INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078

(Volume 7, Issue 2 - V7I2-1300)
Available online at: https://www.ijariit.com

Plant disease detection using ensemble learning – A review

Sadhana
sadhana.dhanjal@gmail.com
DAV Institute of Engineering and Technology, Jalandhar, Punjab

Dr. Vinay Chopra
vinaychopra222@yahoo.co.in
DAV Institute of Engineering and Technology, Jalandhar, Punjab

ABSTRACT

Agriculture is one of the strongest pillars in Indian economy. Plants are important as they are the source of energy supply to mankind. Approximately 70% of Indian economy depends on agriculture for their livelihood. But this is affected by disease which cause the lower agriculture productivity. The farmers encounter difficulties in their detection of plant diseases. Diseases can affect the plant at the time of sowing and harvesting and it leads to low productivity and lowers the economy level. So, there is a need to mitigate this issue by computer techniques and machine learning methods. This paper presents an overview of various classification techniques in machine leaning that helps in plant disease detection. Emergence of accurate techniques leads to impressive results.

Keywords: *SVM, KNN, Decision Tree, Random Forest.*

1. INTRODUCTION

In any developing nations, agriculture plays important role. Agriculture is the spine of economy. Therefore, any harm to harvesting would lead to production loss and ultimately a big loss to economy growth of country. The quantity and quality of crop production is based on the growth of plant. So it is important to detect them at the earlier stage. After detection, some necessary steps should be taken to prevent it to spreading to other regions in field. Dates back ago, the farmers identified diseases by monitoring their changing color and shapes but it is not possible in the large fields. And also it takes more time and much hard work which is not possible now a days. There is a major need to time efficient and automated diagnosis methods for improving the rate of crop production.

2. CLASSIFICATION OF PLANT DISEASES AND THEIR SYMPTOMS

The plant leaves are affected by fungal, bacterial and viral diseases which include rust, powdery mildew, Downey mildew, brown spots etc.[1]. The primary diseases of plants are due to fungus , viral and bacterial diseases like Alternaria, anthracnose, bacterial blight, canker and leaf spot etc[2]. Some common types of diseases appear on vegetables[3] – Spotted Wilt of Tomatoes and Peppers, Southern Blight, Mosaic, Fruit Rot, Rust, Numatode Diseases etc.

1. First type of symptom– These were scattered small and consists of a no. of small lesions or spots spreads over the leaf surface(Fig. 1a). Here, two criteria were used – relatively isolated symptoms were taken individually(Fig.1b), and lesions were part of clusters were taken as group(Fig. 1c)[4].

2.Second type of symptom– These were scattered large and consists of no. of large lesions or spots spread over the leaf surface(Fig. 2a). In half of the cases, the spurious lesions were blacked out(Fig. 2b), and in other half, the spurious lesions were kept unchanged(Fig, 2c). This was done to increase the diversity of conditions[4].

3.Third type of symptom– These were isolated in nature and consists of single lesions or spots(Fig. 3a). In this case, there is only one new figure is spawned from original leaf sample (Fig. 3b). In some cases lesions were split into clearly distinct regions (Fig. 3c)[4].

4.Fourth type of symptom– These are widespread and consists of large lesions that spread over the leaf surface (Fig. 4a). The rest of the subdivisions were done by detecting relatively homogeneous regions within diseased tissues [4](Fig. 4b, 4c).

5.Fifth type of symptom– These are in powdery form and consists of powdery spots on leaf surface(Fig. 5a). When spots were isolated, each one of them generated a new image(Fig. 5b). When disease was more widespread, divisions followed same rule as widespread lesions(Fig. 5c)[4].

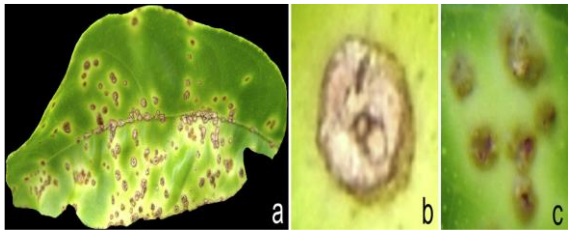

Fig. 1: Example of scattered small symptoms (a), isolated lesion (b), and cluster of lesions (c).

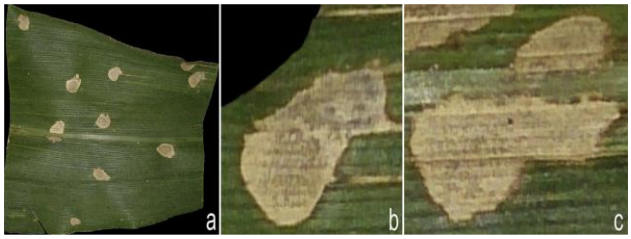

Fig. 2: Example of scattered large symptoms (a), spurious lesion blacked out (b), and spurious lesion unchanged (c).

Fig. 3: Example of isolated symptoms (a), lesion region delimited (b), and lesion with two visible regions (c).

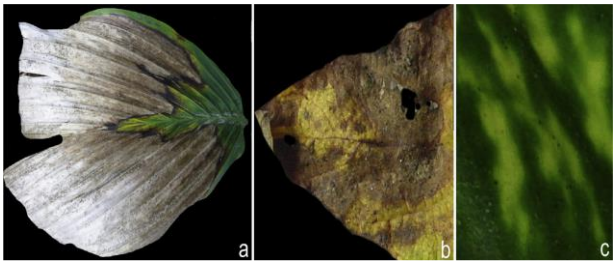

Fig. 4: Example of widespread symptoms (a), lesions delimited by homogeneity (b and c).

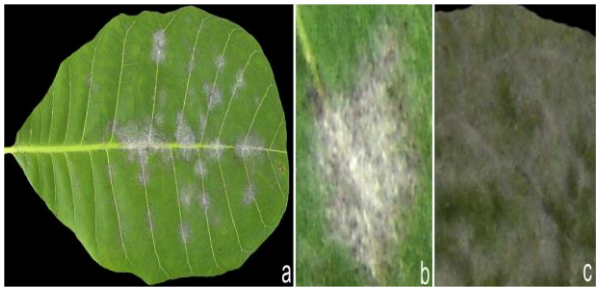

Fig. 5: Example of powdery symptoms (a), isolated spots (b), and widespread spots (c).

3. PROPOSED WORK

Image Processing Techniques– To identify leaf infection, various image processing techniques has been introduced. There are different stages involved in plant disease detection system using image processing. The stages are as follows:

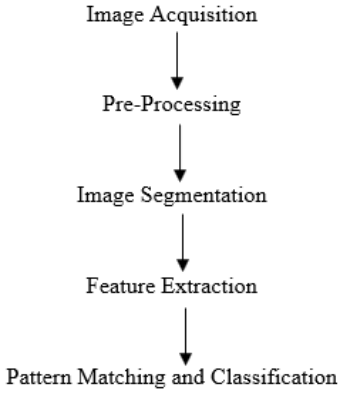

Fig. 6: Stages of Image Processing Technique

- *Image Acquisition*– This is the prior stage of proposed method. For proposed work, the images of plant leaves are collected from Plant village(CrowdAI) dataset and Self-Acquired dataset. The collected images are invariant of their angle, rotation, light etc.[5].
- *Pre-Processing*– This is the second stage in which the collected images are pre-process or refine for noise removal, resizing, contrast or brightness enhancing etc.[5].
- *Resizing*– The captured image or its region of interest is not clearly visible sometimes so all the images are resized under pre-processing method. Mostly KNN method is used for resizing of image.
- *Noise Removal*– It is undesirable effect that alters the color and pixel appearance. It causes variation in color and brightness of clicked image. Noise depends upon the nature and methodology under which the image clicked. Noise is of different type like Gaussian noise, shot noise and salt and pepper noise etc. To remove these noises, various filter are used depending upon the type or nature of noise. Gaussian Filter, Vector Median Filter and hybrid filter are some filters to remove noise in color images[6].
- *Contrast and Brightness*– Image Enhancement process is used to adjust the digital image so that the results are more suitable for display or for further analysis of image.
- *Image Segmentation*– K-means algorithm is used foe image segmentation. It is very simple and efficient method for segmentation. K-means algorithm divides N observations among K clusters by assigning an observation into a cluster using distance function generally an Euclidean distance function.
- *Feature Extraction*– In this stage, the features are extracted from the image like color feature, texture feature. The process of extracting interested set of values called features from image is called feature Extraction. These feature provide information about image for more processing. Mostly Gray Level Co-occurrence matrix(GLCM) is used for texture feature extraction and Global color Histogram, Image bitmap and local color histogram is used for color feature extraction. Color histogram gives the representation of colors in image. In this, the RGB image is converted to HSV. HSV model aligns closely with how the human eye depicts the colors in an image[7]. The other features extracted from leaf image are – Shape features, vein features, Gabor Wavelet Transform Features, Zernika Moment Features etc.[8].
- *Pattern matching and Classification*– Pattern matching is the process of analyzing stated sequence of tokens for the subsistence of components of some pattern. The purpose of matching algorithm is to make comparison of features with index features of image occurring in dataset. The process of recognizing the category of observed pattern is called classification. Classification is of two types: Supervised and unsupervised.

4. CHALLENGES IN PLANT DISEASE DETECTION

1) *Collection of dataset*– Creating an image dataset is the main requirement of image processing. One has to travel to various locations for capturing images of plant disease. It is a challenging task to collect data due to the unavailability of different varieties of crop infections may at some farms. Sometimes, diseases occur only during some seasons.
2) Image Background– an important stage of image processing is Image Segmentation. Mainly required part of the image is separated in this process. It is difficult to apply leaf image segmentation because of the existence of plants, leaves and some other green components in backdrop.

3) *Image Capture condition*– Stable and component results are provided by automatic plant disease detection system only if all the pictures are clicked under similar condition. It is possible to capture image under same condition just in inside laboratories. Due to uncontrollable environment, image capturing under same condition within the farm is a difficult task.
4) *Symptom Segmentation*– Most plant infection indicators have no distinct boundaries. They fade on plants gradually due to which there will not be a good segmentation. This phenomenon affects the ultimate result.
5) *Symptom variations*– Symptoms are based on environment, infection and plant. Any alteration in these elements can cause change in symptoms. Identifying plant infection with symptoms disparity is a difficult task.

5. RESEARCH PROCESS

1) *Machine Learning Methods*-Many classification methods are developed by researchers. SVM is used for classification in plan disease detection. SVM, KNN lies under supervised learning technique.
2) *Support Vector Machine*– SVM is extremely popular and high-performing algorithm to analyse data for classification. It is effective in high dimensional spaces. SVM is versatile in nature. In SVM, data points are scattered into n-dimensional space where n represents the features taken into consideration. SVM is the classification of data on the basis of relation with other related datasets. Hyperplanes are the boundaries that help to classify data points. The main aim is to maximize the margin between the hyperplane and data points.
3) *K Nearest Neighbour*– It is used for statistical estimation and pattern recognition. It is simple, easy, flexible, and robust to noisy training data but its computation cost is higher. It has two phases – training and testing phase. In training phase, the leaf images are labelled with their classes and in testing phase the leaf images are unlabelled and algorithm yields the list of k nearest data points to label the unlabelled points and classifies their classes[9].
4) *Decision Tree*– It is also a supervised classification and regression in supervised learning that builds classifiers by dividing the data into several smaller groups(in a tree like structure). Gini Index and entropy are the main attributes that used as disparity measures. Decision tree can produce less amount of training errors. There are different variants of decision trees – ID3, C4.5and CART etc.
5) *Ensemble Method*– These are the techniques that create multiple models and then combine them to produce improved result. It usually produces more accurate solutions than a single model would. This has been the case in a no. of machine learning competitions, where the winning solutions used ensemble methods.
6) *Random Forest*– RF is an ensemble of randomized decision tree classifier learning method. Rf is operated by constructing multiple decision trees at training time. The class labels of the testing dataset are measured based on the voting of each classification tree. The outcome of classifier depends on the class labels that have maximum voting by the classification trees. RF uses bagging and randomness of features during building of each individual tree.

6. EXPERIMENTAL RESULT ANALYSIS

This section investigates the performance of different classification techniques such as SVM, KNN, DT and RF on plant disease detection and found that the RF classification technique provides higher accuracy than the others[10].

Table 1: Comparison between classification techniques

Classifiers	Accuracy(%)
SVM	77.56
KNN	76.16
DT	74.35
RF	79.23

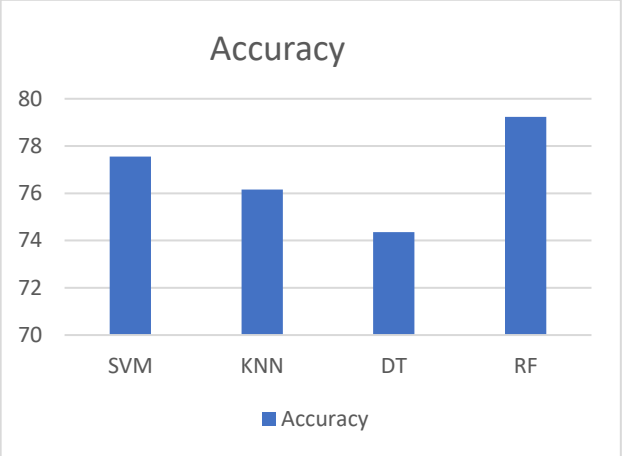

Fig. 7: Classification Accuracy of Classifiers

7. CONCLUSION

This comparative study carried out different machine learning classification techniques on digital image processing to detect plant disease. Different pre-processing and segmentation techniques were discussed that helps to extract features. This paper also compares the number of classifications which is considered by researchers for evaluating accuracy results. From this survey, Random Forest becomes the better option for disease detection.

8. REFERENCES

[1] L. Sherly Puspha Annabel, T. Annapoorani and P. Deepalakshmi "Machine Learning for Plant Leaf DiseaseDetection and Classification – A Review" International Conference on Communication and Signal Processing, April 4-6, 2019, India
[2] Eftekhar Hossain, Md. Farhad Hossain and Mohammad Anisur Rahaman " A Color and Texture Based Approach for the Detection and Classification of Plant Leaf Disease Using KNN Classifier" 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019
[3] Mukesh kumar Tripathi, Dr. Dhananjay D. Maktedar "Recent Machine Learning Based Approaches for Disease Detection and Classification of Agricultural Products" Conference Paper · August 2016 DOI: 10.1109/ICCUBEA.2016.7860043
[4] Jayme Garcia Arnal Barbedo" Plant disease identification from individual lesions and spots using deep learning" Embrapa Agricultural Informatics, Av. Andr_e Tosello, 209, C.P. 6041, Campinas, SP, 13083-886, Brazil
[5] VIPPON PREET KOUR AND SAKSHI ARORA" Particle Swarm Optimization Based Support Vector Machine (P-SVM) for the Segmentation and Classification of Plants" IEEE Access, Digital Object Identifier 10.1109/ACCESS.2019.2901900
[6] J. Pei, H. Fan, X. Yang, and X. Feng, ``A hybrid _lter with neighborhood analysis for impulsive noise removal in color images," Signal Process., vol. 153, pp. 197_209, Dec. 2018.

[7] Shima Ramesh, Niveditha M, Pooja R, Prasad Bhat N, Shashank N, Mr. Ramachandra Hebbar, Mr. P V Vinod "Plant Disease Detection Using Machine Learning" 978-1-5386-7523-6/18/$31.00 ©2018 IEEE DOI 10.1109/ICDI3C.2018.00017

[8] Nilay Ganatra and Atul Patel "A Multiclass Plant Leaf Disease Detection using Image Processing and Machine Learning Techniques" International Journal on Emerging Technologies 11(2): 1082-1086(2020)

[9] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient knn classification with different numbers of nearest neighbors," IEEE transactions on neural networks and learning systems, vol. 29, no. 5, pp. 1774–1785, 2018.

[10] Kshyanaprava Panda Panigrahi, Himansu Das, Abhaya Kumar Sahoo and Suresh Chandra Moharana "Maize Leaf Disease Detection and Classification Using Machine Learning Algorithms" ResearchGate DOI: 10.1007/978-981-15-2414-1_66

