

Data Mining Techniques for Identification and Classification of Various Diseases in Plants

Arun kumar Nakatha, Sathish S. Kumar

Abstract: Data mining is currently being used in various applications; In research community it plays a vital role. This paper specify about data mining techniques for the preprocessing and classification of various disease in plants. Since various plants has different diseases based on that each of them has different data sets and different objectives for knowledge discovery. Data Mining Techniques applied on plants that it helps in segmentation and classification of diseased plants, it avoids Oral Inspection and helps to increase in crop productivity. This paper provides various classification techniques Such as K-Nearest Neighbors, Support Vector Machine, Principle component Analysis, Neural Network. Thus among various techniques neural network is effective for disease detection in plants.

Keywords: Classification, Data Mining, K-Nearest Neighbors, Neural Network, Preprocessing, Principle Component Analysis, Segmentation, Support Vector Machine.

I. INTRODUCTION

Agriculture is the one of the source of Indian economy and agriculture is composed of many crops, employs 60 percent of labor force. Due to plant disease 18% of crop worldwide is lost. To enhance the quality and productivity of the crop and also to reduce the cost, some mechanism need to be proposed to former so that it will help them to detect the disease in early stage. Inspecting every plant and identifying the disease is tedious, Applying data mining techniques to detect the disease plays an important role. As applying these techniques reduce the manual intervention and also detection of disease at early stage reduces usage of pesticide .Thus the crop productivity and quality is increased with help of segmentation and classification techniques .Manual inspection and identification of disease is less accurate and time consuming. Hence agriculture being one of the important source of income for farmers applying new trends and technology will add on to agriculture development

Revised Manuscript Received on December 05, 2019.

* Correspondence Author

Arun Kumar Nakatha, Research Scholar , Computer Science and Engineering, RNS Institute of Technology, Bengaluru, India.

Email: arunnakhate@gmail.com

Dr. Sathish S. Kumar, Professor, Information Science and Engineering, R N S Institute of Technology, Bengaluru, India.,

Email: sathish_tri@yahoo.com

II. RELATED WORK

The automatic investigation of plant disease considered using the machine learning approach [1]. This paper has made an attempt to study machine learning method used by researchers to identify disease and classification based on image .A comparative study [2] is carried out on five types of machine learning classification techniques for recognition of plant disease is done .Techniques such as SVM, CNN applied to detect the disease. CNN has shown high accuracy in detection of disease. Other classification like decision tree and naive based can be used for automatic detection of plant disease. This paper has proposed a real-time detection approach [3] that is based on improved convolution neural networks for apple leaf diseases, the comprehensive detection performance reaches 78.80% MAP. Meanwhile, the detection speed of the model reaches 23.13 FPS. The Proposed system [4] with the help of parameters temperature, moisture, etc. predicts cotton crop diseases using decision based on previous data. Umair Ayub et.al [5] stated analysis of different data mining classifiers on different feature sets to predict the grass grub damages .The classifiers used are Random Tree, Random forest ,Decision Tree ,Support vector machine ,Neural Network Naïve Bayes and K-Nearest Neighbors combination of Decision Tree ,Random Forest and Support Vector Machine has proven as best combination out of all testes combinations and specified deep learning and hybrid approaches can solve the crops related problem.The analysis [6] were made on crop disease by applying random forest model and decision Tree Model ,the best result is by applying the random forest model. The proposed methodology in this paper [7] depends on CNN and neural network techniques which are configured for leaf disease detection. Use of neural network saves a considerable amount of effort. Threshold and neural network algorithms provides high accuracy and consumes very less time for entire processing. The conclusion made in this paper [8] that Artificial neural network is best classifiers for the system with highest accuracy of 92% due to its ability of classification of leaf images. By applying the ANN on the data set Blight Disease in potato was diagnosed at early stage. Features are extracted from the disease affected region used for classification by taking texture into consideration. The Artificial neural network has advantage that it is flexible without imposing any restrictions on input variables to model unseen hidden relationship between variables.

Author has used multi SVM classifier, K-Means clustering in the paper [9] ,Otsu's method gives the high accuracy and save computational time.

Table I: shows plant diseases detection and classification accuracy of different data mining technique.

Index of Citation	Methodology	Limitations	Result
Yingfeng et.al[22],2010	Back Propagation NN	Misclassification rate is 5.9 and modified BPNN accuracy rate less compare to standard BPNN	98%
R.SwarnaLakshmi et.al[28],2014	Fuzzy logic neural network based classifiers	Segmentation and quality determination rate is less	90.30%
Pankaj kumar Sharma et al[26],2015	SVM and Random forest	Classification techniques need to be applied more effectively for detection.	Improvement is increased by 4 %
Hemanth Kumar et.al[27],2017	Naïve bayes Kernel algorithm	The sensors used to detect the plant disease in the field and accuracy rate can be increased	90%
Hyeon park et.al[23],2017	Image based deep learning	The input images fed to system is less.	89.70%
Melike sardogam et.al[25],2018	CNN with LVQ	To improve recognition rate in classification process different filters or different size of convolution can also be used.	Out of 20 crop images for 18 images correctly classified
Hyeon park et.al[24],2018	Image based deep learning	Accuracy of detection is less as dataset consists of less images.	92%
Om Tiwari[21],2019	CNN	It takes a long time to calculate the gradient of the model using the larger dataset. The pre-trained models used, training data not required and same technique can be applied to detect disease.	98%
Soumya Ranjan et.al[9],2019	Multi SVM	Computational time is more.	-
Darshana A et.al [20],2015	Region based segmentation	Focus is only on growing segmentation.	Region growing gave the maximum peaks representing distinct regions having least discrete entropy and highest grey level energy as compared to mean shift segmentation methods.
Sachin D .Khirade et. [10],2015	K Means Clustering Multi SVM Classifier ANN method	The feature extraction and classifying the disease can be improved.	Accurate detection and identification of diseases with less time consuming
santanu Phadikar et al[11],2008	Segmentation boundary detection Zooming algorithm SOM neural Network	Image transformation does not give better classification compare to original image.	Satisfactory classification for test images
Ramakrishna et.al [12],2015	HSV conversion	Work can be carried for more than four disease detection.	Classified four different disease with 97% accuracy

III. METHODOLOGY

Data mining technique is the process of identifying and discovering interesting patterns and knowledge from large of data. The information is retrieved in the form of knowledge and a repository is maintained. As the data from various view point is analyzed. The disease hidden in the crop is identified and extracted from agricultural data and acts as an important role in crop disease detection. The methodology applied is data mining architectural flow and the techniques incubated are advance techniques in order to extract the necessary information.

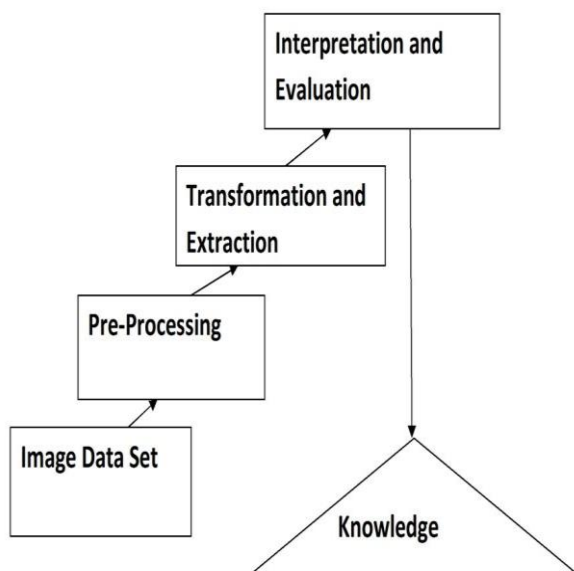


Fig 1. Architecture of Data Mining

The architecture work flow includes five phases:

Phase 1: Image Database- The basic functioning of a Data Mining system starts stage by stage. a database of the images from the relevant sources are collected and stored inside a database.

Phase 2: Preprocessing- The images acquired from these sources are rarely as we desire them. These images need a lot of preprocessing before they become useful for Image Mining applications. Image Pre-Processing techniques are thus applied to the images stored in the database. Pre-processing is an improvement of the image data that suppresses unwanted distortions or enhances some image feature important for processing. Various Preprocessing techniques commonly used are Histogram Equalization, Addition of Filters, Noise Removal, Thresholding, Contrast and Color correction etc.

Phase 3: Transformation and Extraction- Samples within a dataset identified for a particular purpose is known as the Region of Interest (ROI). Extraction of the ROI from the sample data set is the most important step in the data mining system. The preprocessed Images in the database are scanned and transformed. In The Image Transformation, the image are transformed or consolidate into forms appropriate for mining Transformation of images into database like table. In the table each row stands for a pixel. Thus the cardinality corresponds to the total number of pixels in an image. These features can be local variation, grey level, entropy, contrast,

mean, etc. and then finally the recognized patterns are extracted after innumerable iterations.

Phase 4: Interpretation and evaluation: Patterns evaluated depending on the constraints to extract valuable knowledge.

Phase 5: Knowledge: Using knowledge in the decision making of real time applications example medicines, health.

A. Approaches

Support Vector Machine: It is also called as support vector networks. The support vector machine consists of models related to supervised learning with associated learning algorithms which is meant for analysis of data used for classification and regression. SVM performs both Linear and non-linear classification efficiently.

Fuzzy Logic :It is an Approach of Data Mining that involves the data computation which is dependent on probable predictions and clustering as opposed to the traditional "true or false". Fuzzy logic consists of many algorithm, the algorithm that use fuzzy logic are increasingly being applied in several disciplines to help in mining of database.

K-means: It is the vector quantization method, is meant for cluster analysis and partition of cluster takes place. K means aim is to observe the data into K clusters by partitioning. There are various algorithms, the most common in algorithm is the refinement technique is applied iteratively.

Association Rules: The interesting relation between variables in large database are discovered and the Association rule mining is also known as rule based learning. The association rule applied on various type of data.

Neural network: It consists of simple processing units called a Neurons Hence neural network are organized into several layers with fully or partial connections. The activation value received by neuron from the neighbor. The output computed based on the neighbor input parameters and send that output to its neighbor.

IV. RESULTS AND DISCUSSIONS

The advantage of Convolutional neural network are motivated by the fact that they can capture and are able to learn relevant features from an images at different levels similar to human vision The steps involved in Convolutional neural network are as follows

- Input image
- Different filters are applied
- Function are applied are filtering
- Pooling image is applied to Feature Map
- Made into single long vector by flattening the pooled images
- Inputs the vector into fully connected artificial neural network
- Processes the feature
- Train through propagation techniques

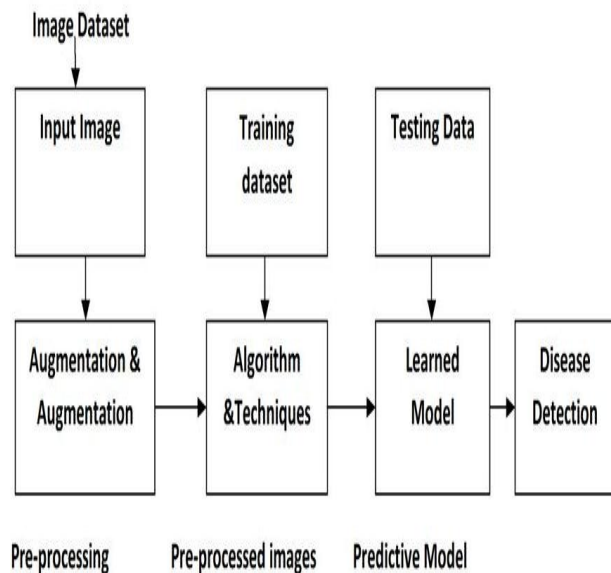


Fig 2. Overview of the System Design

The overview of the system consists of the following aspects:

- Collection of images from different sources and maintaining a dataset, the images are given as input for the system.
- As a part of pre-processing, Annotation and augmentation of data takes place.
- Once pre-processing is done, the algorithms and techniques are applied to train the dataset.
- A predictive model or learned model is applied on the test data.
- Image Analysis takes place to test and validate the images to detect the diseased leaf .Finally results are concluded as an output of system.

IV. CONCLUSION

The results from various papers are compared and concluded that automatic plant disease identification process could offer a great help to improve production process. To overcome the issues in pre-processing, transformation and extraction by applying the advanced techniques and to improve the rate of accuracy by considering large set of images .The Convolution neural Network approach has got the advantage as it takes images as input and the preprocessing required is less as compared to other classification algorithms. The image is processed in such way that the features in image is not lost and is scalable to massive .Hence Convolutional neural network approach can be applied for the identification and classification of diseases in plants.

REFERENCES

1. JyotiShirahatti,Rutujapatil,pooja Akulwar,"A Survey paper on plant Disease identification using Machine learning approach",Proceedings of the International Conference on communication and Electronics Systems (ICCES 2018)IEEE Xplore Part Number :CFPI18AWO-ART;ISBN:978-1-5386-4765-3.
2. Shruthi,Dr.NagaveniDrRaghavendra B KA "Review on Machine Learning Classification Techniques For Plant Disease Detection",2019 5th International Conference on Advanced Computing & Communication System (ICACCS),978-1-5386-9533-3.

3. Peng Jiang, Yuehan Chen, Bin Liu, Dongjian He, And Chunquan, "Real-Time Detection of Apple Leaf Diseases Using Deep Learning ApproachBasedonImprovedConvolutionalNeuralNetwork"10.1109/A CCESS.2019.2914929.
4. Jayraj,SagarNakum,VivekNakrani,HiralRaveshiya,"Cotton Crop Disease Detection using Decision Tree Classifier",10.1109/ICSCET.2018.8537336,IEEE.
5. UmairAyub,SyedAtif ,"Predicting Crop Diseases using data Mining Approaches and classification",978-1-5386-5482,IEEE 2018
6. R.Beulah ,M.punithavalli ,"Prediction of sugarcane Disease using data Mining Techniques" ,2016 IEEE International Conference on Advances in computer Applications (ICACA). 978-1-5090-3770-4, IEEE
7. Er. Saiqa Khan,DarakshanParveen, Khan SaameenPatel Jannat,"Analysis Of Data Mining Techniques For Agricultural Science"10.1109/ICSCET.2018.8537322,IEEE
8. Priyadarshinipatil ,NagaratnaYaligar,"Comparision of performance of classifiers SVM,RF and ANN in Potato Blight Disease detection using Leaf Images" .978-1-5090-6621 2017,IEEE
9. SoumyaRanjansahu ,chandrasekhar panda ,"A Review on leaves disease segmentation and classification Techniques ".,2019 IJSRT ,Volume 6,Issue 3 ,10.32628/IJSRT1196333.
10. Sachin D. Khirade and A.B. Patil "Plant DiseaseDetection Using Image Processing",2015 International Conference On Computing Communication Control And Automation, IEEE, pp. 978-1-4799-6892.
11. SantanuPhadikar And Jaya Sil "Rice Disease Identification Using Pattern Recognition Techniques" 11th International Conference On Computer And Information Technology(ICCIT2008), IEEE, PP. 1-4244-2136-7/08.
12. Ramakrishnan.M AndSahayaAnselinNisha.A "Groundnut Leaf Disease Detection And Classification By Using Back Propagation Algorithm" IEEE ICCSP 2015, conference, pp.978-1-4 799-8081-9/15.
13. H.Al-Hiary,S.Bani-Ahmad,M.Reyalat,M.Braik&Z.AIRahamneh, "Fast and Accurate Detection and Classification of Plant Diseases", International Journal of Computer Applications, Vol. 17, No.1, pp. 31-38.March2011.
14. Prakash M.Manikar, ShreekantGhorpade and MayurAdawadkar "Plant Leaf Disease Detection and Classification Using Image Processing Techniques" International Journal of Innovative and Emerging Research in Engineering (IJERE), volume 2, issue 4, 2015, e-ISSN :2394-3343, p-ISSN: 2394-5494.
15. Pallavi S. Marathe, G.H. Raisoni, SavitriPhule, "Plant Disease Detection Using Digital Image Processing And GSM", IJESC, volume 7, issue 4, April 2017.
16. R. G. Mundada, Dr. V.V. Gohokar, "Detection And Classification Of Pets In Green House Using Image Processing" IOSR Journal Of Electronics And Communication Engineering (ISOR-JECE) volume 5, issue 6, pp. 57-63, 2013.
17. S. Ananthi& S.V. Varthini, "Detection and Classification of Plant Leaf Diseases",International Journal of Research in Engineering & Applied Sciences, Volume 2, Issue 2, pp.763-773, February 2012.
18. W. Abudullakasm& J.Unartngam, "Assessment of the Severity of Brown Leaf Spot Disease in Cassava using Image Analysis", The International conference of the Thai Society of Agricultural Engineering, 2012.
19. Gudino, J. Gudino-Bazaldua, J. L. Rojas- Renterai, V.Rodriguez-Hernandez and V.M. Castano, "Color Image Segmentation Using Perceptual Spaces Through Applets For Determining And Preventing Diseases In Chili Peppers", African Journal of Biotechnology Volume 12, no.7, pp. 679688, 2013.
20. Darshana A., Jharna Majumdar, Shilpa Ankalaki, "Segmentation Method For Automatic Leaf Disease Detection", IJRCCE, Vol. 3, Issue 7, pp.1-7,2015
21. OmTiwari,Vittal goyal, Pramod Kumar,Sonakshi vij,"An experimental set up for utilizing convolutional neural network in automated weed detection ".978-7281-1253-4/2019,IEEE
22. Yingfeng Zhou,Yaming Wang ,Qing Yao,"Segmentation of Rice Disease Spots Based Improved",978-1-4244-5555-3/2010-IEEE.
23. Hyeon Park,Jee-Sook Eun and Se-Han Kim,"Image-based disease diagnosing and predicting of the crops through the deep

- learningmechanism”,978-1-5090-4032/2017 IEEE.
24. Hyeon Park,Jee-Sook Eun and Se-Han Kim” Crops disease diagnosing using Image-based Deep learning Mechanism”.978-1-5386-5928-1/2017 IEEE.
 25. Melike Sardogan ,Adem Tuncer,Yunus Ozen ,”Plant disease Detection and classification Based on CNN with LVQ Algorithm”,978-15386-7893/18/IEEE.
 26. Pankaj Kumar Sharma,Hammad haleem and Tanvir Ahmad “Improving Classificatin by Outlier detection and Removal”,Springer International publishing switzerland :10,1007/978-3-319-13731-5_67
 27. Hemanth Kumar ,Nilima Ashantkar,”An Appropriate Model Predicting Pest/Diseases of crops Using Machine Learning Algorithms”,2017 International Conferenc on Advanced Computing and Communication Systems,978-1-5090-4559-4/IEEE.
 28. R.SwarnaLakshmi ,B.Kanchanaevi,”A Review on fruit Grading systems for Quality Inspection” .IJCSMC,Vol 3,Issue July 2014 pg 615-621.

AUTHORS PROFILE



Arun Kumar Nakatha, Research scholar R N S institute of Technology , Bengaluru , Karnataka, India, He has done BE degree from S L N college of Engineering Raichur ,Karnataka, MTech Degree From Cambridge Institute of Technology, Bengaluru , Karnataka. He is lifetime member of ISTE. His field of Interest includes Data Mining, Big Data, Computer Vision, and networks.



Dr. S. Sathish Kumar has done B.E degree from Madurai Kamaraj University ,Tamilnadu, India,M.E degree from Regional Engineering college ,Tiruchirappali ,Bharathidasan University ,Tamilnadu, India and Doctor of philosophy Degree from Dr. M. G.R Universirty ,Chennai ,Tamilnadu ,India. Currently he is working as a professor in R N S institute of Technology ,Bengaluru ,Karnataka ,India .he is having more than 25 years of experience .He has served as a BOE/BOS member of various Universities and colleges .He has published research Papers in various Universities and colleges .He has published research papers in various journals and conferences .His field of Interest includes Data Mining ,Big Data, Bio Medical Engineering and cloud computing.