PERFORMANCE COMPARISON OF MACHINE LEARNING CLASSIFIERS FOR THE DETECTION OF POTATO LEAF DISEASES

Twinkle
Department of Computer Systems Engineering
Mehran University of Engineering and Technology
Jamshoro, Pakistan
twinklemaheshwari04@gmail.com

Sammer Zai
Department of Computer Systems Engineering
Mehran University of Engineering and Technology
Jamshoro, Pakistan
sammer.zai@faculty.muet.edu.pk

Sanam Narejo
Department of Computer Systems Engineering
Mehran University of Engineering and Technology
Jamshoro, Pakistan
sanam.narejo@faculty.muet.edu.pk

Muhammad Ahsan Ansari
Department of Computer Systems Engineering
Mehran University of Engineering and Technology
Jamshoro, Pakistan
ahsan.ansari@faculty.muet.edu.pk

Abstract**—**Potato is a widely used crop in Pakistan whose production may be significantly declined if proper care is not taken at an early stage due to various diseases. An automatic early detection for diseases such as Early blight and Late blight can be done using various machine learning based algorithms. However, accuracy plays a vital role for an exact detection of these diseases. Therefore, the purpose of this study is twofold. Firstly, we aim to detect the diseases associated with that of the potato crop. Secondly, the classification of the detected disease is performed by using three different classifiers including Naive Bayes (NB), Support Vector Machine (SVM) and Convolutional Neural Network (CNN). Moreover, the performance of these classifiers has been evaluated through various performance measures on a publicly available dataset which exhibits that CNN obtains the highest accuracy of 91% approximately.

Keywords: *Plant disease, Machine Learning, Classifiers, Image Processing.*

I. INTRODUCTION

Potatoes have been an important crop for farmers as well as consumers in Pakistan over the years. By volume of production, it is the fourth most important crop and it gives farmers high yields and high returns [1]. Potato is a vegetable crop of significant economic value in Pakistan. The pests and illness that occur in Potato plants influence these production volumes [1]. Many plant diseases can decrease agricultural production , causing farmers to suffer with tremendous losses. Any of such diseases can't be seen by the human eye. Even if the disease is apparent to humans, the identification and classification of plant disease is not a simple task, it requires constant monitoring that can be exhausting and costly. In addition, waiting until the signs are apparent in order to take any steps to treat them is not helpful. Maybe it's too late to act. As plants are important for both people and animals to survive, the best modern technology should be supplied to farmers. These technologies should be able to recognize and classify, in a short time, a wide range of plant diseases. The identification of plant diseases in the early stages will dramatically reduce crop losses. Use of technology for human aid has been fast and accurate in the past. Researchers have also worked in the area of leaf diseases detection using various machine learning algorithms and they prove to be fast and accurate than conventional manual methods. An example image of potato leaf along with common diseases is shown in Fig. 1.

(a) (b) (c)

Fig. 1. Samples of Potato plant diseases; (a) Early Blight (b) healthy (c) Late Blight.

The goal of this study is to build a model that uses machine learning algorithms to accurately identify leaf diseases specifically early blight and late blight at the early stage of infection. For the classification of diseases from plant leaves different classification methods of supervised machine learning such as NB, SVM, and CNN are being used for this study and a comparison is also made between these classification techniques.

Remainder of the paper is organized as follows. Literature review has been described in Section II. Section III discusses the proposed methodology in detail. Section IV describes the efficiency of the proposed method with the help of experimental results. However, conclusion and discussion for future work has been presented in Section V.

II. RELATED WORK

To detect unhealthy regions in leaf images, the automated pixel-based classification method is proposed [2]. The algorithms underwent extensive studies. To identify each pixel, linear SVM was used and it has also shown how the outcomes of SVM could be better-quality of neighborhood search. Their algorithm primarily relies on color image and takes three phases to complete. First, by segmentation, the image is divided into the foreground and background. In the second stage, to predict the class of each pixel belonging to the foreground, SVM is applied. Finally, by neighborhood-check, the process of refinement is continued to omit all wrongly classified pixels.

Another study in the literature [3] has used the image processing and artificial intelligence models for the detection of leaf diseases. However, their recognition and characterization was limited to only cotton leaf diseases. Moreover, this study has presented a survey on context elimination and segmentation methods which has shown that color space conversion from RGB to HSV is useful for background removal. In their method, color segmentation was performed by masking green pixels in the deleted image in the background and then applying thresholding to the obtained masked image to get a binary image which has proved to be helpful in extracting the correct characteristics of disease. This is helpful in extracting correct disease characteristics.

Deep learning techniques may also be used to classify plant diseases, such as the Convolutional Neural Network (CNN). One such work has been presented in the literature where CNN model was developed to classify potato leaves into three classes: stable leaves, early blight, and diseased leaves with late blight [4]. This paper has used a dataset of almost 1500 images with 500 leaves belong to each class. This model learns the features from the raw images automatically and can identify the images with good accuracy.

Another algorithm has been presented in the literature for leaf disease recognition which is a combination of machine learning and image processing based techniques [5]. Specifically, in their study, their focus is on the use of RGB images, due to the low cost and high availability of digital RGB cameras.

A classifier differentiating healthy leaves as well as those infected by late blight and early blight diseases has been designed by Islam and their co-workers [6]. The Multiclass SVM has been used to identify image datasets into one of three classes based on 10 color and shape features. In their work, they have used a dataset of 300 potato leaves which have been drawn from the Plant Village dataset (https://www.kaggle.com/emmarex/plantdisease). However, it would have been more useful to provide the system's performance on a larger dataset.

To recognize leaf disease, a Back-Propagation (BP) network has been used with various combinations of shape, color and texture characteristics comprising the input feature vector with reduced dimensionality using PCA [7]. However, their algorithm was applicable to grape and wheat leaves only.

Qin et al. [8] have proposed a solution to classify four leaf diseases affecting alfalfa grass. In their method, images were cropped to produce sub-images of one or more lesions and SVM was the best performing classification algorithm, offering 97.64 percent and 94.74 percent recognition accuracy, respectively, on the training and testing sets.

Usama Mokhtar et al. [9] has proposed a method where they identified and detected the diseases of Tomato leaves: Powdery mildew and Early blight. In this approach, numerous methods have been used in image pre-processing, such as surface smoothing, noise removal, image resizing, image isolation and background removal for contrast enhancement. In function extraction for function vectors, the Gabor wavelet transformation has been used in classification. Whereas, in SVM for performance decision and disease identification training, Cauchy Kernel, Laplacian Kernel and Invmult Kernel are used.

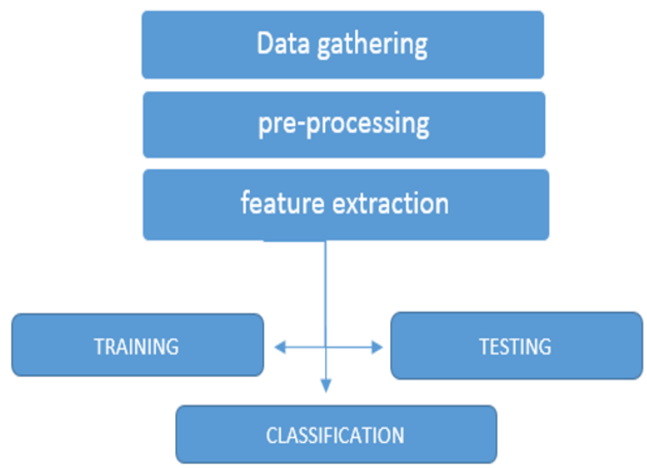
A methodology for the detection of leaf diseases using digital image processing techniques has been proposed [10]. In their study, the experimental findings has shown that four major plant leaf diseases can be successfully identified and categorized by their proposed system.

Researchers have recorded high accuracy of recognition on specific datasets, but in many instances, when evaluated on various datasets or in field environments, the performance of those systems has degraded significantly. In this paper, first we aim to detect the diseases associated with that of the potato crop and then perform the classification of the detected disease using three different classifiers including Naive Bayes (NB), Support Vector Machine (SVM) and Convolutional Neural Network (CNN).

III. Proposed Methodology

This paper presents an efficient approach for the detection and classification of diseases found in the leaves of potato crop. The proposed method is a blend of image processing and Machine learning based techniques. The overall step-by-step process of the proposed methodology has been presented in Fig. 2. However, following subsequent sub-sections provide the details of each steps given in the block diagram of Fig. 2.

Fig. 2. Proposed Methodology



A. Data Gathering

In order to prove the efficiency of any algorithm, data collection plays an important role. As specified by Fig.2, the very first step of the proposed method is to collect the data. In this study, a publicly available dataset has been used known as PlantVillage (https://www.kaggle.com/emmarex/plantdisease). This dataset includes a vast amount of data which is about 50,000 images that belongs to 14 crop species and 26 different diseases. Out of which, 2119 images belong to potato crop. Furthermore, there are three (3) classes of Potato images. Classe-1 consists of images infected with Early Blight disease, Classs-2 comprises of healthy images, and Class-3 consists of images with Late Blight disease.

B. Pre-processing

Every image in the dataset can be unique in its own way and algorithms might give an unreliable output while processing such images. To avoid this uncertainty in the output, pre-processing of these images is required to bring the images to same grounds. Images are resized and brought to same size. To make the quality of images better, the noise in the images is removed highlighting only the leaf area. For this purpose, we have applied the best noise removal algorithm known as "non local means denoising" to remove the Gaussian noise from a color imageand the results are exhibited in Fig.3.

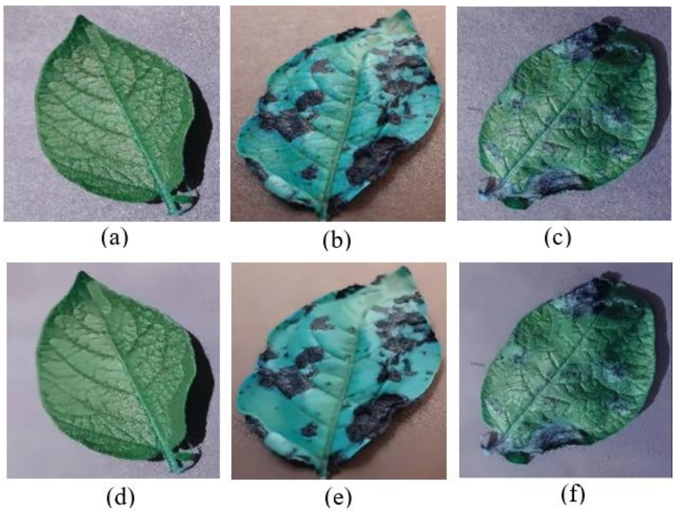


Fig. 3. Noise Removal (a) Healthy Image (b)Image with Early Blight disease (c) Image with Late Blight disease (d) Noise Removal applied on (a). (e) Noise Removal applied on (b). (f) Noise Removal applied on (c).

C. Feature extraction:

After the noise is removed from an input image, feature extraction is the most important step of the whole process. To extract the feature, the given image has to be converted into machine readable language by saving the image as an array. In this way each pixel can be saved and treated separately. Color characteristics such as RGB values and Hue (H), Saturation (S) and Intensity (I) values are derived to ensure the precision in the identification of the disease.

D. Design and development

After getting the features the images are run for training and subsequent testing for all the three classes of Potato.

E. Evaluating the performance of different classifiers.

In the final step the classification is carried out using three different classifiers NB, SVM and CNN. The results of each of these classifiers are compared with each other to find out the most suitable classifier. The results are shown in terms of confusion matrix for all these three classifiers.

IV. EXPERIMENTAL RESULTS

This segment discusses the performance of different classification methods, such as NB, SVM, and CNN on Potato disease detection dataset. The dataset for this study is divided into training data (80 percent) and testing data (20 percent). Potato plant disease dataset comprises a total of 552 images and three class labels. Details of the classes of dataset that are being used are as follows: late blight Potato, early blight Potato and healthy are 200, 200 and 152, respectively as shown in Fig. 4 and Fig. 5, respectively.

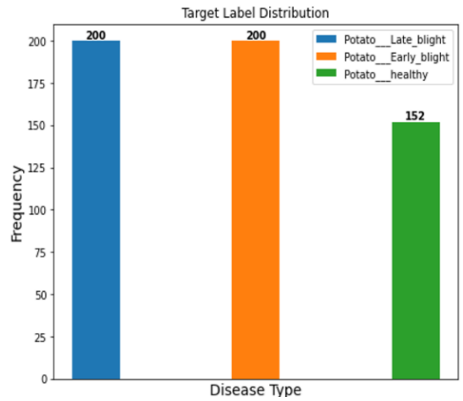


Fig.4. Number of images for each label class.

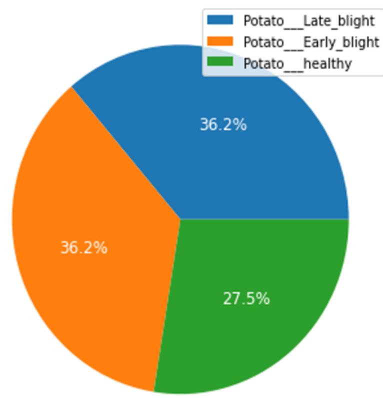


Fig.5. Number of images for each label class.

The performance of Naive Bayes classifier is shown in Fig. 6 as confusion matrix. From this graph it can be seen that early blight is predicted most accurately as compared to that of late blight and healthy. Whereas late blight is predicted poorly with good portion being detected as healthy. Also more number of healthy images are misinterpreted as late blight then early blight.

Figure 7 represents the confusion matrix of SVM classifier. It is evident form the graph that it has followed the same pattern as Naive Bayes classifier with early blight being most accurate and healthy images being poorly detected. But in this model no amount of healthy images is being misinterpreted as early blight.

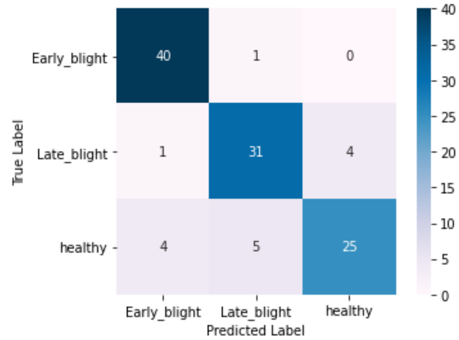


Fig.6. Confusion matrix for Naive bayes classifier

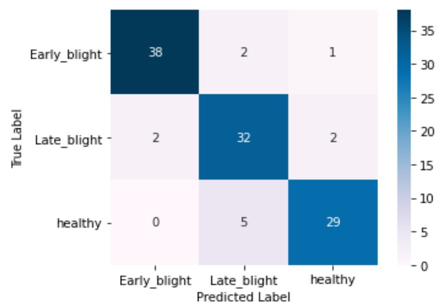


Fig 7: Confusion matrix for SVM

The confusion matrix for CNN classifier is shown in Fig. 8, where it can be observed that it follows quite the reverse pattern than that of those other two models. In this model healthy images are being detected most accurately with no image being misinterpreted as Early blight and only a few as Late blight.

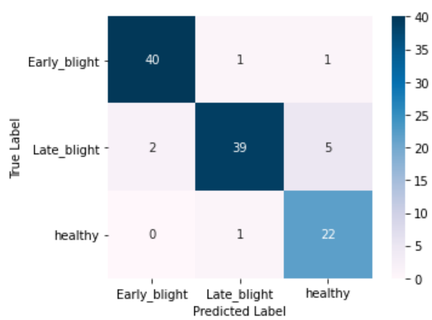


Fig 8: Confusion matrix for CNN

Here early blight is detected poorly as compared to other two sample types. Similarly, Late blight is also predicted very accurately with only one image each being misinterpreted as healthy and early blight. Fig. 9 and Fig. 10 shows the accuracy and loss of CNN model.

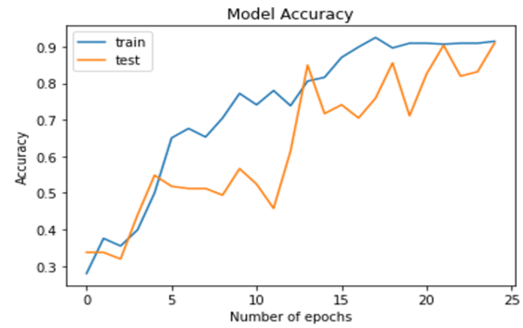


Fig 9: Accuracy of training and testing for CNN model

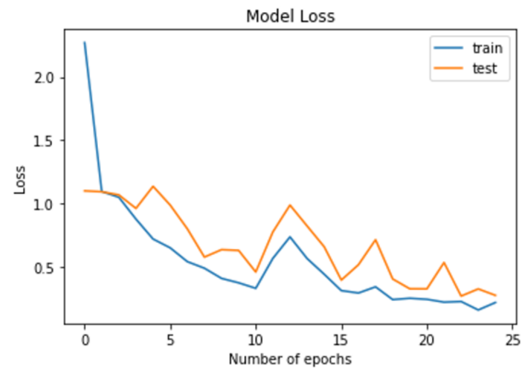
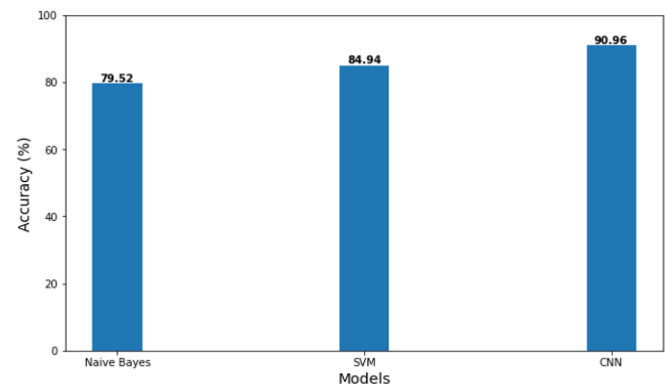


Fig 10 : Loss of training and testing loss CNN model

The classification accuracy of all the three models is



presented in Fig.11. It is clear from the graph that CNN model is giving best accuracy (90.96%) in comparison to SVM and Naive Bayes model.

Fig 11. Accuracy comparison of three models

V. CONCLUSION

In this work, disease detection algorithm has been proposed. The proposed method detects and classify the two common Potato leaf diseases: Early blight and Late blight. This study has also compared the three machine learning algorithms namely Naïve Bayes, Support Vector Machine, and Convolutional Neural Network. It has been observed from the experimental results that CNN classifier attains the highest accuracy.

Acknowledgment

The authors would like to thanks Mr. Vikash Choudhary for his assistance during this research. The authors would also like to extend their thanks to Mehran University of Engineering and Technology, Jamshoro, Pakistan, for providing us the resources necessary to conduct this research.

References

[1] Khokhar, khalid, "Present 2status and prospects of tomatoes in pakistan", Agriculture corner,10.13140/rg.2.1.1482.5449,2013.

[2] Madhogaria, Satish, M. Schikora, W. Koch and D. Cremers. "Pixel-based classification method for detecting unhealthy regions in leaf images." GI-Jahrestagung ,2011.

[3] B. S. Prajapati, V. K. Dabhi and H. B. Prajapati, "A survey on detection and classification of cotton leaf diseases", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai, 2016, pp. 2499-2506.

[4] Rahul Sharma ,Amar Singh,Vinus Sharma "Potato Leaf Diseases Identification Using CNN", International Journal of Emerging Technologies and Innovative Research, ISSN:2349-5162, Vol.5, Issue 12, December 2018,pp 519-527.

[5] Ngugi, Lawrence C., Moataz Abelwahab, and Mohammed Abo-Zahhad. "Recent advances in image processing techniques for automated leaf pest and disease recognition A review", Information processing in agriculture,2020.

[6] Islam, Monzurul, Dinh, Anh, Wahid, Khan ,Bhowmik, Pankaj, "Detection of potato diseases using image segmentation and multiclass support vector machine", IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE),2017,pp. 1-4.

[7] Wang, Haiguang, Guanlin Li, Z. Ma and X. Li. "Image recognition of plant diseases based on backpropagation networks." 2012 5th International Congress on Image and Signal Processing ,2012 pp.894-900.

[8] F. Qin, D. Liu, B. Sun, L. Ruan, Z. Ma, H. Wang, "Identification of alfalfa leaf diseases using image recognition technology", PLoS ONE,2016, pp. 1-26.

[9] Usama Mokhtar, Mona A. S. Alit, Aboul Ella Hassenian, Hesham Hefny, "Tomato leaves diseases detection approach based on support vector machines", IEEE , 2015,pp. 978-1-5090-0275-7/15.

[10] Min oo, Yin; Chi htun, nay, "Plant leaf disease detection and classification using image processing", International Journal of Research and Engineering, Vol. 5, n. 9, Nov. 2018. , pp. 516-523.