

Role of Machine Learning Approach for Detection and Classification of Diseases in Cotton Plant

¹Sandhya N. dhage, ²Dr. Vijay Kumar Garg

¹Research Scholar, Computer Science & Engineering, Lovely Professional University, Phagwara, India

²Associate Professor, School of Computer Science & Engineering, Lovely Professional University, Phagwara, India

¹dsand8@gmail.com, ²vijay.garg@lpu.co.in

Article History: Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

Abstract: Qualitative and quantitative agricultural production leads to economic benefits which can be achieved by periodic monitoring of crop, detection and prevention of crop diseases and insects. Quality of crop production is reduced by pest infection and crop diseases. Existing measures involves manual detection of cotton diseases by farmers and experts which requires regular monitoring and detection manifest at middle to later stage of infection which causes many disadvantages such as becoming too late for diseases to be cured. Lack of early detection of diseases causes the diseases to be spread in nearby crops in the field and also spraying of pesticides is done on entire field for minimizing the infection of disease. The main goal of proposed research topic is to find the solution to the agriculture problem which involves detecting disease in cotton plant at early stage and classify the disease based on symptoms. Early detection of disease at an early stage prevent it from spreading to another area and preventive measures can be taken by farmers by spraying pesticides to control its growth which helps to increase the cotton yield production. Automatic identification of the different diseases affecting cotton crop will give many benefits to the farmers so that time, money will be saved and also gives healthy life to the crop. The contribution of this paper is to present the machine learning approach used for cotton crop disease diagnosis and classification.

Keywords: Cotton plant, Crop disease, Machine learning, detection, classification.

I. INTRODUCTION

Agriculture field plays a key role in India. Crop diseases and Pest control is one of the major problem faced by farmers at nation and also at local level. Crop diseases are identified by farmers by naked eye observation which is very time consuming process and requires continuous monitoring of the farm. Sometimes incorrect identification of disease causes spraying of wrong pesticides by farmers on entire field causes economic and environmental loss. The diseases in crop are cause of problem of production loss and it produces economic fall in agricultural industries. Therefore early prediction of disease and its percentage of severity is considered as the major environmental and economic challenge for all the farmers in India.

Incorrect detection of disease by expert and farmer can result in incorrect application of pesticides which results in crop production losses for our farmers. Also it causes spraying of pesticides entirely on field on healthy cotton plant for minimizing the infection of disease. This unnecessary use of pesticides causing environmental problems because many other insects and birds can die because of eating such plants. Also this leads to reduction in the quality and productivity of the plants. Therefore, the detection of leaf disease of the plant is very essential in the initial stage and taking the corrective action at the beginning can prevent it from spreading to other parts of the field.

Machine learning is the method in which machine learns from past experiences and perform the task. Real world problems are solved using machine learning by building a learning model that is good and useful approximation to the data. The study of machine learning has been increased to explore learning capability of computer similar to the human brain. Learning process in machine learning model is divided into two steps as training and testing as shown in figure 1. In training process, learner or learning algorithm learn the features from input samples in training data and build the learning model. In the testing process, test or production data is tested by learning model using the execution engine to make the prediction. Learning model produces tagged data as output which gives the final classified result.

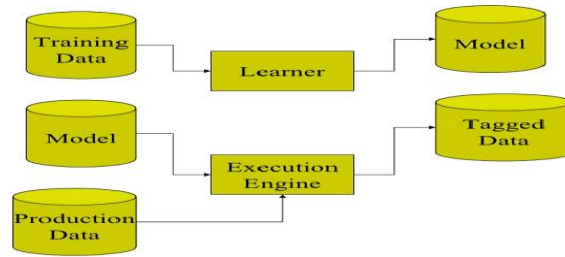


Fig.1 Machine Learning Model

Machine learning algorithms are categorized in areas of supervised and unsupervised learning based on the nature of learning signal. Supervised learning is performed on labeled training data which consist of input and output values. Supervised learning is further divided into classification and regression problem depending on discrete and continuous value of labeled data. Unsupervised learning is used to find unknown patterns from unlabeled data. It is further divided into clustering and association problem. Similar patterns in large data set are grouped together in clustering while association is used to find the association rules among the data objects in large data set.

Micro-organism like fungi and microorganism causes plant diseases whereas some micro-organism cannot be predicted at early stage by manual process by farmers. So, automatic identification of plant disease is important task in the research area of agriculture. Several attacks and damaged caused by diseases on cotton plant which is caused due to changing environment related to temperature, humidity, soil fertility. Bacterial, viral, and fungal disease are most common diseases in cotton plant. Different physical characteristics of the leaves are shown in diseased plant. Similar patterns on the leaves cause difficulty to detect changes in leaves at earlier stage so that earlier detection and recognition of disease becomes challenge in agriculture field and automatic identification of plant diseases is a most important analysis topic. In viral disease, virus enters the plant by a lesion which affects the natural growth of plant. Examples of viral diseases are mosaic, leaf curl, leaf roll, blue disease etc. In cotton leaf curl, infected cotton leaves curl upwards which is shown in figure 3. Figure 2 & 4 shows blue and mosaic disease respectively.



Fig.2 Blue Disease



Fig.3 Leaf Curl



Fig.4 Mosaic Disease

In bacterial disease,during all growth stages of cotton plant all sections such as roots, leaves, bracts, and bollards can be infected with bacteria. It causes seedling blight, leaf blot, stem blackarm and petioles, black vein, and rot in the boll. Examples of bacterial diseases are bacterial blight, crown gall which are shown in figure 5 & 6 respectively.



Fig.5 Crown Gall



Fig.6 Bacterial Blight

In fungal diseases, fungus occurs on entire plant. Black root rot, Boll rot, Fusarium Vilt, Verticillium wilt, Grey mildew are some examples of fungal diseases which are shown in figure 7,8 and 9.



Fig.7 Fusarium Vilt



Fig.8 Verticillium wilt



Fig.9 Grey mildew

II. LITERATURE SURVEY

This section describes study and work that have been already done by researchers in detection and classification of different crops for resolving challenges in agriculture automation. The table I describe comparisons of machine learning methods used for detection of disease detection and classification in various plants for resolving challenges in agriculture automation. Different machine learning methods such as random forest, decision tree, navie bayes, K-nearest neighbor, support vector machine have been used for detection and classification and also different deep learning methods have been developed by researchers in recent years for achieving accurate prediction and classification of crop diseases. From the comparative study of literature survey, it has been investigated that accuracy of CNN classifier is more than SVM. Moreover SVM gives high predictive accuracy among all other machine learning methods.

Reference Paper	Machine Learning Technique used	Dataset of plant images	Accuracy
[1]	Convolutional Neural Network	Dataset of 87,848 images 25 different plants in a set of 58 distinct classes of plant, disease combination	99.53%
[2]	Two-stage architecture of neural network	Dataset of 54,323 images of 14 different crops	93.67%
[15]	Convolutional Neural Network	Less number of affected leaf images of different plants	98%
[12]	Random Forest	160 images of Papaya leaves	70.14%
[13]	Support vector machine, Decision tree, Random Forest, Navie Bayes	3.823 images of corn plant	SVM has high accuracy among other methods
[4]	Support vector machine, K-nearest neighbour	Leaf images of cotton plant	More than 96% accuracy
[22]	Support vector machine, K-nearest neighbour	190 images of cotton plant	SVM has high accuracy among other methods
[21]	Support vector machine	145 images of rice plant	Not specified
[6]	K-nearest neighbour	237 leaf images of 5 types of diseases of different plants	96.76%
[7]	Convolutional Neural Network	Dataset of 500 images of rice plant	99.53%
[17]	Support vector machine	Images of sugarcane borer disease	96%
[18]	Support vector machine	900 images of five types of diseases of cotton plant	83.26%

Table I Comparative study of crop disease detection

III. ROLE OF MACHINE LEARNING IN CROP DISEASE DETECTION AND CLASSIFICATION

Machine learning techniques are used in variety of applications but it plays a key role in agriculture applications for early detection of and classification of diseases in crops. As compared to traditional methods used by farmers and experts, automatic detection of disease on crops is possible using machine learning which

gives accurate results. Spectroscopic techniques for disease detection are svery expensive and can only be used by qualified people. To resolve challenges related to crop diseases for improvement in yield production, machine learning techniques are used by researchers in recent era. Many researchers have studied and implemented different machine learning techniques and algorithms for disease identification on crops, fruits, and vegetables.

The system which uses machine learning is divided into two modules that is image processing and image classification shown in figure10. First module consist of image acquisition and image processing which requires capturing of images of healthy and diseased plants followed by preprocessing of image. Image preprocessing is required to remove noise and distortion in image for getting good quality image for further stages. Effect of such distortion can be removed using different noise removal filters. Image segmentation is needed to divide the image data into region of interest so that one can extract the useful features from the data required . The feature extraction focuses on identifying similar characteristics of features present within an image.. Features are extracted in three categories as color, shape, and texture. The color is an important feature because it can differentiate one disease from another. Furthermore, each disease may have different shape; thus system can differentiate diseases using shape features. Texture means how color patterns are scattered in the image. Classification is required to maps the data into specific groups or classes.. For crop disease, classification is performed for classifying the image into two labels such as healthy and diseased plants.

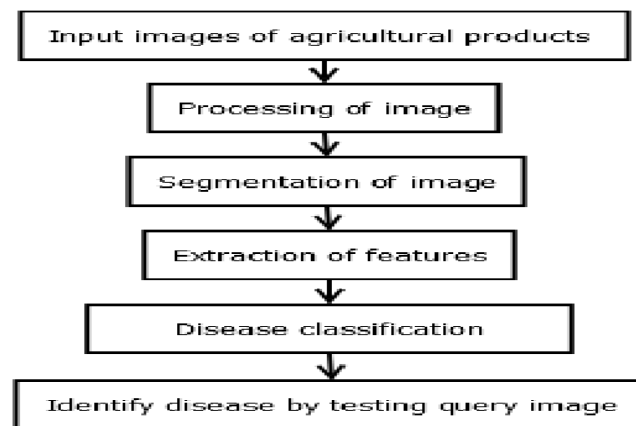


Fig. 10 Stages of Image detection and classification

IV. COMPARISONS OF MACHINE LEARNING TECHNIQUES

The review of literature has been studied for investigating the findings and research gap to implement the system for the early detection and classification of cotton diseases using machine learning techniques. Previous research study showed that less work is carried out using hybrid of two or more algorithm for disease detection. So future work can be carried out using hybrid combination of deep learning algorithms along with good feature extraction methods for improved accuracy of classification and to reduce time. Classification is done on small size of dataset in most of the earlier research work so that accurate detection and classification cannot be achieved for new infected images of plants. Large volume of dataset is required for classification to achieve highest accuracy.

Supervised and unsupervised techniques have been studied and comparison of algorithms is given in table II& III. In supervised learning, model is learned based on labeled data. Supervised learning is divided into classification and regression problem. Random Forest, Navie Bayes, support vector machine, decision trees, linear and logistic regression are supervised algorithms.

Algorithm	Type of Problem	Type of Feature value	Predictive accuracy	Training speed	Prediction speed	Data set needed
Decision Tree	Works for classification and Regression problem	Works on both continuous or discrete (categorical) feature values	Average	Fast	Fast	Require small number of training dataset
Navie Bayes	Works for classification problem	Works on binary or categorical feature values	Lower	Fast	Fast	It handles more features easily.
Support Vector	Works for classification	Works on discrete	good	Slow for large	Fast for linearly	Requires more features than

Machine	problem. Usually used for binary classification.	(categorical) feature values		dataset	separable problems	required for DT but do not works well for large number of features as like navie bayes.
Random Forest	Works for classification and Regression problem	Works on both continuous or discrete (categorical) feature values	It runs efficiently on large data bases.	Faster than decision tree	Prediction speed is faster than training speed because it saves forests for future use.	It can handle thousands of input variable
Linear Regression	Works for Regression problem	Works when dependent variable(output variable) is continuous.	Prediction is accurate when strong corelation exists between two or more variable.	Fast	Fast	Works well for small data set.
Ordinary Least Squares Regression	Works for Regression problem	Works when dependent variable is beyond the range means continuous.	Prediction accuracy is lower than logistic regression	Fast	Fast	Works well for small data set.
Logistic Regression	Works for binary classification problem	Works when dependent variable(output variable) is discrete or categorical and independent variable can be continuous or categorical	Prediction accuracy is more than ordinary least square regression	Fast	Fast	Works well for small data set.

Table II Comparative study of Unsupervised algorithms

Unsupervised technique is used to find the unknown patterns in dataset where data has no label. Learning model works on its own and discover the patterns in the data. Unsupervised learning is divided into clustering and association problem. Hierarchical clustering, K means clustering, K nearest neighbor, principal component analysis, independent component analysis, singular value decomposition are clustering algorithms whereas apriori algorithm is association algorithm.

K means clustering	It is used for finding fixed number of clusters in dataset	It is used for clustering problem	It partition the data into distinct non-overlapping regions.	It is simple to implement and uses large dataset	Uses Ecludian distance to calculate distance of each data point from center of centroid.
Apriori algorithm	It is technique to uncover how items are associated to each other.	It is used for association problem	Join and Prune steps are used to find frequent items in large dataset	It requires high computation when items are large.	Confidence, support measures are used to find association among items.

Principal Component analysis	It is used to find principal components based on covariance matrix	It is used for dimensionality reduction problem	It is used to reduce number of features while preserving the maximum variance.	Principal components are orthogonal.	If $d+1$ dimensions are given the mean, covariance matrix, eigenvalue, eigenvectors are used to find $d \times K$ dimensional matrix.
Singular Value decomposition	It is used to reduce large set of data to smaller subset of features based on covariance matrix.	It is used for dimensionality reduction problem	Largest variance occurred in the direction of first column of principal component.	The largest variance on subspace orthogonal to the first principal component is direction of second column.	covariance matrix, eigenvalue, eigenvectors are measures.
Independent Component Analysis	It is used to find independent components which are independent of each other.	It is used for dimensionality reduction problem	It does not focus on variance.	Independent components are not orthogonal.	Covariance matrix, eigenvalue, eigenvectors are used as measures.

Table III Comparative study of Unsupervised algorithms

From the literature review, comparative analysis is performed. The fig. 11 shows machine learning techniques used in referred literature for Crop disease detection and classification. Graph shows that utilization of SVM is more.

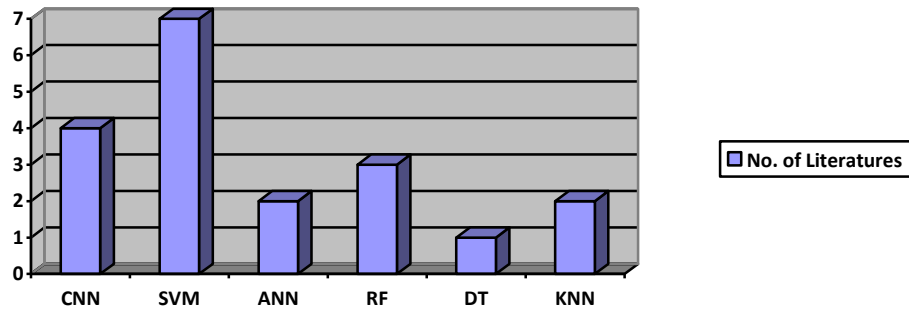


Fig.11 Graphical representation of ML techniques used

The fig. 12 shows type of crop used for disease detection and classification using machine learning. From the comparative analysis, it is found that machine learning techniques are implemented for various plants disease detection.

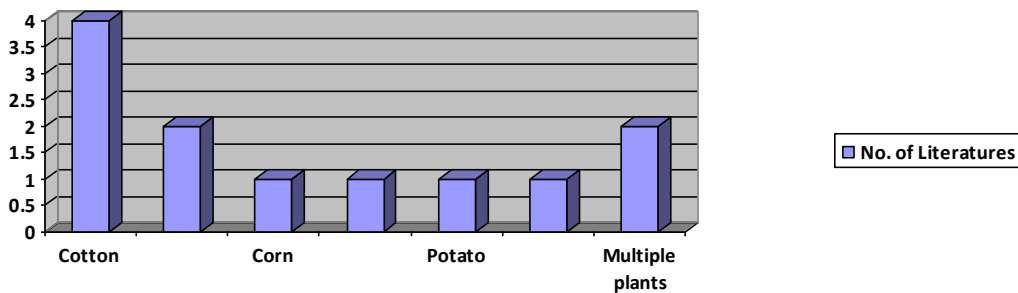


Fig. 12 Graphical representation of considered crop

The fig. 13 shows dataset used for detection of disease in multiple crops using machine learning in given reference paper. It has been found that large dataset is required for more accurate prediction of disease using CNN.

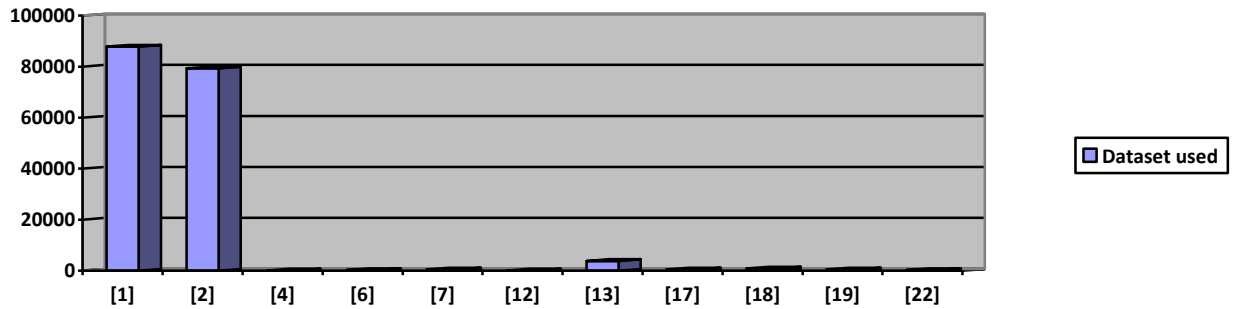


Fig. 13 Graphical representation of dataset used per reference

V. CONCLUSION

Crop diseases cause reduction in agriculture production which is one of the major problem faced by farmers. This paper summarized that different machine learning and deep learning algorithms are used for crop disease identification and classification to gain optimum agriculture production. Based on literature review, It has been found that cotton plant disease detection and classification have been carried out by using machine learning methods by less number of researchers so it requires more work to be performed in future for focusing mainly more number of cotton plant diseases. Hybrid combination of different machine learning algorithms can be applied in the system to achieve highest accuracy. Hence the paper proposed that hybrid machine learning techniques can be used for accurate prediction and classification of cotton crop diseases. Deep learning which is the advanced approach of machine learning can be implemented as its accuracy is higher than the basic machine learning algorithms.

REFERENCES

1. Konstantinos P. Ferentinos,” Deep learning models for plant disease detection and diagnosis,” *Computers and Electronics in Agriculture* 145 (2018) 311–318
2. Marko Arsenovic , Mirjana Karanovic, Srdjan Sladojevic, Andras Anderla , Darko Stefanovic,” Solving Current Limitations of Deep Learning Based
3. Approaches for Plant Disease Detection”, *Symmetry* 2019
4. Kirtan Jha, Aalap Doshi, Poojan Patel, Manan Shah, “Comprehensive review on automation in agriculture using artificial intelligence”, *Artificial Intelligence in Agriculture* 2 (2019) 1–12.
5. Kapil Prashar, Rajneesh Talwar, Chander Kant, “CNN based on Overlapping Pooling Method and Multi-layered Learning with SVM & KNN for
6. American Cotton Leaf Disease Recognition”, *International Conference on Automation, Computational and Technology Management (ICACTM)*, 978-1-5386-8010-0/19/2019 IEEE.
7. L. Sherly Puspha Annabel, T. Annapoorani and P. Deepalakshmi, “Machine Learning for Plant Leaf Disease Detection and Classification – A Review”,
8. *International Conference on Communication and Signal Processing*, April 4-6, 2019, India.
9. Eftekhari Hossain, Md. Farhad Hossain ,Mohammad Anis Rahaman, “A Color and Texture Based Approach for the Detection and Classification of Plant
10. Leaf Disease Using KNN Classifier”, *International Conference on Electrical, Computer and Communication International Conference on Electrical, Computer and Communication*, 978-1-5386-9111-3, 7-9, February, 2019.
11. V. Vanitha,” Rice Disease Detection Using Deep Learning”, *International Journal of Recent Technology and Engineering (IJRTE)*, 2277-3878, Volume- 7, Issue-5S3, February 2019.
12. Nikhil Shah, Sarika Jain, “Detection of Disease in Cotton Leaf using Artificial Neural Network” ,978-1-5386-9346-9,2019,IEEE.
13. Shruthi U, Dr. Nagaveni V, Dr. Raghavendra B K, “A Review on Machine Learning Classification Techniques for Plant Disease Detection”, *5th International Conference on Advanced Computing & Communication Systems*, 978-1-5386-9533-3,2019.
14. Sona Pawara, Dnyanesh Nawale, Kunal Patil, Rakesh Mahajan, “Early Detection of Pomegranate Disease Using Machine Learning and Internet of Things”, *3rd International Conference for Convergence in Technology (I2CT)*, Pune, India. Apr 06-08, 2018.

15. M. Vengateshwaran, E.V.R.M. Kalaimani,, “Deep Learner Based Earlier Plant Leaf Disease Prediction and Classification Using Machine Learning Algorithms”, IOSR Journal of Engineering (IOSRJEN),2250- 3021, 45-51,2018.
16. Shima Ramesh, Ramachandra Hebbar, Niveditha M, Pooja R, Prasad, Bhat N, Shashank N, Mr. P V Vinod, “Plant Disease Detection Using machine Learning”, International Conference on Design Innovations for 3Cs Compute Communicate Control, 978-1-5386-7523-6/18/2018 IEEE.
17. Budiarianto Suryo Kusumo, Ana Heryana, Oka Mahendra, Hilman F. Pardede, “Machine Learning-based for Automatic Detection of Corn-Plant Diseases Using Image Processing”, International Conference on Computer, Control, Informatics and its Applications , 978-1-5386-5741-6/18, 2018 IEEE.
19. Endang Suryawati, Rika Sustika, R. Sandra Yuwana, Agus Subekti , Hilman F. Pardede, “Deep Structured Convolutional Neural Network for Tomato Diseases Detection, ICACSYS 2018,IEEE,978-1-7281- 0135-4,2018.
20. A. Blessyl, Dr. D.C. Joy ,Winnie Wise, “Detection of Affected Part of Plant Leaves and Classification of Diseases Using CNN Technique”, International Journal of Engineering and Techniques - Volume 4 Issue 2,Mar-Apr 2018.
21. Na Wu, Miao Li, Lei Chen, Yuan Yuan, Shide Song, “A LDA-based segmentation model for classifying pixels in crop diseased images”, Proceedings of the 36th Chinese Control Conference, July 26-28, 2017, Dalian, China.
22. Tisen Huang, Rui Yang, Wenshan Huang, Yiqi Huang, Xi Qiao, “Detecting sugarcane borer diseases using support vector machine”, Information Processing in Agriculture, 74–82,2017
23. Adhao Asmita Sarangdhar , Prof. Dr. V. R. Pawar, “Machine Learning Regression Technique for Cotton Leaf Disease Detection and Controlling using IoT”, International Conference on Electronics, Communication and Aerospace Technology, 978-1-5090-5686-6,2017.
24. Priyadarshini Patil, Nagaratna Yaligar, Meena S M, “Comparision of Performance of Classifiers - SVM, RF and ANN in Potato Blight Disease Detection using Leaf Images”, IEEE International Conference on Computational Intelligence and Computing Research, 978-1-5090- 6621-6,2017.
25. Thomas Truong, Anh Dinh, Khan Wahid, “An IoT Environmental Data Collection System for Fungal Detection in Crop Fields”, IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), 978-1-5090-5538-8,2017.
26. Jitesh P. Shah, Harshadkumar B. Prajapati, Vipul K. Dabhi, “A Survey on Detection and Classification of Rice Plant Diseases”, 978-1- 090-1936-6/16/2016 IEEE.
27. Bhumika S.Prajapati, Vipul K.Dabhi, Harshadkumar B.Prajapati, “Asurvey on Detection and Classification of Cotton Leaf Diseases”, International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 978-1-4673-9939-5/16/2016 IEEE.