A Wide Survey on Data Mining Approach for Crop Diseases Detection and Prevention

Vikas N Nirgude [a,1], Dr.Sandeep Malik [a]
aDepartment of CSE, Oriental University, Indore, India

Abstract. India is agriculture land and major revenue manufacturing sector. However, because of amendment in temporal parameters and uncertainty in climate directly have an effect on quality and amount of the assembly and maintenance of crops. Also, quality even a lot of degrade once the crops area unit infected by any malady. The main focus of this analysis in agriculture is to increment the crop quality and potency at lower price and gain profit as result of in India the majority of the population depends on agriculture. Big selection of fruits is growing up in India such as apple, banana, guava, grape, mango, pomegranate, orange is the main one. Fruit production gives around 20% of the country's development. However, because of absence of maintenance, inappropriate development of fruits and manual investigation there has been scale back in generate the standard of fruits.So, Data Mining Approach used in the agriculture domain to resolve several agricultural issues of classification or prediction. During this paper complete survey of several data mining approach for crop disease management has been done. Detection of disease in early state will improve in quality of crop still as decrease the production cost. Also, we can improve the production of the particular crop. Several major parameters are used for the crop disease classification or prediction.

Keywords: Data Mining, Support Vector Machine, Classification, Naïve Bayes Machine

1. Introduction

India is agriculture country; whereas major population depends on it. Indian farmers have a choice to decide their crops. Our farmers work hard to feed large population. But due to drastic change in climate condition crop failure will occur which leads more farmers are committed suicides. So, by using different data mining approaches, can predict the different diseases which occur on early stage so that we can increase the crop quality and quantity. In agriculture fields Vegetables and Fruits are most significant products. To increment the crop quality and potency at lower price and gain more profit as result of in India the majority of the population depends on agriculture due to this main focus is given on this analysis. Wide range of fruits is growing up in India such as apple, banana, guava, grape, mango, pomegranate, orange is the main one. Fruit production gives around 20% of the country's development. However, because of absence of maintenance, inappropriate development of fruits and manual investigation there has been scale back in generate the standard of fruits [1].

---

[1] Vikas N Nirgude, Ph.D Scholor, Department of CSE, Oriental University, Indore, India
Email: vikas.nirgude2006@gmail.com.

In Agriculture quality of seed and soil is important for producing good quality of product. The quality of agriculture product is reduced if crop is infected by some malady. The operation like crosspollination transpiration, fertilization, and germination etc., will affects the plant which may cause the disease. The pathogen like bacteria, fungi, viruses and also because of unfriendly conditions will produce the diseases on the crops [2]. In plants different parts are like fruits, leaves, stem, flowers on which disorder can be found. So, without knowing the proper disease unnecessary pesticides are sprays and used which cause huge amount of air and soil pollution, that also effect on human health. So, it is necessary to predict the right disease depend on the atmospheric condition. Those natural details could be particular climate condition like humidity, temperature, leaf wetness and rainfall. Sometimes experience is insufficient. For specific prediction of potentially fruit plant disease, and right time for protection the fruit plant using data mining. To extract meaning information from huge database the Data mining is the technique used. Data mining is rising exploration field in agriculture plant protection too. Compare the available data with historical datasets as well as previous results of pathogen disease has been recorded [3].

Some factors of soil and environmental parameters like temperature, humidity, moisture etc. are affecting the quality and quantity of fruits. Just by observing the all-environmental parameters manually unnecessary pesticides are used by farmers sometimes which are actually not needed which will increase the production cost and degrade the quality of the fruit. So, there is need of new approach which detect the actual disease of fruit, which lead to improvement in quality and quantity.

2. Literature Survey on Crop Disease Management

Till dates several research works carried by different researchers such as, Dnyanesh Nawale et al [4] proposed machine learning framework with the help of IOT to early detect pomegranate disease, the framework uses Hidden Markov model to uncover disease and alert farmer initial phase. The framework also uses sensors to keep track of temporal details. The above framework required more time for testing and giving the result.
B. Balaji Bhanu et al [5] the temporal details have been taken trough different sensors for improving crop production wireless sensor networks system was proposed. S. R. Rajeswari et al [6] uses random forest method with the help of temporal details crop yield prediction and cost of product will be able to decide. Basically, this approach work for Maize and Rice type crop. Proposed system work has shown on static data set and basically it works on small area.
Geraldin B. Dela Cruz et al [7] uses principal component analysis (PCA) and genetic algorithm (GA) for crop classification. Basically, proposed framework uses crop images of maize. The proposed framework shows the improvement in classification further we can improve the classifying result by applying a novel data mining method.
X.E. Pantazi et al [8] proposed ANN model for detecting wheat yield prediction within field variation. This system uses soil parameters for checking the crop growth, also the factor which affecting the yields has been discussed. The temporal parameters which affect the quality of crop well as yield production is not considered.
Sunil More et al [9] suggested a framework for farmers about pesticides and nutrients that can be used, also it uses wireless sensor network for measuring the different

temporal details. Proposed system shows the better result as compared to traditional methods.

Manisha Bhange et al [10] proposed a framework by using web-based tool to detect fruit disease The proposed system having already processing the images of different diseases based on that image it gives the results. The proposed framework takes the online image as input further it extract the features and compare it with available images. The proposed framework used SVM classifier to classify the disease and according to that it will give the result. As the proposed system will work based on available database images it will not give accurate result for newly introduced disease.

Yuguang Huang et al [11] proposed a Naive Bayes Classification method which work on small data set. Proposed method is most effective in text classification. The proposed framework work well for large training sample set. For large training sample set put heavy work for classification and more storage is required.

Dhiman Mondal et al [12] proposed a framework by combining Image processing and Naive Bayesian classifier.

The proposed method is useful for detect as well as classify the present disease in Okra leaf of YVMV disease. The proposed system only focuses on okra leaf and limited disease prediction and classification.

G. Prem Rishi Kranth et al [13] proposed different machine learning framework for analysis of plant disease prediction. By using the feature of crop such as size, wilting, dryness, shape the disease of the crop is predicted. The above framework further extended to large area and large image dataset. There are several factors are there which will leads to different disease. Also, there is chance in improvement for shadow images because above framework will not give good result for shadow images.

Khumukcham Robindro et al [14] proposed Naive Bayes machine learning framework for detecting the diseases of rice plant. The proposed framework is a future for the detecting common diseases in rice plant may occur during entire life of rice plant. The framework will give the result by doing the assumptions of symptoms and precondition of diseases. Sometimes leads to prediction of wrong disease will affect the quality and also increase the cost productivity.

Chithra, P. et al [15] proposed K-means framework for detecting the defected apple, the proposed framework uses the images for analysis and extract the different features like colour.This system does not take any temporal details which will affect the quality of fruits.

Varughese et al (2016) [16] proposed K-mean framework to detect the apple disease. This framework used the segmentation approach to detect the affected area. The result can be further improved by using Artificial neural network.

Dubey, S et al [17] design a K-mean system for detection and classifying the disease of apple fruit. In this approach mainly focus given on scab, blotch and rot types of diseases. This framework work on images but disease may occur due to temporal details. So, for achieving better accuracy temporal details will be considered.

Mitunkumar Balar et al [18] proposed neural network framework for crop prediction. In this framework rain water and soil parameters such as PH, nitrogen, soil temperatures are considered. In this research focus given on cotton crop. It works on only static data sets and temporal details are not considered while predicting the production

Siddique et al [19] designed a framework by using multiple linear regression and k-nearest neighbour regression. This framework proposed for detecting which crop is

cost effective as well as give maximum yield for this soil parameters are considered. The result is calculated for small area and limited data sets.

3. Challenges in Recent Technology

CR-I. Which method of data mining is more suitable for crop disease management?
CR-II. Which are the major challenges that are faced during crop disease prediction as well as detection?

CR-I: Crop Disease Management by using different Data Mining approaches

Here we discuss various Data Mining approaches based on SVM, Naïve Byes that are used for crop disease
a. Identification
b. Classification
c. Prediction.

1. SVM approach for crop disease prediction:

Support vector machines are supervised learning models that analyse data for regression analysis and classification. They come with associated learning algorithms. A Support Vector Machine classifies data by determining the hyperplane that maximises the difference between two classes. It is used in machine learning applications.

A support vector machine used for classification and give better result if the problem is two class problem but in crop having more number of diseases are there then it not able to classify that problem properly.

2. Naïve Bayes approach for crop disease detection:

Naive Bayes assumes that all features which we will going to considered are independent, but in real life or for disease prediction or detection most of the disease features are dependent on each other.so again there is challenge if factors are more then it will not give accurate result of disease.

Naïve Bayes will give equal importance to each feature but in case of disease detection or prediction each factor having different values.

CR2-II: Crop disease prediction and identification are difficult tasks.

- Diseases are appearing on crops due to a variety of aspects like soil, temporal details, rainfall, seeds, and weeds because of this take decision of automatic real-time disease management is difficult. Crop diseases and quality are influenced by these factors, also prediction of correct disease is difficult.
- The majority of researchers have only looked at single parameters such as temperature, water, soil factors, Here the agricultural decision-making process takes into account all of them. As a result, by taking into account all of these considerations, crop disease accuracy can be enhanced.
- Due to the dynamic crop production method, real-time field data collection based on these parameters is critical.

- Most of the researchers considered only detected or predicted one or two diseases. But if more than two diseases are there then how to perform correct classification of these diseases.
- Very less research has been carried out on Indian soil therefore in India farmers are not get expected quality and quantity of the crop.
- In India because of the less research on the disease so farmers are not able to produce good quality of the crop even though they are using more pesticide unnecessary cost will get increase as not understanding the proper disease.

4. Analysis of Different Research Paper

Literature review was conducted for different paper which is shown in following Table 1. Based on a review of these research papers and data mining techniques, it was discovered that very little research has been carried out on real-time environmental temporal parameters and crop disease management.

Table 1. Analysis performance of Data Mining technique on Different Dataset

Sr. No.	Paper	Author	Explanation Features/ Algorithm	Application/ Advantage	Gap Identify
[1]	"Comparative Study of Knowledge in Crop Diseases Using Machine Learning Techniques"	P. Revathi et al. (2011)	The techniques of data mining as, • C4.5 • SVM machine learning • ABC algorithm • Artificial neural networks were presented	Application of data mining technique. Accuracy and Performance obtained	A Multidisciplinary approach to combining computational with agriculture can aid in the accurate forecasting and management of agricultural crops.
[2]	Disease Prediction in Data Mining Technique – A Survey	Laharika et al.	1.The training dataset Is used for reference. 2.Different algorithm used such as C4.5, Decision Tree, Naïve Byes, KNN	Here, some statistical tests are performed to find out possible output prediction with the help of some static data set and compared with this technique with data sets	1.Accuracy is different for each sample dataset 2.Can perform the Analysis on image dataset
[3]	"Early Detection of Pomegranate Disease Using Machine Learning and IoT"	Dnyanesh Nawale et al.	In this system Hidden Markov Model is used for crop disease Detection also sensor network, is used.	It predicts disease in early stage	It required more time in testing.
[4]	"Naive Bayes Classication Algorithm Based on Small Sample Set"	Yuguang Huang et al (2011)	For classification purpose Naive Bayes algorithm is used	For Text classification	Get Accurate Result only in case of large dataset sample
[5]	"Application of Data Mining Techniques for Medical Data Classification: A Review"	Saima Anwar et al	Data Mining algorithm KNN, SVM, Naïve Bayes	It predicts medical disease	Accuracy is depending on cleanness of dataset
[6]	"Heart disease Prediction Using fuzzy logic System"	Long,et al	Rough set based on • Attribute reduction • Fuzzy logic system used	Heart Disease diagnostic system	--
[7]	"Decision Support System for Rice Plant Disease Diagnosis using Naive Bayes' Algorithm"	Khumu et al (2017)	Naïve Bayes Algorithm	The designed system is intended for the diagnosis of common diseases in rice plant occurred during the life span	

5. Conclusion

A broad survey of data mining approaches used in the agriculture domain for crop disease management was conducted in this paper. Using the most up-to-date classification, prediction methods in the context of data mining and applying them to age-old agricultural practises would be a novel approach to solving some of the farmer's most pressing issues. According to the survey, the majority of researchers have focused on crop disease identification and classification, but there has been less work on crop disease prediction using real-world data. So by using Hybrid data mining approach we can predict correct disease on real filed dataset. We can increase the productivity and quality of the fruit.

References

[1]. Dr. Kamaljit Kaur, Manpreet Kaur," Prediction of Plant Disease from Weather Forecasting using DataMining" IJFRCSCE | April 2018, Available @ http://www.ijfrcsce.org

[2]. Wilfried Baudoin, Remi Nono-Womdim, NeBambi Lutaladio, Alison Hodder , "Good Agricultural Practices for greenhouse vegetable crops", Food and Agriculture Organization of the United Nations Rome, 2013

[3]. Hui Wang, Jorge Antonio Sanchez-Molina,Ming Li and Francisco Rodríguez Díaz ,"Improving the Performance of Vegetable Leaf Wetness Duration Models in Greenhouses Using Decision Tree Learning" Water 2019, 11, 158; doi:10.3390/w11010158

[4]. Dnyanesh Nawale, Kunal Patil, Rakesh Mahajan, Sona Pawara, "Early Detection of Pomegranate Disease Using Machine Learning and IoT." International Journal of Innovative Research in Science, Engineering and Technology 2018. Issue 6, June 2018.

[5]. B. B. Bhanu, K. R. Rao, J. V. N. Ramesh and M. A. Hussain, "Agriculture field monitoring and analysis using wireless sensor networks for improving crop production," 2014 Eleventh International Conference on Wireless and Optical Communications Networks (WOCN), Vijayawada, 2014, pp. 1-7, doi: 10.1109/WOCN.2014.6923043.

[6]. S.R.Rajeswari , Parth Khunteta, Subham Kumar,Amrit Raj Singh,Vaibhav Pandey,"Smart Farming Prediction Using Machine Learning" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-7 May, 2019.

[7]. Geraldin B. Dela Cruz, Member, IACSIT, Bobby D. Gerardo, and Bartolome T. Tanguilig III,"Agricultural Crops Classification Models Based on PCA-GA Implementation in Data Mining",International Journal of Modeling and Optimization, Vol. 4, No. 5, October 2014

[8]. Potharaju, S. P., Sreedevi, M., & Amiripalli, S. S. (2019). An Ensemble Feature Selection Framework of Sonar Targets Using Symmetrical Uncertainty and Multi-Layer Perceptron (SU-MLP). In Cognitive Informatics and Soft Computing (pp. 247-256). Springer, Singapore.

[9]. S. More and M. Nighot, "An agro advisory for pomegranate field using wireless sensor network," 2016,International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, 2016, pp. 557-562, doi: 10.1109/ICACDOT.2016.7877647.

[10]. Manisha Bhange, H.A.Hingoliwala,"Smart Farming: Pomegranate Disease Detection Using Image Processing." Second International Symposium on Computer Vision and the Internet IEEE 2015.

[11]. Y. Huang and L. Li, "Naive Bayes classification algorithm based on small sample set," IEEE International Conference on Cloud Computing and Intelligence Systems, Beijing, 2011, pp. 34-39, doi:10.1109/CCIS.2011.6045027.

[12]. D. Mondal, A. Chakraborty, D. K. Kole and D. D. Majumder, "Detection and classification technique of Yellow Vein Mosaic Virus disease in okra leaf images using leaf vein extraction and Naive Bayesian classifier," 2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI),Faridabad, 2015, pp. 166-171.

[13]. Potharaju, S. P. (2018). An Unsupervised Approach For Selection of Candidate Feature Set Using Filter Based Techniques. Gazi University Journal of Science, 31(3), 789-799.

[14]. G. Prem Rishi Kranth, M. Hema Lalitha, Laharika Basava ,Anjali Mathur,"Plant Disease Prediction using Machine Learning Algorithms." International Journal of Computer Applications (0975 – 8887) Volume182 – No. 25, November- 2018

[15]. Chithra, P. L., & Henila, M. (2017)," Defect identification in the fruit apple using k-means color image segmentation algorithm", International Journal of Advanced Research in   Computer Science, 8(8).

[16]. Suraj Khade., et al(2016),". Fruit quality evaluation using k-Means clustering Approach", International Journal of Advances in Science Engineering and Technology,4(2),2321-9009.

[17]. Arlimatti, S. R. (2012),"Window based method for automatic classification of apple fruit", International Journal of Engineering Research and Applications, 2(4), 1010-1013.

[18]. Mitunkumar Balar1 Ronakkumar Patel2," Crop Production Estimator", IJSRD - International Journal for Scientific Research & Development| Vol. 6, Issue 11, 2019 | ISSN (online): 2321-0613

[19]. T. Siddique, D. Barua, Z. Ferdous and A. Chakrabarty, "Automated farming prediction," 20 Intelligent Systems Conference (IntelliSys), London, 2017, pp. 757-763

[20]. P. Revathi , R. Revathi and Dr.M.Hemalatha ,"Comparative Study of Knowledge in Crop Diseases Using Machine Learning Techniques",(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (5) , 2011, 2180-2182.