Detection of Potato Disease Using Image Segmentation and Machine Learning

Md. Asif Iqbal[1] and Kamrul Hasan Talukder[2]

[1,2]Computer Science and Engineering Discipline, Khulna University, Khulna-9208, Bangladesh
[1]asifnwu1993@gmail.com, [2]k.h.t@alumni.nus.edu.sg

***Abstract*—Potato is one of the prominent food crops all over the world. In Bangladesh, potato cultivation has been getting remarkable popularity over the last decades. Many diseases affect the proper growth of potato plants. Noticeable diseases are seen in the leaf region of this plant. Two common and popular leaf diseases of the potato plants are Early Blight (EB) and Late Blight (LB). However, if these diseases were identified at an early stage it would be very helpful for better production of this crop. To solve this problem by detecting and analyzing these diseases image processing is the best option. This paper proposes an image processing and machine learning-based automatic system that will identify and classify potato leaf diseases. In this paper, image segmentation is done over 450 images of healthy and diseased potato leaf, which is taken from publicly available plant village database and seven classifier algorithms are used for recognition and classification of diseased and healthy leaves. Among them, The Random Forest classifier gives an accuracy of 97%. In this manner, our proposed approach leads to a path of automatic plant leaf disease detection.**

***Index Terms*—Image segmentation, Global feature descriptors, Training, Classification, Machine learning.**

I. Introduction

Potato is one of the most requisite food crops in Bangladesh. It is also a good source of carbohydrates in our daily life. However, nowadays the production of potato is tremendously hampered due to various diseases, which are affecting the potato plants. Many diseases can attack potato plants and symptoms of diseases are evident in different parts of this plant. Common disease problem includes foliage (leaf) diseases. Two common leaf diseases of the potato plant are early blight and late blight. Early blight on potato leaf is caused by the fungal pathogen *Alternaria Solani* and the organism responsible for Late blight on potato leaf is *Phytophthora Infestans*. Both of them are fungal diseases, which occurred on a potato leaf, causing substantial loss and it is harmful to the economy of a country. Therefore, for optimum use of pesticides and to minimize the yield loss detection of these diseases are necessary. Usually, the most practiced approach for detection and identification of plant leaf disease is performed by naked eye observation by farmers or local experts. However, many instances are found where this approach is proven unfeasible due to excessive processing time and lack of experts at farms and often give inappropriate results. To overcome this situation, an automatic leaf disease detection system is needed.

As noted above, for better growth of this plant and successful cultivation of potato, a computer-based automatic leaf disease detection system is required. Various techniques namely image processing, IoT, Big Data are used for the recognition of leaf disease. The machine learning technique is also found efficient for this purpose.

Machine learning technique gives system the opportunity to learn by itself and it can give decisions. Three kinds of machine learning algorithms are available namely supervised learning, reinforcement learning and unsupervised learning. In this paper, various machine-learning classifiers are trained to recognize diseases on potato leaf and the classification process will be done by them also.

The rest of this paper is organized as follows: Section II describes the previous works done by others on leaf disease detection of different plants and Section III represents the dataset description. The methodology is explained in section IV. Section V contains system architecture. In section VI result and discussion are represented and section VII concludes this work with our future plans.

II. Literature Review

There are many works done in leaf disease detection. Here several types of methods that many researchers have analyzed on leaf disease detection are reviewed.

Monzurul Islam et al. in [1] developed an image-processing, machine learning-based effective and error-free leaf disease detection system for potato plant. Their automated system can classify disease affected and healthy potato leaf from publicly accessible dataset "plant village" and their system is tested over 300 images. The accuracy of their system is 95%. Their future work is to develop a smartphone-assisted system.

Malvika Ranjan et al. in [2] described a simple disease detection system for the cotton plant using its leaf image. Image of the disease-affected leaf is captured. Then using diverse image processing and Artificial Neural Network (ANN) distinguishing is performed over healthy and diseased samples. Here ANN classification accuracy is 80%.

Libo Liu et al. in [3] proposed a system for classifying diseased and non-diseased part of rice leaves using Back Propagation Neural Network (BPNN) as a classifier. Here the color features of disease-affected and healthy regions were served as input values to BPNN. Testing over 400 images of diseased and non-diseased rice leaves with BPNN, they got an accuracy of more than 90%.

Prajwala TM et al. in [4] proposed a work to detect and identify diseases in tomato leaves with the help of the slightly

978-1-7281-5284-4/20/$31.00 ©2020 IEEE

changed model called LeNet of Convolutional Neural Network (CNN). Automatic feature extraction technique is employed by the neural network model, which gives aid to the classification. Their proposed system has achieved an average accuracy of 94-95% in identifying and detecting the leaves, which indicates the feasibility of the neural network.

Sandika Biswas et al. in [5] proposed a system to determine the severity of potato late blight disease on potato leaves by using image processing and neural network. Fuzzy C-means clustering segments disease-affected area, and adds background with the same color characteristics to the images. Their proposed algorithm achieves an accuracy of 93% for 27 images.

Aparajita et al. in [6] raised an automated algorithm combining image processing and adaptive thresholding to detect late blight diseases of potato from leaf image. Statistical features of images are used to calculate the threshold value. Their proposed method is tested over 100 healthy and 100 disease affected leaves of potato crops which are obtained from plant village database and it achieved an accuracy of 96%.

Harshal Waghmare et al. in [7] developed an identification technique of grape leaf disease based on texture analysis and pattern recognition. The classification is done by support vector machine. Two types of major diseases are classified of grape plants namely downy mildew and black rot. The proposed approach achieved an accuracy of 96.6%.

Farhana Tazmim Pinki et al. in [8] worked on the classification of three common paddy leaf diseases (Brown spot, Leaf blast and Bacterial blight) and advised fertilizers or pesticides. K-means clustering is used for segmentation of the disease-affected part and Visual contents (texture, color and shape) are used as features in this automated system. Then Support vector machine classifier does the classification process. The overall accuracy of the system is 92.06%.

Namrata R. Bhimte et al. in [9] have precisely classified cotton leaf diseases. K-means segmentation (which is a color-based segmentation technique) is performed to obtain disease-affected part of a leaf image. The classification is based on extracting appropriate features such as the color and texture of the segmented part of an image. The accuracy of the classification has been rated up to 98.46%.

Md. Selim Hossain et al. in [10] proposed an image processing system that identified and classified two diseases namely brown blight and algal leaf diseases of the tea plant. Support vector machine classifier was used for recognition of these diseases. Eleven features were used for the classification process and the system gave an accuracy of more than 90%.

R. Meena Prakash et al. in [11] implemented a work, which can analyze and classify citrus leaf diseases. The whole system consists of four parts (Image preprocessing, Segmentation using k-means clustering, feature extraction and classification). The classification is done using support vector machine where 35 diseased and 25 normal images of citrus leaves are used to train and test the classifier. The proposed system claims an accuracy of 90 to 100%.

Md. Rasel Howlder et al. in [12] developed an automated Deep Convolutional Neural Network (D-CNN) based approach to recognize major guava leaf diseases namely Algal leaf spot, Whitefly and rust. They have created their dataset that contained 2705 images with four different categories of three infected and one healthy leave category. This proposed method presents eleven layers based D-CNN model, which was developed following the AlexNet framework. The proposed system produced an average accuracy of 98.74%.

III. Dataset Description

The dataset is collected from publicly available plant village database, which contains more than fifty thousand images of 14 different crop species. Three sample images of potato leaf are shown in Fig. 1. We have analyzed 450 images of potato leaves, which have following class labels assigned to them:

- Early blight affected potato leaf
- Late blight affected potato leaf
- Healthy or non-diseased potato leaf

(a) (b) (c)

Fig. 1. Sample images of potato leaf: (a) Late blight (b) Early blight (c) Healthy.

IV. Proposed Method

The detection of potato disease is followed by these stages of image processing, image segmentation, feature extraction, training and classification.

A. Image Processing

This consists of two steps: image normalization and color space conversion. In image normalization, the input images have been converted to a fixed size (i.e. 500×500 pixels).

After that color space of these normalized images has been converted. As noted here, the OpenCV library in python by default reads an image in BGR format. Therefore, it needs to be converted into an RGB coloring format for further operation. Finally, the RGB images are converted to HSV color space, which is essential for segmentation purpose here. Fig. 2. represents the color space conversions.

(a) (b) (c)

Fig. 2. Color space conversion: (a) BGR (b) RGB (c) HSV.

B. Image Segmentation

Segmentation divides an image into parts, which got a strong correlation along with the region of interest. Features acquired from a properly segmented image can lead to easy identification of healthy and diseased leaf samples. For segmentation purpose, we choose a technique based on generating masks by using color information, the intensity of color and the brightness of the HSV color space. We threshold (which is the simplest method of image segmentation) HSV image for the range of green and brown color thus it separates the region of interest from the images. Green signifies healthy and brown signifies disease in the image samples. Thresholding HSV image generates mask for a healthy and diseased potato leaf image in RGB color space. Feature descriptors then extract features from these segmented images. Fig. 3. shows the masking of a healthy potato leaf and Fig. 4. shows the masking of a disease affected potato leaf.

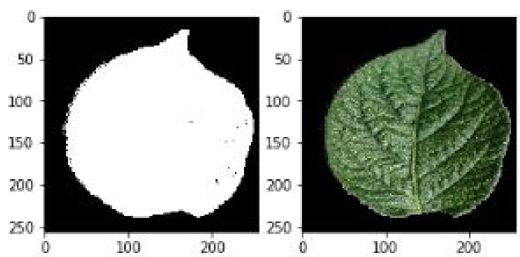

Fig. 3. Masking of healthy leaf.

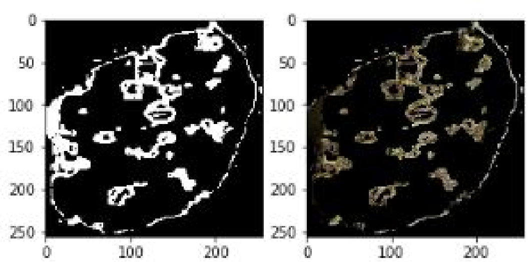

Fig. 4. Masking of disease affected leaf.

C. Feature Extraction

Features are the initial information extracted from images to identify them. We could possibly think of color, texture and shape features, which are impactful for potato leaf image.

Global Feature Descriptors (GFD)

Features of images are extracted globally with global feature descriptors. Rather than taking the interest point of an image, GFD takes the whole image for processing and feature extraction.

Here three feature descriptors are used:

1. Hu Moments:
   The shape of an object in an image is quantified by Hu Moments. It normally signifies the outlines of the object. A color image is converted to a greyscale image and computation of moments of the image happened. Then it returns vectors of shape feature.
2. Haralick Texture:
   Texture features are extracted by the Haralick Texture feature descriptor. To extract features from an image, conversion of the color image to greyscale image will happen, as Haralick feature descriptor expects images in greyscale format. The fundamental concept involved in computing the Haralick texture feature is the Grey Level Co-occurrence Matrix (GLCM).
3. Color Histogram:
   Represents the color distribution in an image and shows up the number of pixels in each color range. Color histogram descriptor computes the color intensity of an image.

Fig. 5 Represents the graphical view of the color histogram of the healthy and color histogram of a diseased potato leaf is shown in Fig. 6.

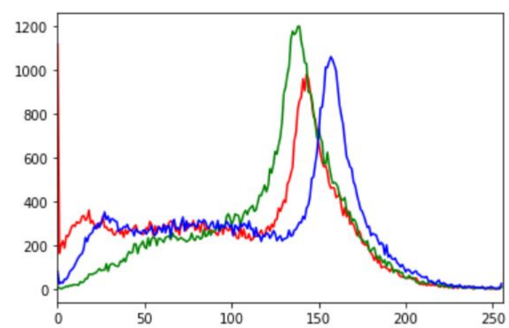

Fig. 5. Color histogram of a healthy potato leaf.

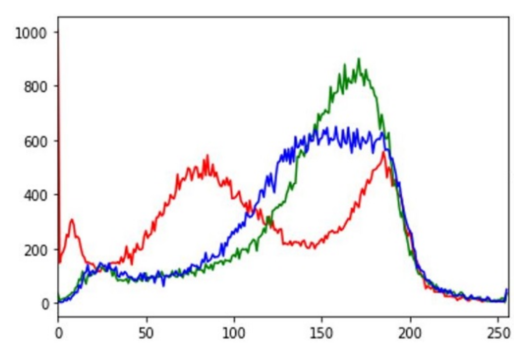

Fig. 6. Color histogram of a disease affected potato leaf.

D. Training

After extraction of three global features, feature vectors and image labeling on behalf of each image will be created. Then Random Forest (RF), Logistic Regression (LR), k-Nearest Neighbors (KNN), Decision Trees (DT), Naive Bayes (NB), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) classifiers are trained for classification purpose.

E. Classification

Global feature extraction creates feature vector on behalf of all image samples. Among them, feature vectors correspond to test dataset are supplied to our trained model that consisted of seven classifiers for prediction of result.

V. System Architecture

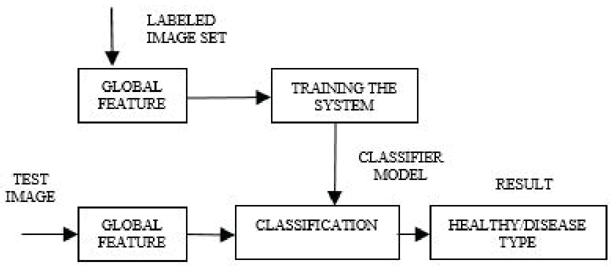

Fig. 7. System architecture.

Fig. 7 visualizes our proposed system. Here image dataset of potato leaf is firstly labeled with three different classes namely early blight, late blight and healthy. After that, global features are extracted from these labeled images by global feature descriptors and stored as feature vectors. We use 450 images of diseased and healthy leaves as our potato leaf dataset. These labeled and feature extracted images split into a training dataset and testing dataset. In this way, we train the classifier models with training dataset and test those classifiers using unseen test data.

VI. Ressult and Discussion

In this experimental study, a dataset of 450 images of potato leaves is fed to the system. Our dataset consists images of 150 early blight affected, 150 late blight affected and 150 healthy leaves. In our experiment, the dataset has been split into a training phase, which contains 80% images and the rest of the images are used for the testing portion. For classification purposes, seven classifier models are utilized. At 80%-20% train-test split Random Forest classifier gives the highest accuracy of 97% over the testing dataset. The reason for its higher accuracy than other classifiers is, it creates multiple decision trees and merges them to get more accurate predictions. Besides, the Random Forest classifier handles missing values of data which keeps the accuracy higher. For understanding the performance of our system better, 10 fold cross-validation method is applied over the classifier algorithms. The accuracy value achieved from different classifier algorithms is illustrated in Table I.

TABLE I
Accuracy Achieved From Different Classifiers

Algorithm/classifier	Accuracy value
Random Forest (RF)	97%
Logistic Regression (LR)	94%
k-Nearest Neighbors (KNN)	91%
Decision Trees (DT)	91%
Naive Bayes (NB)	84%
Linear Discriminant Analysis (LDA)	78%
Support Vector Machine (SVM)	37%

For performance evaluation of Random Forest classifier, performance parameters such as precision, recall, f1-score and support are calculated and averaged. The performance measures are illustrated in Table II.

TABLE II
Performance Measures of Classification for Random Forest Classifier

Class labels	Precision	Recall	F1-score	Support
0	1.00	0.94	0.97	32
1	0.91	1.00	0.95	31
2	1.00	0.96	0.98	27
Macro average	97	97	97	90
Weighted average	97	97	97	90

Fig. 8 visualizes the confusion matrix having test data for early blight (class label 0), late blight (class label 1) and healthy (class label 2). It represents the performance of the Random Forest classifier over the testing dataset.

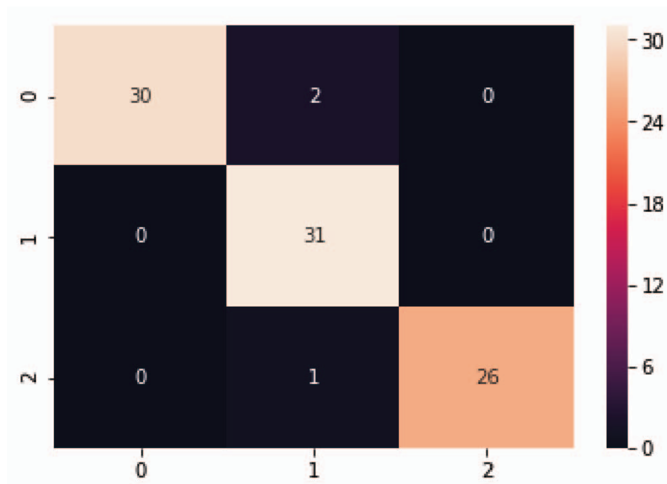

Fig. 8. Confusion matrix for test data.

Table III represents the performance comparison between different leaf disease recognition techniques.

TABLE III
Performance Comparison Between Various Recognition Techniques of Leaf Disease

Reference	Techniques	Disease affected plants	Accuracy
Proposed system	Proposed method with Random forest classifier	Potato	97%
Sandika Biswas et al. [5]	fuzzy C-means clustering and Back Propagation Neural Network	Potato	93%
Farhana Tazmim Pinki et al. [8]	K-means clustering and Support Vector Machine(SVM) classifier	Paddy	92.06%
Md. Selim Hossain et al. [10]	Image processing with Support Vector Machine(SVM) classifier	Tea	90%

VII. Conclusion

In this paper, machine learning classifiers are trained to detect and classify the two most common leaf diseases of potato plant namely early blight and late blight with the help of image segmentation. Besides, healthy leaves are also identified and classified in this process. Among the seven classifiers, the Random Forest classifier shows up with better accuracy in the detection and classification of potato leaf disease. However, we are planning to add more plant species in our system for leaf disease detection in the future.

Acknowledgment

We would like to acknowledge the ICT Division of Ministry Of Posts, Telecommunications and Information Technology, Government of the People's Republic of Bangladesh and also Computer Science and Engineering Discipline, Khulna University for supporting this work.

References

[1] M. Islam, Anh Dinh, K. Wahid and P. Bhowmik, "Detection of potato diseases using image segmentation and multiclass support vector machine," *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, Windsor, ON, 2017, pp. 1–4.

[2] Malvika Ranjan, Manasi Rajiv Weginwar, NehaJoshi, A.B. Ingole, "Detection and Classification of Leaf Disease Using Artificial Neural Network," *International Journal of Technical Research and Applications*, 2015, pp. 331–333.

[3] L. Liu and G. Zhou, "Extraction of the Rice Leaf Disease Image Based on BP Neural Network," *2009 International Conference on Computational Intelligence and Software Engineering, Wuhan*, 2009, pp. 1–3.

[4] P. Tm, A. Pranathi, K. SaiAshritha, N. B. Chittaragi and S. G. Koolagudi, "Tomato Leaf Disease Detection Using Convolutional Neural Networks," *2018 Eleventh International Conference on Contemporary Computing (IC3)*, Noida, 2018, pp. 1–5.

[5] S. Biswas, B. Jagyasi, B. P. Singh and M. Lal, "Severity identification of Potato Late Blight disease from crop images captured under uncontrolled environment," *2014 IEEE Canada International Humanitarian Technology Conference—(IHTC)*, Montreal, QC, 2014, pp. 1–5.

[6] Aparajita, R. Sharma, A. Singh, M. K. Dutta, K. Riha and P. Kriz, "Image processing based automated identification of late blight disease from leaf images of potato crops," *2017 40th International Conference on Telecommunications and Signal Processing (TSP)*, Barcelona, 2017, pp. 758–762.

[7] H. Waghmare, R. Kokare and Y. Dandawate, "Detection and classification of diseases of Grape plant using opposite colour Local Binary Pattern feature and machine learning for automated Decision Support System," *2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN)*, Noida, 2016, pp. 513–518.

[8] F. T. Pinki, N. Khatun and S. M. M. Islam, "Content based paddy leaf disease recognition and remedy prediction using support vector machine," *2017 20th International Conference of Computer and Information Technology (ICCIT)*, Dhaka, 2017, pp. 1–5.

[9] N. R. Bhimte and V. R. Thool, "Diseases Detection of Cotton Leaf Spot Using Image Processing and SVM Classifier," *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2018, pp. 340–344.

[10] S. Hossain, R. M. Mou, M. M. Hasan, S. Chakraborty and M. A. Razzak, "Recognition and detection of tea leaf's diseases using support vector machine," *2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*, Batu Feringghi, 2018, pp. 150–154.

[11] R. M. Prakash, G. P. Saraswathy, G. Ramalakshmi, K. H. Mangaleswari and T. Kaviya, "Detection of leaf diseases and classification using digital image processing," *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, Coimbatore, 2017, pp. 1–4.

[12] M. R. Howlader, U. Habiba, R. H. Faisal and M. M. Rahman, "Automatic Recognition of Guava Leaf Diseases using Deep Convolution Neural Network," *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Cox'sBazar, Bangladesh, 2019, pp. 1–5.