ANALYSIS AND 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

Diseases in fruit cause wrecking issue in economic misfortunes and production in agricultural industry around the world. Disease management by physically is a challenging task. A large portion of the diseases are seen on the leaves or stems of the plant. Hence agriculturist needs to discover the efficient techniques. To reduce the cost of production and enhance the quality and amount of any fruit. It will be extremely valuable for the farmers to detect the disease in beginning time. Data mining is the process of discovering and extracting of intriguing patterns and knowledge from a lot of data. There are various data mining strategies to anticipate the distinctive fruit diseases. In this paper the k-means clustering, k-nearest neighbor, support vector machine algorithms are examined.

Keywords: Data mining, k-means clustering, k-nearest neighbor, support vector machine algorithm, Feature Extraction.

1. Introduction

Data mining in agriculture applications includes the conceptualization, design, improvement, estimation and application of current ways for using the information and communication technologies (ICT) in rustic domain including with the significant objective on agriculture productivity. Distinctive displaying processes and simulation strategies have been executed for dynamic systems in agriculture [1]. The real challenge in agriculture is that no specific apportions have been taken with the huge arrangements

of agricultural data. Data mining is the strategy for discovering beforehand unknown and possibly intriguing patterns with regards to enormous datasets. The mined information is typically spoken to as a model of semantic structure of database, wherein the model might be utilized on new data for prediction or classification of agricultural data. In present day years, the outer environmental effects of traditional agriculture have been evacuated and bringing about the organic and reasonable agriculture developments. Nowadays, fruit diseases and land degradation are the most significant issues in agriculture [2]. In this way, the improvement of agricultural is monitored by data mining techniques achieved through improved information and communication processes. In this field, numerous strategies have been discovered to help farmers in recognizing diseased fruit [3].

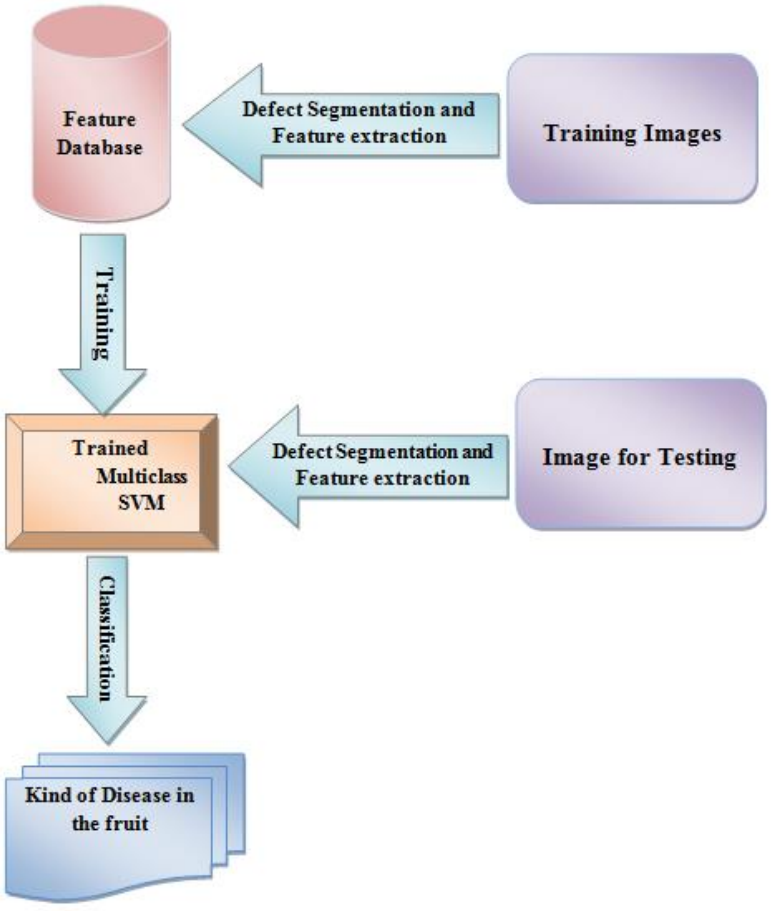

Figure 1: Fruit Disease Identification

Monitoring of wellbeing and detection of diseases is critical in fruits and trees for economical agriculture. Supposedly, no sensor is accessible commercially for the constant appraisal of trees wellbeing conditions. Scouting is the most broadly utilized strategy for monitoring worry in trees, yet it is costly, time-consuming and work concentrated process. Polymerase chain reaction which is a molecular technique utilized for the identification of fruit diseases however it requires point by point testing and processing. The different kinds of diseases on fruits decide the quality, amount, and soundness of yield [5]. The diseases in fruits lessen the yield just as fall apart the grouping and its withdrawal from the development. Early identification of malady and harvest prosperity can encourage the control of fruit diseases through real administration methodologies, for example, vector control through fungicide applications, infection explicit substance applications and pesticide applications; and improved profitability. Fruit diseases can cause significant misfortunes in yield and quality showed up in reaping. For instance, soybean rust (a parasitic disease in soybeans) has caused a significant economic misfortune and just by evacuating 20% of the infection, the farmers may profit with an around 11 million-dollar. An early detection of fruit diseases can help in decreasing such misfortunes and can stop additionally spread of diseases.

2. Literature Survey

Yogeswararao Gurubelli, Malmathanraj Ramanathan, Palanisamy Ponnusamy (2018) Proposed the Fractional fuzzy 2DLDA approach for pomegranate fruit grade classification. The methodology is to use four component extraction based numerical reproduction systems, incorporates conventional 2DLDA. The proposed system holds the most discriminative highlights, by rethinking the fluffy between class dissipate matrix of F2DLDA as fractional fluffy between-class disperse network. Out of four calculations the proposed fractional fluffy 2DLDA gives best acknowledgment rate on the grounds that the disperse networks are fusing the covering test's dissemination data (based on fuzzy membership esteems) and by redefining the SB as fuzzy between class framework, the effect of edge class issue is weakened for the selection of ideal projection direction. The FF2DLDA gives the best recognition rate as compared to the next three techniques. **S. Abirami and M. Thilagavathi** (2019) Proposed the Classification of Fruit

Diseases utilizing Feed Forward Back Propagation Neural Network. Thresholding segmentation separates the affected piece of the fruit images and the features of the portioned region is extracted by Local Binary Pattern (LBP) technique and is utilized as input data in the feed forward back propagation neural network to classify the bacterial and fungal disease of fruit images. Two periods of processing algorithms are utilized in classification, training and testing. In the training stage, characteristic properties of the picture features are segregated and based on these, an extraordinary description of each training class is created. In the testing stage, classification of picture features is done using these feature-space partitions. In the proposed system the classification is done with Feed Forward Backpropagation Neural Network. In this network, the information moves from the input hubs through the concealed hubs to the yield hubs. The accuracy obtained is 92% for bacterial disease and 86% for fungal diseases with 10 shrouded layers. The quantity of concealed layers are fluctuated and the performance of the network is estimated. As the quantity of shrouded units are decreased the percentage of accuracy is decreased. **Dr. N. Sasirekha, N. Swetha** (2015) Proposed an Identification of Variety of Leaf Diseases Using Various Data Mining Techniques. This technique will improves productivity of crops. The Knowledge Discovery in Databases (KDD) process to recognize disease co-occurrences at various degrees of abstraction from a public wellbeing dataset consisting of codes. Utilizing space learning to make an idea of chain of command and fusing the progressive system into the data mining process by abstracting data to an increasingly huge level idea can help constrain the search space and enhance performance and aftereffects of the mining process. The utilization of summed up principle mining for generalizing data to a more elevated level of abstraction in a public wellbeing dataset and compared disease co-occurrences. It maintaining links from the summed up data to the crude data may help constrain the search space and guideline space, minimize information misfortune, and discover new associations. In addition, utilization of various concept hierarchies may affect the outcome. **Dr. Kamaljit Kaur, Manpreet Kaur** (2018) Proposed framework use division methods, for example, k-implies bunching and profound neural network figuring out how to anticipate the malady based on climate feature of the orange plant. This system causes the rancher to understand the disease of orange plant and additionally increase the yield of orange plant.

To build up the system to achieve high accuracy and to predict the disease of orange plant. To used the division method, for example, k-implies bunching and profound neural network figuring out how to recognize the malady part and sickness name of the orange plant based on picture data and climate feature of plant. The proposed gives accurate outcome and give 99% accuracy. This system causes the rancher to understand the disease of the orange plant. To utilize the segmentation techniques such as k-means clustering, SGLDM and deep neural network to detect the disease of the orange plant based on accessible data. **Samiksha Bhor, Shubha Kotian, Aishwarya Shetty, Prashant Sawant** (2017) Proposed to building up a rural passage for yield sickness forecast. This is accomplished by building a web stage in which farmers can interact with master, share their experiences and knowledge. It which eventually helps in data collection that can be utilized to distinguish different crop diseases and averts them. This gateway can be utilized for multiple reasons where agro based industries can utilize the data to launch their products just as acquire feedbacks. Agricultural institutes can investigate new patterns in crop diseases and utilize expected technology to anticipate them. The early detection of diseases on plants is much required as few diseased crops can spread the infection to the entire crop in the field and in this way affects further stockpiling and offers of agriculture products. This system can be utilized for multiple reasons where agro based industries can launch their products and acquire feedbacks. Agricultural institutes can investigate new inventions and technologies for farmers. This system will increase income from by and large crop production. This system will be an interactive one which overcomes communication hindrance through interaction among farmers and specialists.

3. Fruit disease identification using data mining algorithms

3.1 K means algorithm

The k-means is a data mining technique for clustering. Given a lot of data with unknown classification, the point is to find a partition of the set in which comparable data are gathered in a similar cluster. The proportion of similitudes between data tests is given using a reasonable distance: tests that are close to each other are considered comparable.

The parameter k in the k-means algorithm assumes a significant job as it specifies the quantity of clusters in which the data must be partitioned .

The center of the cluster can be considered as the delegate of the cluster, because the center is very close to all examples in the cluster, and in this manner. It pursues that a cluster contains comparable data if every one of its examples are closer to its center and not to the center of some other cluster [8]. Consequently, when tests belonging to a cluster are closer to the center of an alternate cluster, the k-means algorithm moves the corresponding data tests from their original cluster to the new cluster.

3.2 K Nearest Neighbor algorithm (KNN)

The k nearest neighbor (k-NN) is a technique for classification. A training set is known, and it is utilized to classify tests of unknown classification. The basic assumption in the k-NN algorithm is that comparative examples ought to have comparative classification [12]. As in the k-means approach, the likenesses between tests are estimated using appropriate distance functions.

K-Nearest Neighbors or otherwise called K-NN belong to the group of managed machine learning algorithms which means to utilize marked (Target Variable). The K-NN algorithm is a strong classifier which is often utilized as a benchmark for increasingly complex classifiers such as Artificial Neural Network (ANN) or Support vector machine (SVM). To assess any technique by and large look at 3 significant aspects:

1. Simplicity to interpret yield
2. Calculation time
3. Predictive Power

3.3 Support Vector Machine (SVM) algorithm

A Support Vector Machine (SVM) is a discriminative classifier formally characterized by an isolating attention rplane As it were, given marked training data (administered learning), the algorithm yields an ideal publicity rplane which categorizes new models [13]. In two dimensional spaces this hyper plane is a line partitioning a plane in two segments where in each class lay in either side.

SVMs can be even utilized for classifying data which are not linearly detachable. Normally these transformations are not custom fitted to the particular case at hand. Kernels functions are commonly utilized, which implicitly apply a reasonable transformation on the data space. The Gaussian and the polynomial kernels are much of the time utilized for this reason. Commotion and anomalies are taken into account by allowing a portion of the examples in the training set not to be considered if there is a gain in the margin between the two classes.

Another approach is the one-against-one approach. In this case, a single SVM is considered for each pair of classes. Accordingly, n(n - 1)/2 SVMs are required for considering every conceivable pair. During the classification, all SVMs are combined together through a lion's share voting scheme to give the final estimation.

4. Applications of Fruit disease identification in Agriculture

- There are a few applications of Data Mining techniques in the field of agriculture. The K Nearest Neighbor (KNN) is applied for simulating day by day precipitations and other climate factors, and various potential changes of the climate scenarios are examined using SVMs.
- Data mining techniques are applied to concentrate sound recognition issues. For instance, utilizes SVMs to classify the sound of winged animals and other various sounds. There are unit 2 noteworthy kinds of predictions: one will either attempt and predict some inaccessible information esteems or unfinished patterns, or predict a category name for a couple of information. The last is attached to classification. Once a classification model is constructed supported a coaching set, the category mark of associate object might be predicted supported the property estimations of the article and thusly the quality estimations of the categories.
- Prediction is nevertheless additional typically remarked the forecast of missing numerical qualities, or increase/decrease inclines in time connected information. The premier arrangement is to utilize an outsized scope of past qualities to think about plausible future qualities.

5. Experimental Results

Intensity Ratio

K means algorithm	K Nearest Neighbor algorithm	Support Vector Machine algorithm
50	60	78
55.4	65	85
59.6	70	92
60.9	75	95
70.3	81	99

Table 1: Comparison table of Intensity Ratio

The comparison table of Intensity ratio of K means algorithm, K Nearest Neighbor algorithm, Support Vector Machine algorithm shows the different values. While comparing the Intensity ratio of K means algorithm, K Nearest Neighbor algorithm, Support Vector Machine algorithm the Support Vector Machine algorithm is better than the other two algorithms. The K means algorithm value starts from 50 to 70.3, the K Nearest Neighbor algorithm values starts from 60 to 81 and Support Vector Machine algorithm values starts from 78 to 99. Every time the Support Vector Machine algorithm gives the great results.

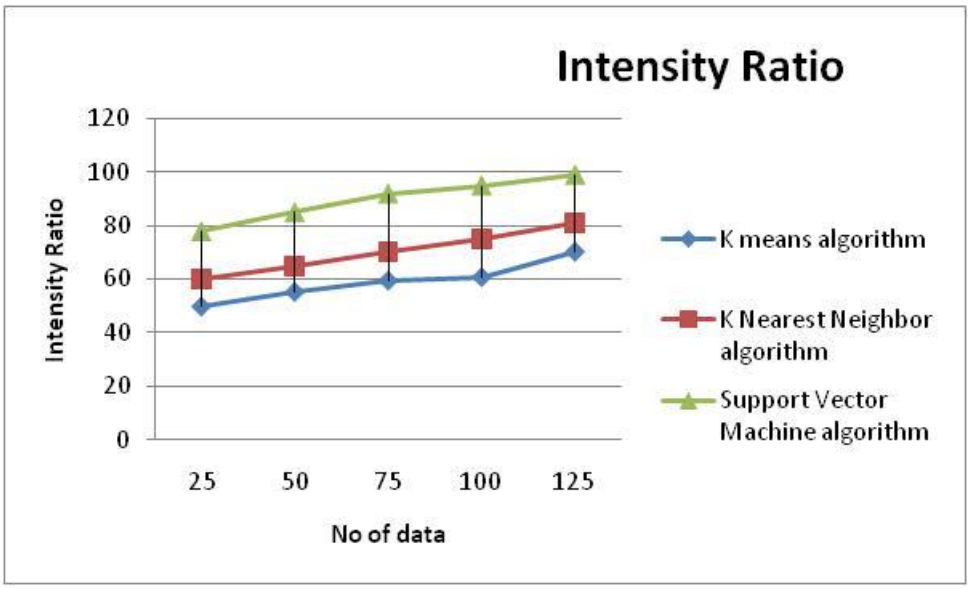

Figure 2: Comparison Chart of Intensity Ratio

The comparison Chart of Intensity Ratio of K means algorithm, K Nearest Neighbor algorithm, Support Vector Machine algorithm demonstrates the different values. No of data in x axis and Intensity Ratio in y axis. The Support Vector Machine algorithm is better than the other two algorithms. The K means algorithm value starts from 50 to 70.3, the K Nearest Neighbor algorithm values starts from 60 to 81 and Support Vector Machine algorithm values starts from 78 to 99. Every time the Support Vector Machine algorithm gives the great results.

Specificity Ratio

K means algorithm	K Nearest Neighbor algorithm	Support Vector Machine algorithm
30.6	45.6	60
36.7	50	65
39.2	54.8	70
45.6	60.9	75
49.5	68.3	80

Table 2: Comparison table of Specificity Ratio

The comparison table of Specificity Ratio of K means algorithm, K Nearest Neighbor algorithm, Support Vector Machine algorithm shows the different values. While comparing the Intensity ratio of K means algorithm, K Nearest Neighbor algorithm, Support Vector Machine algorithm the Support Vector Machine algorithm is better than the other two algorithms. The K means algorithm value starts from 30.6 to 49.5, the K Nearest Neighbor algorithm values starts from 45.6 to 68.3 and Support Vector Machine algorithm values starts from 60 to 80. Every time the Support Vector Machine algorithm gives the great results.

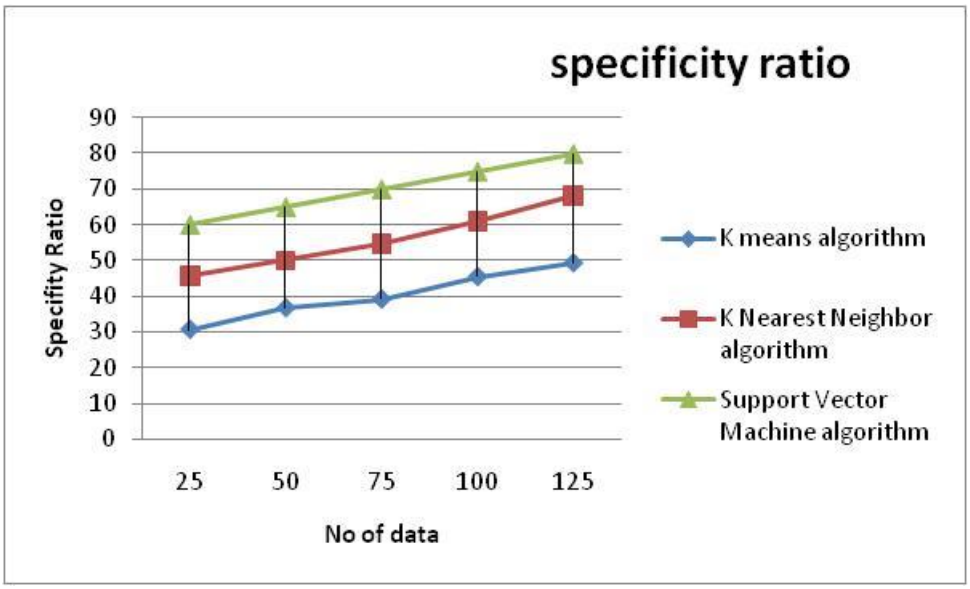

Figure 3: Comparison Chart of Specificity Ratio

The comparison Chart of Specificity Ratio of K means algorithm, K Nearest Neighbor algorithm, Support Vector Machine algorithm demonstrates the different values. No of data in x axis and Specificity Ratio in y axis. The Support Vector Machine algorithm is better than the other two algorithms. The K means algorithm value starts from 30.6 to 49.5, the K Nearest Neighbor algorithm values starts from 45.6 to 68.3 and Support Vector Machine algorithm values starts from 60 to 80. Every time the Support Vector Machine algorithm gives the great results.

Probability Ratio

K means algorithm	K Nearest Neighbor algorithm	Support Vector Machine algorithm
40.6	49.6	55
47.3	56.2	62
55.5	62.1	69
59.3	68.3	77
66.9	77.5	86

Table 2: Comparison table of Probability Ratio

The comparison table of Probability Ratio of K means algorithm, K Nearest Neighbor algorithm, Support Vector Machine algorithm shows the different values. While comparing the Intensity ratio of K means algorithm, K Nearest Neighbor algorithm, Support Vector Machine algorithm the Support Vector Machine algorithm is better than the other two algorithms. The K means algorithm value starts from 40.6 to 66.9, the K Nearest Neighbor algorithm values starts from 49.6 to77.5 and Support Vector Machine algorithm values starts from 55 to 86. Every time the Support Vector Machine algorithm gives the great results.

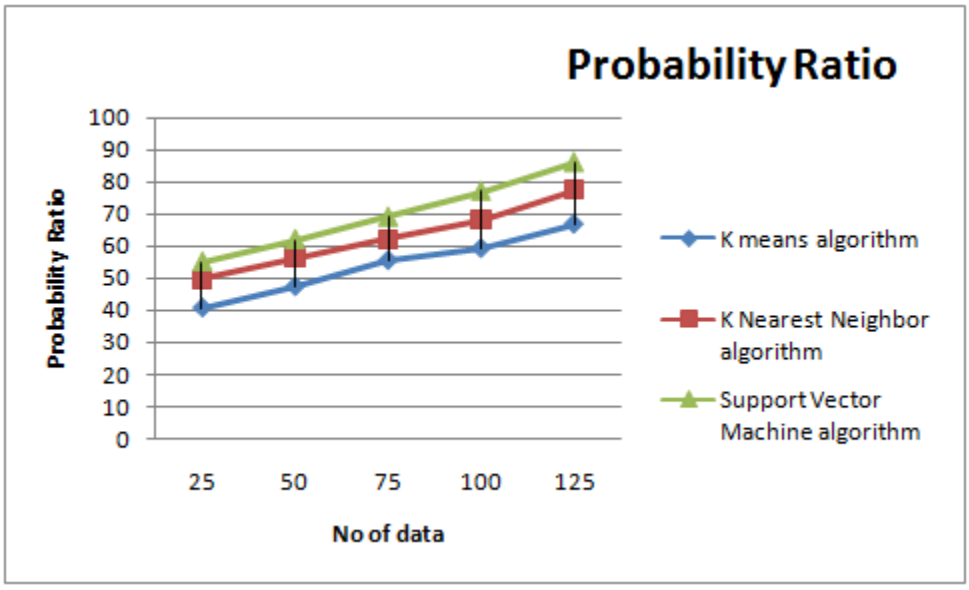

Figure 4: Comparison Chart of Probability Ratio

The comparison Chart of Probability Ratio of K means algorithm, K Nearest Neighbor algorithm, Support Vector Machine algorithm demonstrates the different

values. No of data in x axis and the Probability Ratio in y axis. The Support Vector Machine algorithm is better than the other two algorithms. The K means algorithm value starts from 40.6 to 66.9, the K Nearest Neighbor algorithm values starts from 49.6 to77.5 and Support Vector Machine algorithm values starts from 55 to 86. Every time the Support Vector Machine algorithm gives the great results.

Conclusion

There is enormous 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. The improvement of agricultural is monitored by data mining techniques achieved through improved information and communication processes. In this paper there are various algorithms to actualize data mining based Fruit disease identification in Agriculture. These various algorithms for K means algorithm, K-nearest neighbor algorithm (K-NN) and Support Vector Machine (SVM) are broke down and a portion of the algorithms are compared depending on their performance.

References:

[1] Yogeswararao Gurubelli , Malmathanraj Ramanathan, Palanisamy Ponnusamy," Fractional fuzzy 2DLDA approach for pomegranate fruit grade classification", Computers and Electronics in Agriculture © 2019 Elsevier

[2] S. Abirami and M. Thilagavathi," Classification of Fruit Diseases using Feed Forward Back Propagation Neural Network", International Conference on Communication and Signal Processing, April 4-6, 2019

[3] Dr. N. Sasirekha, N. Swetha," An Identification of Variety of Leaf Diseases Using Various Data Mining Techniques", International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 10, October 2015

[4] Dr. Kamaljit Kaur, Manpreet Kaur," Prediction of Plant Disease from Weather Forecasting using Data Mining", International Journal on Future Revolution in Computer Science & Communication Engineering April 2018

[5] Samiksha Bhor, Shubha Kotian, Aishwarya Shetty, Prashant Sawant," Developing an Agricultural Portal for Crop Disease Prediction", IJIRST –International Journal for Innovative Research in Science & Technology| Volume 3 | Issue 11 | April 2017 ISSN (online): 2349-6010

[6] Saradhambal , Dhivya , Latha , R.Rajesh," PLANT DISEASE DETECTION AND ITS SOLUTION USING IMAGE CLASSIFICATION", International Journal of Pure and Applied Mathematics Volume 119 No. 14 2018, 879-884

[7] Jordi Gene-Mola, Eduard Gregorio, Javier Guevara, Fernando Auat, Ricardo Sanz-Cortiella, Alexandre Escola, Jordi Llorens, Josep-Ramon Morros, Javier Ruiz-Hidalgo, Veronica Vilaplana, Joan R. Rosell-Polo," Fruit detection in an apple orchard using a mobile terrestrial laser scanner", ScienceDirect © 2019 IAgrE.

[8] S Sathiamoorthy, R Ponnusamy, M Natarajan," Sugarcane Disease Detection Using Data Mining Techniques", International Conference "INTELINC 18", 12th & 13th October 2018.

[9] Wenjiang Huang, Qingsong Guan, Juhua Luo, Jingcheng Zhang, Jinling Zhao, Dong Liang, Linsheng Huang, and Dongyan Zhang," New Optimized Spectral Indices for Identifying and Monitoring Winter Wheat Diseases", IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, VOL. 7, NO. 6, JUNE 2014.

[10] Mokashi , ghodke, A.S.chadchankar," Crop Disease Prediction Using Data Mining Method", IJIRMPS | Volume 2, Issue 3, 2014.

[11] Shambhavi Dighe, Anuja Jagdale, Amar Chadchankar," Crop Production Prediction System", International Journal of Research in Engineering, Science and Management Volume-1, Issue-12, December-2018.

[12] A. Petrakova, M. Affenzeller and G. Merkurjeva, "Heterogeneous versus homogeneous machine learning ensembles," Information Technology and Management Science, vol. 21, pp. 135-140, 2015.

[13] D. Moshou et al., "Crop health condition monitoring based on the identification of biotic and abiotic stresses by using hierarchical selforganizing classifiers," in Precision agriculture'15, 2015. Wageningen Academic Publishers, 2015, vol. 8, pp. 619-626.

[14] M. Stoces, J. Vanek, J. Masner and J. Pavlek, "Internet of things (iot) in agriculture-selected aspects," Agris on-line Papers in Economics and Informatics, vol. 8, pp. 83-88, 2016.

[15] B. Milovic and V. Radojevic, "Application of data mining in agriculture," Bulgarian Journal of Agricultural Science, vol. 5, pp. 18- 21, 2017.