). A Multiclass Plant Leaf Disease Detection using Image Processing and Machine Learning Techniques. *International Journal on Emerging Technologies*, **11**(2): 1082–1086.