Grenze International Journal of Engineering and Technology, Jan Issue

Comparative Analysis of Machine Learning based Plant Leaf Disease Detection Systems

Riyazahemed A Jamadar[1], Dr. Anoop Sharma[2] and Dr. Kishor Wagh[3]
[1]Research Scholar Department of Computer Science& Engineering Singhania University Rajasthan, India
[2]Professor, Department of Computer Science& Engineering Singhania University Rajasthan, India
[3]Professor, Department of Computer Engineering, AISSMS IOIT Pune, India

***Abstract*—Recently the availability of Machine Learning(ML) algorithms through state of the art programming interfaces & tools has lead the research community to employ ML based object detection/classification algorithms & techniques for the research work with improved accuracy & performance. There is a significant amount of research work that employs ML based approaches for plant leaf disease detection & classification. A systematic study and comparative analysis of these ML algorithms provide a good insight to researchers and guide them to select & blend these ML algorithms for higher accuracy and performance. In this paper an effort is made to establish comparative analysis of the most recent work done in this domain, The ML algorithms have been analyzed against parameters like the data set, preprocessing techniques, feature extraction methods and accuracy, and the type of diseases that could be detected. Every analysis concisely discusses the characteristics and scope for further extensions of the proposed technique/method.**

***Index Terms*— Leaf Disease Detection, Machine Learning, Image Processing, Accuracy.**

I. INTRODUCTION

Agriculture is one of the major sources of revenue building for country like India. The efforts and toil of farmers they take in cultivations should be proportionately rewarded. Due to climate change, the weather conditions have become erratic, eventually, the plants are more susceptible to common plant diseases, which primarily affect the leaves and fruits. The diseased plants deteriorate the quantity and quality of production by 10% to 50% of the total crop production yearly.[1]. The use of modern technology must mitigate this by early detection of the plant diseases and help the farmer to take corrective measures and save the crop. In this regard there is a lot of research going on. The primary work is to identify the diseased leaf/fruit through its image. Thus there are image processing techniques that classify the input image as healthy or unhealthy. As these methods are not producing the desired accuracy level of classification, researchers have opted to blend these image processing techniques with machine learning & deep learning algorithms. Here an effort is been made to compile such work which is based on machine learning algorithms. As it's the most trending and promising technology, the recent work in precision agriculture is invariably employing a blend of machine learning & deep learning algorithms. There is tremendous work going on in this field and has wide future scope with every day's new arrivals of state of the art algorithms that are optimizing the resources required with outperforming in accuracy and precision. In this work an effort is made to analyze the machine learning algorithms, so that off-the-shelf one can pick up the algorithm and know its performance in terms of accuracy and precision. The spectrum of machine learning algorithms that

Grenze ID: 01.GIJET.9.1.507_2

are reviewed here used primarily for classification and some are used for image segmentation as well. Section-II provides brief introduction of the most common and widely used machine learning algorithms, with merits and demerits of each technique. Section – III provides a detailed work carried out by authors, Section-IV discussion on machine learning based leaf disease detection techniques and finally Section-V discusses the conclusion of the review work carried out here with challenges and future scope.

II. Discussion on Machine Learning Algorithms

Machine Learning is a branch of Information Technology that primarily deals with Artificial Intelligence where the machine (computer) learns from the previous experience/input data and is able to predict the future output/outcome. Basically there are two broad categories, Namely Supervised Learning and Unsupervised Learning. In Supervised Learning labeled data is used for training, where as in Unsupervised Learning, there is no labeled data. Besides these two, there is another type of learning process called Reinforcement Learning where the machine automatically determines the ideal behavior of the system. Fig.1 shows the steps involved in Machine Learning

***Data Collection/Gathering*:** This is the primary work the researcher has to do. It is an important task because further phases depend on the quantity and quality of this data. Normally researcher use off-the-shelf available data from databases available on internet. This data has lot many features, it can be processed and required features are obtained efficiently. For plant leaf disease detection, the public data set like Plant village data set can be used or the researcher can create his/her own data set of images by taking real-time images of diseased and healthy leaves.

***Processing the raw Data*:** Normally the raw data has lot of noise and unwanted information. To train the model precisely, utmost care should be exercised to clean it. Here comes the need of processing the data. Here for in plant leaf disease detection, the raw dataset of images is processed through various image processing techniques and a region of interest (ROI) is obtained.

***Training Model*:** In this phase , depending upon the requirement, the system is trained using any of the relevant machine learning algorithm. Through training dataset, the model made to learn all the possible scenarios.

***Testing/Evaluation Model*:** Once the model is trained with the training data set, it is tested against the test dataset and it's evaluated for the expected outcomes.

***Performance Improvement*:** The classification accuracy or performance can be enhanced to the expected level, by blending different machine learning algorithms, with augmented datasets.

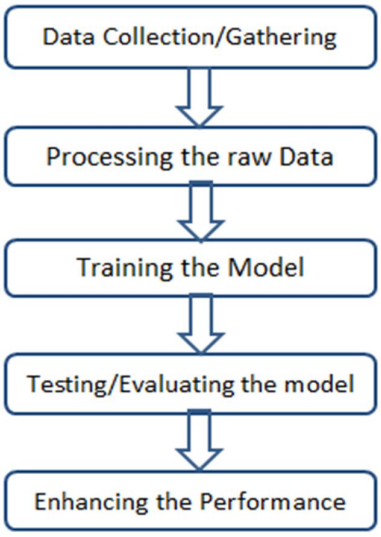

Fig.1 Steps involved in Machine Learning

The most commonly used classification algorithms are Support Vector Machine, K-nearest neighbor, Nai`ve Bayes, Decision Tree and Random Forest. Most of the research work carried out in plant leaf disease detection use invariably support vector machine algorithm for classifying diseased leaves and healthy leaves K-NN belongs to supervised machine learning where the dataset has labeled record. Decision tree is the most common

machine learning algorithm used for classification sort of things. The naive Bayes algorithm is again a very common algorithm that performs classification tasks in the field of machine learning. It is one of the best suited algorithm for the dataset with large size.

III. Review of Existing Work

In this section, works of various researches in this domain are discussed, in regards to the kind of dataset, image processing technique, and classification algorithms with accuracy obtained. The authors Kuldeep Singh et al. [4] suggested a method to identify the early symptoms through which a plant's rust disease of leaves at microscopic level could be detected. Primarily they collected the image data, from Hill Agricultural Research and Extension Centre, Dhaulakuan, Himachal Pardesh, India.Gaussian filters were used to remove the unwanted particulars and clear the noises etc.

Through log transform techniques the enhancement of the images, was carried out so that contrast of lower intensity values is increased. Further, In order to get the images in clear contrast, a binary threshold segmentation method was used by authors. They enhanced the quality of the image for feature extraction by applying the output Guassian filters & Log transformation to discrete wavelet transform as input. As a final step of obtaining quality image data , The segmented image was used to produce wavelet features and it is stored for further processing . Now in order to categorize healthy and rusted leaves, they applied Support Vector Machine classification algorithm, and the authors claim that they have obtained an accuracy of 89.60%. The important findings of this paper are that, the approach is cost effective and simple, but has major limitation that the rust diseases with little variations/changes cannot be detected. Moreover carrying out image processing at microscopic level requires lot of intricate technical analysis.

The authors V. Singh and AK. Misra [5]. have proposed a method to categorize the diseased leaves of the Banana plant. They obtained 106 images of healthy and unhealthy banana leaves through high resolution camera. Further they divided this set into two sets namely training and testing data set. The authors have employed Genetic algorithms for processing the images, and perform image segmentation to extract the diseased part of the unhealthy leaves. As a final step these extracted features were applied to the Machine learning algorithms, Support Vector Machine , for classifying the diseased leaves and they claim that classification was performed with an accuracy of 95.7%. The major thing to be discussed here is that, the authors have used very small size of data set. Moreover the system consumes more time in preprocessing as it is genetic algorithm for it.

The authors M. Islam et al.[6]. have proposed a method for identifying diseased/infected potato leaves. Their work is based on the public dataset Plant village. The authors have picked up a very small data set from the plant village data set and divided into training dataset and testing dataset. The obtained the region of interest(ROI), through image segmentation by applying thresh-holder application in Matlab software. Further, the extracted texture and color features from the segmented images were stored in the feature vector and This is given to the Multiclass Support Vector Machine classifier which provided an accuracy of 95%. Here again the major drawback of this work is that, the data set used is of very meager size. So input with little variation from the training set would provide incorrect classification results.

The authors M. H. Ahmed et al.[7] proposed a method named GAACO (Genetic Algorithm & Ant Colony Optimization) with TSVM(Transductive Support Vector Machine) for detecting pomegranate leaf disease, thereby minimizing the drawbacks of present systems based on machine learning. Primarily they rectified the performance affecting aspects, for instance poor image processing, inconsistent results in classification. Further the method uses RGB image to the gray-scale image conversion, and developed the histogram of the input image. In the next step, the proposed method segmented the images by applying their above said proposed algorithm. Subsequently, the extracted texture features are stored in a feature vector. Lastly, helpful features were given as input to the Transductive Support Vector Machine TSVM classifier to classify pomegranate leaves with an accuracy of 91.3%. The major drawback of this work is that again it uses a very small dataset, due to which the model is not trained for different types of variations in plant diseases, and thus the results have poor accuracy for such cases.

The authors A. Iqbal and K. H. Talukder[8]. Have proposed system to detect two common leaf diseases of the potato plants namely Early Blight and Late Blight. Primarily, the dataset comprises of only 450 images. Through, image segmentation they cleaned all the images. Again the region of interest(ROI) of the image is extracted using the color threshold segmentation method and , and Histogram features were extracted. The interesting part of their work is that, the authors have used seven machine learning algorithms such as Logistic Regression (LR), Random Forest (RF), KNN, Decision Trees, Nai`ve Bayse , Linear Discriminant Analysis (LDA), and Support Vector Machine to perform the classification. The Random Forest Machine learning

classifier has the highest accuracy of 97% compared to rest of the six algorithms. The major limitation of this method is that again they are using very small data set, of 450 images, which is not the sufficient number for the training model to be robust and accurate. Moreover the preprocessing of images by the use of Histograms affects the rate of disease detection rate.

The authors P. R. Rothe and J.P. Rothe[9]. proposed a method that intends a scheme which can be used for detection of plant leaf diseases like Alternaria, Bacterial leaf blight, and Myrothecium especially for diseased leaves of cotton crop. They developed their own dataset of fom the fields of Nagpur. . For the region of interest (ROI), Otsu segmentation algorithm is used that basically extracts a diseased/unhealthy part/region from the preprocessed images. Further, nine color, four shape, and twenty-two texture features were derived from the segmented images and then inputted to the Feed Forward Back Propagation Neural Network (FFBPNN) classifier. Lastly, the classifier provides the classification results with an accuracy of 95.48%. The major shortcoming of this work is that, it has used hardly 305 images , due to which the proposed training model underperforms.

The authors G. Sun, X. Jia and T. Geng[10]have proposed a method to detect the plant leaves diseases of various corps using Multiple Linear Regression (MLR). Primarily, the proposed work pre-processes images, selecting image edges by edge detection algorithms to separate out the leaf from the background. As a next step the authors clean image & the noise using spatial domain techniques. The segmentation phase, uses an improved histogram method that calculates the threshold value automatically. Thus here ,segmentation method is used to separate out the unhealthy leaf region. Then, the authors extracted the texture, shape, and color features from the infected/diseased area of the leaf image. Lastly, this method uses all the extracted features and performs classification with an accuracy of 90%. Again like other proposed works, this method too uses very small size of data set.

The authors R. Meena Prakash, et al.[11] have proposed the GLCM (Gray Level Co-occurrence Matrix) algorithm to detect various diseases in the citrus leaf using texture features. Primarily, the authors performed the color space conversion to increase quality of the images. Further, the diseased part of the image was selected by the k-means clustering algorithm. Lastly , the four extracted texture features from the GLCM method were applied to the Support Vector Machine classifier and which produced an accuracy of 90%. Table.1 summarizes our work of review of various commonly used machine learning algorithms for plant leaf disease detection with key features and limitation of each machine algorithm.

TABLE I. SUMMARY OF VARIOUS COMMONLY USED MACHINE LEARNING ALGORITHMS FOR PLANT LEAF DISEASE DETECTION

S. no	ML algorithms	Key feature	Limitation
1	Support Vector Machine(SVM).	Simple, yet robust algorithm with good scalability when blended with specific algorithms.	The major drawback of this algorithm is that , the training process become complex with larger data sets. Moreover it is not compatible segmentation methods which are slower.
2	K-Nearest Neighbor(KNN).	It has lower computation overhead.	Deciding the value of K is difficult.
3	Na'ıve Bays(NB).	Simple classifier, with higher accuracy levels, and faster relative to other algorithms.	It cannot handle online data.
4	Decision Trees(DT).	Good accuracy levels.	Normally suffers with the over-fitting issues.

IV. DISCUSSION ON MACHINE LEARNING BASED LEAF DISEASE DETECTION TECHNIQUES

***Datasets*:** From the above review of existing work in plant leaf disease detection using machine learning algorithms has witnessed that, in all most all of the proposed works lesser number of inputs is used in their dataset. Normally the plant village dataset is the natural and obvious choice of researchers. The limitation of the plant village dataset is that the images were captured in a controlled environment. It is obvious that the trained model in the controlled environment is unfit to test the input which is collected in uncontrolled environment. This review work suggests researchers to use the real-time images.

Segmentation Methods: K-means is the most popular segmentation method that is used for the segmentation process. This method successfully detects the diseased portion of the leaf images, but, it is quite tedious and trick to predict the value of ' k'. Providing the value of k iteratively is a difficult for larger data sets during the training phase

To overcome prediction of K value during training phase, a improved image segmentation technique is used called, the Otsu segmentation algorithm , which works quite faster compared to the Active contour technique. Moreover finding the threshold value manually is again prone to errors and thus the likelihood of getting inaccurate results is probably high. Again to avoid this, an improved histogram techniques is used by researchers to detect the leaf 's diseased portion there by finding the threshold value automatically. It suffers from lower rates of detection of diseased part of the leaf image.
To overcome this issue of lower rate of detection of diseased potion of the image, a Genetic Algorithm is used by researchers during segmentation phase. The major advantage of this method is that it is fully automatic, detects the diseased portion accurately. As it requires more processing time, It makes the overall the processing slower. The shortcoming of the above method is minimized in another novel method of image segmentation called, Grabcut which enhances the classification accuracy significantly.

V. CONCLUSION

The review work carried out here clearly discussed the most commonly used machine learning algorithms vis-à-vis datasets used, segmentation methods and training models used. It is clear from the above discussion that machine learning algorithms are unable to process new data(inputs), i.e. they lack the ability of learning from new data. Moreover for accurate and precise training , larger datasets are required, for which the machine learning algorithms cannot be trained. It is evident from the above discussion that, most of the research work is invariably adopted a classifier as Support Vector Machine, for the reason being it is simpler to implement. Moreover, here it is witnessed that, most of the proposed work commonly uses plant village dataset for training and testing. In order to improve the accuracy, authors/researchers must work on real-time image dataset developed in a blend of controlled and uncontrolled environments. From this review work, it is clear that there is no sufficient work done to detect the pomegranate plant leaf disease detection. As India is agriculture rich and especially the southern part of India like Maharashtra and Karnataka cultivates pomegranate crop, there should be a good research on pomegranate plant leaf disease detection using the state of the art learning techniques using deep learning algorithms, which overcome the issues of machine learning algorithms.

REFERENCES

[1] E. Oerke, "Crop losses to pests," *The Journal of Agricultural Science"*, vol. 144, p. 31, 2006
[2] Hilman F. Pardede, Endang Suryawati and Dikdik Krisnandi. "Machine Learning Based Plant Diseases Detection: A Review", *IEEE International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications*, 2020.
[3] Majji V Applalanaidu, G. Kumaravelan. "A Review of Machine Learning Approaches in Plant Leaf Disease Detection and Classification", *IEEE International Conference on Intelligent Communication Technologies and Virtual Mobile Networks* 2021.
[4] K. Singh, S. Kumar and P. Kaur. "Support vector machine classifier based detection of fungal rust disease in Pea Plant (Pisamsativam)", *International Journal of Information Technology.* 2018
[5] V. Singh and AK. Misra. "Detection of plant leaf diseases using image segmentation and soft computing techniques", *Information Processing Agriculture*. Vol. 4, No.1,pp. 41–49, 2017.
[6] M. Islam, A. Dinh, K. Wahid, and P. Bhowmik. "Detection of potato diseases using image segmentation and multiclass support vector machine", *Electrical and Computer Engineering (CCECE), 2017 IEEE 30th Canadian Conference on. IEEE*. pp. 1–4, 2017
[7] M. H. Ahmed, T. Islam and R R. Ema. "A New Hybrid Intelligent GAACO Algorithm for Automatic Image Segmentation and Plant Leaf or Fruit Diseases Identification Using TSVM Classifier", *2nd IEEE International Conference Electronic Computer. Communication Engineering.* pp. 1–6, 2019.
[8] A. Iqbal and K. H. Talukder. "Detection of Potato Disease Using Image Segmentation and Machine Learning", *IEEE international conference*. pp. 43–47, 2020
[9] P. R. Rothe and J.P. Rothe. "Intelligent Pattern Recognition System with Application to Cotton Leaf Disease Identification", *Innovations in Computer Science and Engineering. Springer*. Pp. 19-27, 2019.
[10] G. Sun, X. Jia and T. Geng. "Plant Diseases Recognition Based on Image Processing Technology", *Journal of Electrical and Computer Engineering*. pp. 1–8, 2018
[11] R. Meena Prakash, GP. Saraswathy, G. Ramalakshmi, KH. Mangaleswari and T. Kaviya. "Detection of leaf diseases and classification using digital image processing", *Proceedings of 2017 IEEE International Conference Innovation Information Embedded Communication System*. pp. 1–4, 2018.