SURVEY ON FRUIT DISEASE IDENTIFICATION IN AGRICULTURE USING DATA MINING

[1]P. Kanjana Devi, [2]Dr. M. Rathamani

[1]Research Scholar, [2]Associate professor

[1, 2]Dept of master of computer applications,

[1,2]N.G.M. College of arts and science, pollachi.

Abstract

The field of agriculture needs to manage a lot of data and processing and recovery of significant data from this abundance of agricultural information is necessary to support the farmers. In this way, proper strategies and techniques are required for overseeing and sorting out this data to increase the efficiency and agricultural productivity. Data mining can process and convert this crude data into helpful information for improving agriculture. Diseases in fruit cause real production and economic misfortunes just as reduction in both quality and amount of agricultural products By and by a day's fruit diseases detection has gotten expanding consideration in checking colossal field of harvests. Data mining is the process of discovering and extracting of intriguing patterns and knowledge from a lot of data. In this paper there are different data mining techniques utilized for processing of agricultural information/data such as k-means clustering, k-nearest neighbor, artificial neural networks, support vector machine, naive Bayesian classifier and fuzzy c-means are described.

Keywords: fruit diseases, data mining, k-means clustering, k-nearest neighbour, artificial neural networks, support vector machine, naive Bayesian classifier, fuzzy c-means.

1. Introduction

Data mining is the collection of exponentially developing techniques which are utilized to locate some valuable information, patterns and knowledge from datasets. This valuable information advances existing research and improve productivity [1] . The applications of data mining are uncountable; it is utilized nearly in each aspect of life. Some huge uses of data mining are Healthcare, Market Analysis, Finance, Education, Manufacture Engineering, Corporation Surveillance and Agriculture. The data mining techniques can broadly recognize into two kinds predictive and descriptive. The predictive family can further classify into classification, regression and time arrangement examination. In healthcare industry, data mining assumes a significant job in predicting diseases. In early days, the monitoring and examination of plant diseases were done physically by the mastery person in that field. This process requires colossal measure of work hence requires excessive processing time [2]. Utilization of information technology in agriculture can change the situation of decision making. Data mining assumes a crucial job for decision making on a few issues identified with agriculture field. Data mining techniques can be basically isolated in two gatherings: classification and

clustering techniques. Classification techniques are designed for classifying unknown examples utilizing information given by a lot of classified examples. This set is typically alluded to as a preparation set, because, when all is said in done, it is utilized to prepare the classification technique how to play out its classification. For instance, neural networks and support vector machines endeavor preparing sets for tuning their parameters so as to tackle a particular classification issue. At the end of the day, these two classification techniques gain from a preparation set how to classify unknown examples, i.e., tests whose classification is unknown. Another classification technique, the k nearest neighbor, doesn't have any learning stage, because it utilizes the preparation set each time a classification must be performed. To extract the knowledge, data mining algorithms initially dissect the data at that point generate the meticulous kinds of patterns. The data mining model can utilize this outcome to select and characterize the ideal parameters. These parameters are utilized to coerce the significance full information. The mining model of an algorithm produces from the data, it can take different structures, such as clusters, decision trees, mathematical models, and as a lot of regulations.

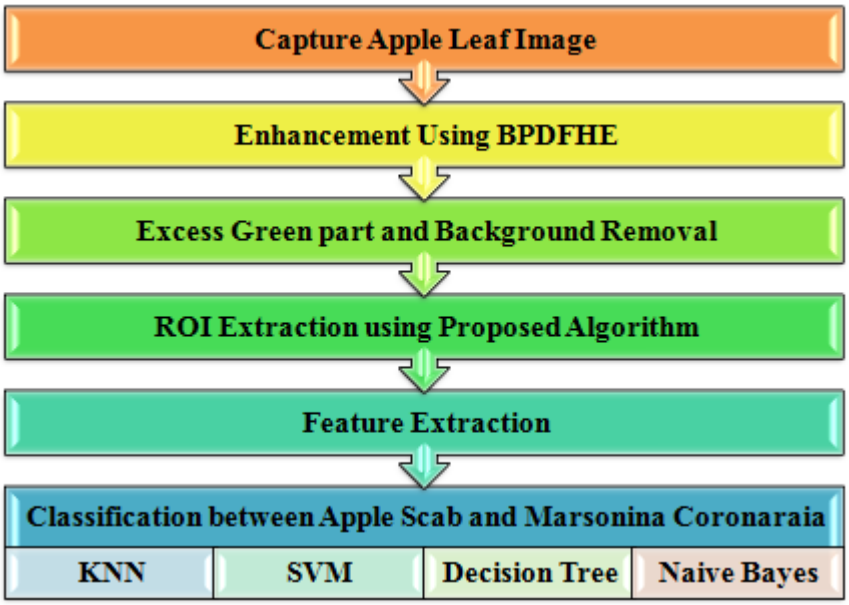

Figure 1: Fruit disease Identification in agriculture.

With an increasing population and a commensurate requirement for increasing agricultural production, there is a critical need to improve management of agricultural resources. Almost 66% of population in India directly relies upon agriculture for its work and consequently, agriculture is the backbone of the Indian economy. The productivity of agriculture is exceptionally low because only 33% of the cropped part is inundated. So as the demand of nourishment is increasing, the farmers, agricultural scientists and government are attempting to put an additional exertion by actualizing techniques for more nourishment production [3]. Yet at the same time today, farmers are performing agriculture-related tasks physically and a not many farmers are utilizing the new strategies, instruments and techniques of cultivating for better agriculture production. In addition, in traditional crop field management, uniform information application not just consider the concept of spatial and fleeting inconstancy inside a crop field, yet in addition brings about environmental pollution and

reduction of homestead profits. The need of site-specific management or precision agriculture has been advocated by researchers, producers and farmers in the around the world. Advanced information technology that can give quick and cost-effective approaches to distinguish spatial inconstancy inside crop fields is the premise of precision agriculture. Additionally, remote sensing technologies have advanced quickly in recent years and have become effective apparatuses for site-specific management in crop protection and production. Data mining speaks to a lot of specific strategies and algorithms pointed exclusively at extracting and patterns from crude data. It helps in the process of discovering already unknown and possibly intriguing patterns with regards to huge datasets. Data mining, which is additionally named as knowledge discovery, is the process of breaking down data from alternate points of view and abridging it into profitable information. This information produced from data mining can be utilized for assortment of purposes like research, future forecasting or prediction, classification etc. in agriculture [4]. Investigation of data in effective manner requires understanding of proper techniques of data mining. The objective of this paper is to describe various data mining techniques in perspective of agriculture area.

2. Literature Survey

Milos Ilic, Sinisa Ilic, Srdjan Jovic , Stefan Panic (2018) Proposed the cherry fruit pathogen ailment detection based on data mining prediction. The utilization of various techniques actualized in MATLAB for conceivable fruit ailment infection prediction. Adjacent to the prediction of potential diseases Mat lab techniques were utilized for data processing, visualization, exception detection and correction. By predicting the time for infection occurrences, the chemical protection is progressively efficient so farmers could set aside cash and on the opposite side, which is unquestionably increasingly significant, it gives more beneficial nourishment because of reduced number of chemical medicines. **Emny Harna Yossya , Jhonny Pranataa , Tommy Wijayaa , Heri Hermawana , Widodo Budihartoa** (2017) Proposed the Mango Fruit Sortation System utilizing Neural Network and Computer Vision. The mango fruit sortation system can use neural network and computer vision strategy to classify mangoes suitably with high accuracy reach 94%. Two aspects that decide the speed and accuracy of the mango clipping are finding the ideal number of shrouded layers while for the learning rate the littler worth will make the learning process more slow however the accuracy will increase generally the more noteworthy the learning rate will make the preparation process quicker yet the worth little accuracy. The preprocessing process is utilized to detect and extract character includes on mango fruit, feed forward process used to configure mango that is ready, juvenile, spoiled, and doable for fare. To utilize this application, the client put the mango fruit in front of the camera and at that point on the off chance that it is proper the client can begin preparing by utilizing a certain button. The system will process the mango fruit through the camera and show on the screen whether it is experienced or youthful. **Md. Rasel Howlader, Umme Habiba, Rahat Hossain Faisal and Md. Mostafijur Rahman** (2019) proposed the automatic recognition of guava leaf diseases utilizing profound convolution neural network. The proposed model applies to classify significant diseases of guava leaf such as Algal Leaf spot, Whitefly, and Rust. The complete process

of D-CNN technique isolated into a few necessary strides for identification of guava leaf diseases detection. The identification accuracy indicates that the proposed D-CNN model is increasingly effective and gives a superior solution to recognize and control guava leaf diseases. To add more categories in the dataset and will prepare and test various types of leaf diseases utilizing this proposed model. Moreover, it consider interfacing the model on advanced cells to make a quick and responsible judgment which can help the farmers immediately detecting and averting diseases. **M.G. Hill, P.G. Connolly,P. Reutemann , D. Fletcher** (2014) Proposed to create models to forecast the outcome of leafroller nuisance monitoring decisions on 'Hayward' kiwifruit crops. Utilizing industry splash journal and vermin monitoring data accumulated at an orchard block level for compliance purposes, 80 characteristics (autonomous factors) were created. Models with a combination of qualities from every one of the three trait categories produce the most accurate forecasts, however orchard management credits seem, by all accounts, to be more powerful than insecticide characteristics. With bigger data sets, it might be conceivable utilizing this technique to pick out combinations of insecticides and application times that give prevalent leafroller control. The outcomes point to the more prominent importance of post-blooming insecticide application traits. **Youssef Es-saady, Ismail El Massi, Mostafa El Y assa,Driss Mammass, Abdeslam Benazoun** (2016) Proposed to automatic recognition of plant leaves diseases dependent on consecutive blend of two SVM classifiers. The proposed framework depends on consecutive blend strategy of two SVM classifiers. The primary classifier uses the shading to arrange the photos; it considers, at this stage, the diseases with relative or closest shading having a place with a comparable class. By then, the subsequent classifier is used to isolate between the classes with relative shading as per the shape and surface highlights. Thusly, the calculation is taken a stab at six classes of plant leaves diseases including three sorts of aggravation insects harms (Leaf diggers, Thrips and Tuta absoluta) and three kinds of pathogens signs (Early curse, Late scourge and Powdery mold), which basically attack the vegetable yields. As such, it will give a choice emotionally supportive network that empowers the ranchers in examination to organize.

Yang Lu, Shujuan Y, Nianyin Zeng , Yurong Liud, Yong Zhang (2017) Proposed to Identification of rice diseases utilizing deep convolutional neural networks. proposed a novel rice diseases identification strategy dependent on profound convolutional neural networks (CNNs) methods. Utilizing a dataset of 500 regular pictures of unhealthy and solid rice leaves and stems captured from rice. The proposed CNNs-based model can viably order 10 regular rice diseases through pictures recognition. The application to the rice sickness identification that the proposed CNNs model can accurately and viably perceive rice diseases through picture recognition. The proposed strategy has a superior preparing performance, quicker convergence rate, just as a superior recognition capacity than the other model. Deep convolution neural networks have achieved incredible performance breakthroughs in machine learning fields. To improve rice diseases identification accuracy, the need to give thousands of top notch rice diseases pictures tests. Moreover, there are countless parameters in CNNs. **Anne-Katrin Mahlein & Erich-Christian Oerke & Ulrike Steiner & Heinz-Wilhelm Dehn** (2012) Proposed the Recent advances in sensing plant diseases for precision crop protection. Precision crop protection focuses on the improvement of sensors for close go (proximal) sensing of crops and diseases

continuously. The interpretation of sensor data, especially spectral data, by the by, is as yet a restricting factor for the detection, identification and quantification of diseases in crop production. **Md. Rasel Mia, Sujit Roy, Subrata Kumar Das, Md. Atikur Rahman** (2019) Proposed the mango leaf diseases recognition utilizing neural network and support vector machine. To detect the manifestations of plant diseases effectively with machine learning than a manual monitoring system. Here, Trained data are produced by classification technique collecting pictures of leaves that were different malady affected. The proposed system could successfully detect and classify the inspected ailment with normal accuracy of 80%. To detect four sorts of infection and one typical among absolute five gatherings. Each gathering consists of four unique sorts of pictures. **Mr. Ketan D. Bodhe, Mr. Himanshu V. Taiwade, Mr. Virendra P. Yadav, Mr. Nikesh V. Aote** (2018) Proposed to Detection and Diagnosis of Cotton Leaf Diseases utilizing Rule Based System for Farmers. By utilizing the actualized system, farmers can without much of a stretch distinguish the ailment on cotton plant and they can apply the correct control measure right away. Hence this system basically amplifies the productivity of crop and it improves the crop yield. There are four malady infected cotton leaf pictures (separated pictures which covering the all conditions of ailment) are utilized. In case of analysis the system will ask a bunch of questions about the side effects and risk factors to the system client and client should give bipolar (yes or no) answer. And it utilizes a score accumulation technique to decide the degree of impact of cotton leaf infection. According to the answer, the system will make decision level of the illness and accordingly pesticides are prescribed. The simple detection from the Knowledge Base Expert System makes the system progressively efficient to utilize. The proposed system detects the patterns of infection in their beginning time.

Hilman F. Pardede, Endang Suryawati, Rika Sustika, and Vicky Zilvan (2018) Proposed the Unsupervised Convolutional Autoencoder-Based Feature Learning for Automatic Detection of Plant Diseases. The usage of convolutional autoencoder has two crucial preferred position. To begin with, the utilization of handcrafted highlights isn't necessary as the network itself may figure out how to produce discriminative highlights. Secondly, the procedure is conducted in an unaided way and hence, no marking of the data are required. Here, to utilize the yield of the autoencoder as contributions to SVM-based classifiers for automatic detection of plant diseases. Normally, completely connected layers are replaced by convolutional layers for picture data. A typical convolution layer consists of a convolution operation and pooling. The yield of the auto encoders are then utilized as contribution for SVM classifiers. By utilizing convolutional layers are more effective than utilizing a completely connected networks for this task. **Milos Ilic, Petar Spalevic, Mladen Veinovic, Abdolkarim Abdala M. Ennaas** (2015) Proposed the Data mining model for early fruit diseases detection. For this reason, various sorts of data mining techniques were assessed on exceptional data sets. One of the most unpredictable and complex task is chemical protection. The key factor for successful chemical fruit protection from diseases and bugs is only the correct minute. This means the selection of chemicals isn't as complex as timing determination for protection. Early fruit infection detection has a great deal of advantages. From the point of farmers, strategies like proposed one give significant information to successful chemical protection. Second advantage for the farmers is economical. They can set aside cash on the off chance that they

reduce quantities of chemical medicines. This is because model indicates when conditions for diseases advancement are not satisfied. In that case chemical treatment isn't required. From the perspective of sound nourishments, reduced number of chemical medications is significant. With suitable detection and prediction it get successful chemical protection and sound nourishment. **Ashwini Awate, Damini Deshmankar, Prof. Samadhan Sonavane** (2015) Proposed the Fruit Disease Detection utilizing Color, Texture Analysis and ANN. OpenCV library is applied for execution. K-means clustering technique is applied for picture segmentation, the pictures are catalog and mapped to their respective illness categories on premise of four element vectors color, morphology, surface and structure of opening on the fruit. The system utilizes two picture databases, one for implementation of question pictures and the other for preparing of already put away illness pictures. Artificial Neural Network (ANN) concept is utilized for example matching and classification of diseases. For Grape - Black Rot, Powdery Mildew, Downy Mildew; For Apple - Apple Scab, Apple Rot, Apple Blotch; For Pomegranate - Bacterial Blight, Aspergillus Fruit Rot, Gray Mold diseases are detected and classified. Once diseases are detected legitimate medicines are recommended accordingly.

Conclusion

Recent technologies can give a great deal of information on agricultural-related activities, which can then be investigated so as to discover significant information. There is huge measure of data in agriculture that is currently accessible from numerous resources and numerous applications of data mining techniques are recently being utilized in agriculture. There are various algorithms to actualize data mining based Fruit infection identification in Agriculture. These various algorithms for K means algorithm, K-nearest neighbor algorithm (K-NN) and Support Vector Machine (SVM) are discussed. In this survey totally 100 papers were selected out of 30 papers has been reviewed for further research.

References:

[1] Milos Ilic, Sinisa Ilic, Srdjan Jovic , Stefan Panic," Early cherry fruit pathogen disease detection based on data mining prediction", Computers and Electronics in Agriculture © 2018 Elsevier

[2] Emny Harna Yossya , Jhonny Pranataa , Tommy Wijayaa , Heri Hermawana , Widodo Budihartoa," Mango Fruit Sortation System using Neural Network and Computer Vision", 2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI 2017, 13-14 October 2017, Bali, Indonesia

[3] Md. Rasel Howlader, Umme Habiba, Rahat Hossain Faisal and Md. Mostafijur Rahman," Automatic Recognition of Guava Leaf Diseases using Deep Convolution Neural Network", 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019

[4] M.G. Hill, P.G. Connolly,P. Reutemann , D. Fletcher," The use of data mining to assist crop protection decisions on kiwifruit in New Zealand", Computers and Electronics in Agriculture 2014 Elsevier

[5] Youssef Es-saady, Ismail El Massi, Mostafa El Y assa,Driss Mammass, Abdeslam Benazoun," Automatic recognition of plant leaves diseases based on serial combination of two SVM classifiers", 2nd International Conference on Electrical and Information Technologies ICEIT'2016

[6] Yang Lu, Shujuan Y, Nianyin Zeng , Yurong Liud, Yong Zhang," Identification of rice diseases using deep convolutional neural networks", Neuro computing © 2017 Elsevier

[7] Anne-Katrin Mahlein & Erich-Christian Oerke & Ulrike Steiner & Heinz-Wilhelm Dehn," Recent advances in sensing plant diseases for precision crop protection", Eur J Plant Pathol (2012) 133:197–209

[8] Md. Rasel Mia, Sujit Roy, Subrata Kumar Das, Md. Atikur Rahman," Mango Leaf Diseases Recognition Using Neural Network and Support Vector Machine", Md. Rasel Mia et al. / International Journal of Computer Science Engineering (IJCSE) Vol. 8 No.01 Jan-Feb 2019

[9] Mr. Ketan D. Bodhe, Mr. Himanshu V. Taiwade, Mr. Virendra P. Yadav, Mr. Nikesh V. Aote," Implementation of Prototype for Detection & Diagnosis of Cotton Leaf Diseases using Rule Based System for Farmers", Proceedings of the International Conference on Communication and Electronics Systems (ICCES 2018) IEEE Xplore Part Number:CFP18AWO-ART; ISBN:978-1-5386-4765-3

[10] Hilman F. Pardede, Endang Suryawati, Rika Sustika, and Vicky Zilvan," Unsupervised Convolutional Autoencoder-Based Feature Learning for Automatic Detection of Plant Diseases", 2018 International Conference on Computer, Control, Informatics and its Applications

[11] Milos Ilic, Petar Spalevic, Mladen Veinovic, Abdolkarim Abdala M. Ennaas," Data mining model for early fruit diseases detection", 23rd Telecommunications forum TELFOR 2015

[12] Ashwini Awate, Damini Deshmankar, Prof. Samadhan Sonavane," Fruit Disease Detection using Color, Texture Analysis and ANN", c 2015 IEEE.

[13] Manisha A. Bhange, Prof. H. A. Hingoliwala "A Review of Image Processing for Pomegranate Disease Detection" International Journal of Computer Science and Information Technologies, Vol. 6 (1), 2015, 92-94.

[14] Vinita Tajane, Prof. N.J. Janwe "Medicinal Plants Disease Identification Using Canny Edge Detection Algorithm, Histogram Analysis and CBIR" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 6, June 2014.

[15] Tejal Deshpande, Sharmila Sengupta, K. S. Raghuvanshi "Grading & Identification of Disease in Pomegranate Leaf and Fruit" International Journal of Computer Science and Information Technologies, Vol. 5 (3), 2014, 4638-4645.