Fruit Disease Prediction Using Machine Learning Over Big Data

M.T Vasumathi, M. Kamarasan

Abstract:Big Data Analytics (BDA) offers a stupendous part where there is a want of rebellious performance in managing massive quantity of facts th at handles 4 traits such as Volume Velocity ,Variety and Veracity. Agriculture is one of the fields which generate information
constantly protecting all four traits with excellent growth. There are a number of challenges in processing agricultural records which deals with variety of structured and unstructured format. One of the challenges in agriculture industry comprises of fruit disease detection and control. For this purpose farmers had to monitor fruits continuously from harvest till its growth period. But this task is not aneasy one. Hence it requires proposing an efficient clever farming method which will help for better yield and growth with less human efforts. Image processing is a technique which will diagnose and classify external sickness within fruits through various images. For the control of the disease in the initial stage itself several images of the day to day condition of the fruit has to be monitored where a slight change calls for a remedy. As the number of images increases obviously big data come into play. This paper discusses the existing sytem in fruit disease detection and also proposes disease prediction using machine learning over big-data.

Index Terms: Big Data Analytics, Machine Learning Fruit Disease Detection.

I. INTRODUCTION

Fruits are being affected because of uneven climatic stipulations leading to reduced agricultural yield. This impacts world agricultural economy.Moreover, situation will become even worst when the fruits are infected by any disease. Also, growing population burdens farmers to extend yield. This is where current agricultural strategies and structures are wished to become aware of and prevent the fruits from being affected by extraordinary diseases. Big data analytics[1] approach helps farmers for identifying fruit ailment by using uploading fruit picture to the system. The machine has an already trained dataset of photos for the fruit. Input image given with the aid of the user undergoes numerous processing steps to discover the severity of ailment by way of evaluating with the trained dataset images

Revised Manuscript Received on December 22, 2018.
M.T Vasumathi, Department of Computer and Information Science Annamalai University, Annamalainagar-608002,Tamil nadu, India
vasumathi_mt@yahoo.com
M. Kamarasan,Assistant Professor, Department of Computer and Information Science, Annamalai University,Annamalainagar-608002,Tamil nadu, India
smkrasan@yahoo.com

In order to detect fruit disease two image databases are used .One database is used for training[2] and the other database is used for testing[3]. The fruit disease detection involves two phases namely training phase and testing phase.

In the training phases input images are received then image pre-processing [4]is done. Subsequently feature extraction[5] and clustering [6]is carried out and finally classification process is accompanied.

In the testing phase,input image is obtained from the user and then sequentially the pre-processing and feature extraction processes are done to the image followed by the conclusion of diseased fruit or normal fruit. These steps are shown in the block diagram fig1.

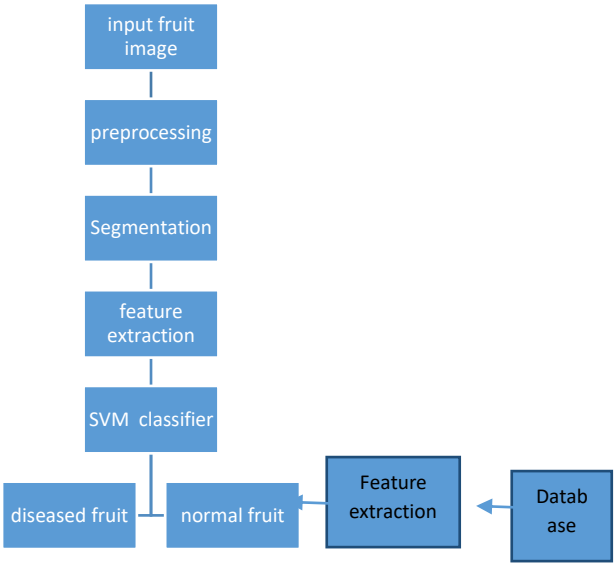


Fig.1 Block diagram for disease detection in fruit

Retrieval Number:F10980476S519/19©BEIESP
Published *By:*
Blue Eyes Intelligence Engineering
& Sciences *Publication*

II. IMAGE PRE-PROCESSING

The input images obtained by sensors on a satellite may contain errors related to form or brightness may be adjusted using suitable mathematical models. Original images acquired from satellites will not have a clear visual effect,hence certain adjustments are carried on to improve the quality of the image.The different enhancement approaches include

A.**Contrast Stretching**: Contrast stretching[7] also called as Normalisation is a simple image enhancement technique that tries to enhance the contrast in an image through `stretching' the range of intensity values it carries to span a preferred range of values.

B.**Noise Filtering**: Noise means, the pixels in the image show different intensity values instead of true pixel values that are obtained from image. Noise filtering[8] is the process of removing or reducing the noise from the image.

C.**Histogram Modification**:The histogram [9] of an image represents the relative frequency of occurrence of the various gray levels in the image.It provides a total description of the appearance of an image. The type and degree of enhancement obtained depends on the nature of the specified histogram.

III. FEATURE EXTRACTION

Feature extraction is a dimensionality reduction approach whichreasonably represents vital parts of an image as a precise feature vector. This technique is useful when image sizes are really large.

IV. CLUSTERING

Clustering is a process of grouping of objects that are similar.K-means clustering is one of the widely used algorithm for clustering. Every cluster in a partition is characterised by its member objects and by its centroid. The centroid is a point to which the sum of distances from all the objects in that cluster is kept minimum.

k-means clustering [10]is an iterative algorithm and its steps are given as follows

Consider an image with resolution of x×y and the image has to be cluster into k number of clusters. Let p(x, y) be an input pixels to be cluster and C_k be the cluster centers. The algorithm for k-meansclustering is following as:

1. Initialize number of cluster k and centre.

2. For each pixel of an image, calculate the Euclidean distance d, between the center and each pixel of an image using the relation given below. $d = p(x, y) - C_k$ (3)

3. Assign all the pixels to the nearest centre based on distance d.

4. After all pixels have been assigned, recalculate new position of the centre using the relation given below. $C_k = 1$ k $y \in C_k$ $x \in C_k$ p(x, y) (4)

5. Repeat the process until it satisfies the tolerance or error value.

6. Rearrange the cluster pixels into image.

V. TRAINING AND CLASSIFICATION

Training and Classification process are carried out using Support Vector Machine (SVM)Algorithm.

Support Vector Machine (SVM) [10]is a supervised machine learning algorithm .Here we plot each data object as a point in n-dimensional space where n represents the number of features we have with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well as shown below.

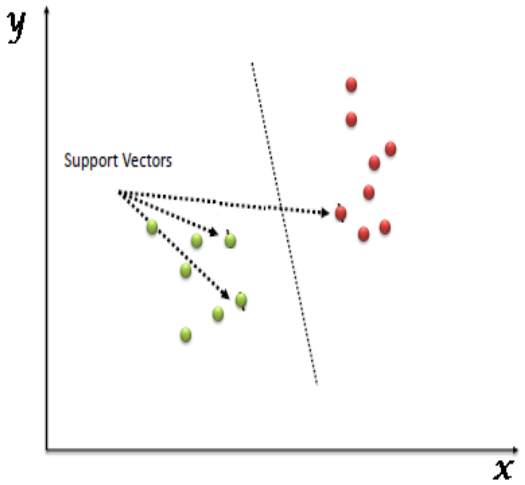


Fig.2 SVM Classifier

Analysis Of Images

VI. CONSTRUCTION OF DATASET:

The training dataset and the testing dataset is obtained by capturing images from digital camera and a mobile phone. Approximately 200 different images of different fruits were captured.The images include fruits which are not affected by disease that is a normal fruit and images of diseases fruits with the degree of disease affected being low to high.

Retrieval Number:F10980476S519/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Three stages of diseased fruit are considered namely low, medium and high.

The low level diseased fruits contain few spots on the surface of the fruit.

The medium level diseased fruits contain more spots which are darker.

The high level diseased fruits contain cracked surface and the colour of the fruit been changed.

The sample images of the fruits are shown in the following figure.

Fig.3 High level diseased fruit

Fig.4 Medium level diseased fruit

Fig.5 Low level Diseased Fruit

Fig.6 Normal Fruit

Table 1 shows the disease detection precision of guava fruit

Fruit Input image class	No. of input images	Disease detection precision in percentage
Low level	15	61%
Medium level	10	70%
High level	25	84%
Normal	30	89%

Table 1.Disease detection accuracy of guava fruit

The precision percentage of low level disease and medium level is less.The detection accuracy is based on the quality of the input images and the size of training dataset.

In order to keep in vigil on the condition of the fruit the data sets of the image has to be checked everyday so that diseases detection accuracy is kept high.But this would result in a huge volume of datasets that could not be handled by the existing system and hence the concept of big data analytics is applied to the problem under consideration.

VII. PROPOSED SYSTEM

Our proposed system will be cheaper than the existing system. Decision Tree Machine Learning[12] Algorithm predicts diseases as well as all sub diseases. Map Reduce Algorithm [13]is applied to increase operational efficiency. It considerably saves Query retrieval time. Also high accuracy is assured.Decision Tree Datasetdetermines diseases well as all the other possible sub diseases. Map Reduce algorithm subdivides the data so that request would be inspected only in the explicit partition, which will increase efficiency but reduce query retrieval time.

Retrieval Number:F10980476S519/19©BEIESP

Published *By:*
Blue Eyes Intelligence Engineering
& Sciences *Publication*

Fig87 Shows the flowchart of fruit disease detection

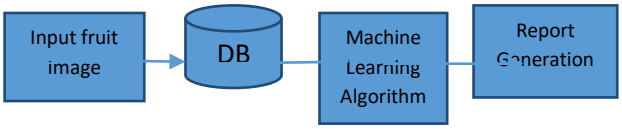
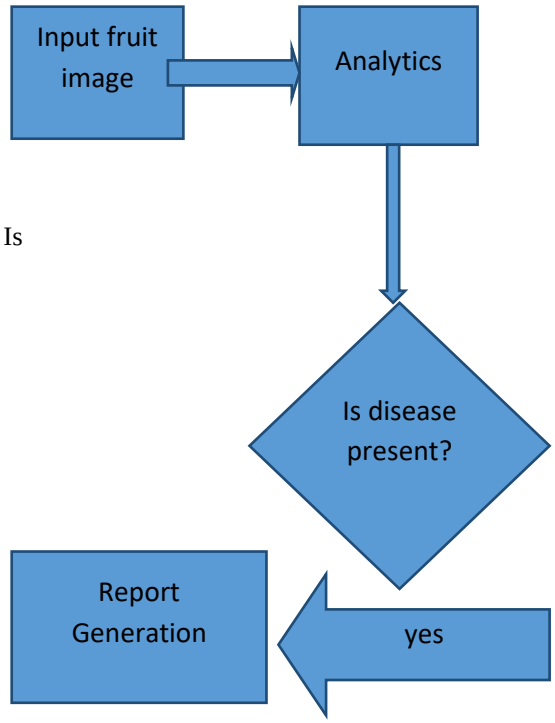


Fig.7 Flowchart showing the sequence of steps in fruit disease detection

Activity Diagram



VIII. CONCLUSION

This paper projects disease detection in fruits in existing system and the proposed system.In the existing system the dataset size used to detect fruit disease is restricted due to the limitations of the algorithms.This in turn the system does not provide accuracy in disease detection and also fail to find the sub diseases .In contrast to existing system,the proposed system that uses big data analytics ensures more accuracy,lesser query retrieval time.It can also result in cheaper cost.Big Data Analytics is a powerful concept that enables to provide huge dataset for training and also day to day condition of the fruits can be evaluated continously.The future scope of the paper is to detect the disease of a fruit more accurately by providing everyday data of the fruit to the system and hence detect and control the disease,The proposed system in future will use machine learning algorithms over big data,

REFERENCES

1. https://www.ias.ac.in/article/fulltext/reso/021/08/0695-0716
2. 5 January 2015; Revised 26 May 2015; Accepted 16 June 2015 Big Data Analytics in Healthcare Ashwin Belle,Raghuram Thiagarajan, S. M. Reza Soroushmehr,Fatemeh Navidi, Daniel A. Beard, and Kayvan Najarian.
3. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 17 (2017) pp. 6338-6346 © Research India Publications. http://www.ripublication.comA study on Deep Machine Learning Algorithms for diagnosis of diseases Dinu A.J.*, Ganesan R1 , Felix Joseph and Balaji V
4. 28 May 2018Hyeok-June Jeong ; Kyeong-Sik Park ; Young-Guk HaImage Preprocessing for Efficient Training of YOLO Deep Learning Networks
5. 18 July 2016 Yushi Chen ; Hanlu Jiang ; Chunyang Li ; Xiuping Jia ; Pedram GhamisiDeep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks
6. https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/
7. Beilei Xu ; Yiqi Zhuang ; Hualian Tang ; Li ZhangObject-based multilevel contrast stretching method for image enhancement
8. 11 August 2003 Noise reduction by fuzzy image filteringD. Van De Ville ; M. Nachtegael ; D. VanderWeken ; E.E. Kerre ; W. Philips ; I. Lemahieu
9. https://nptel.ac.in/courses/117104069/chapter_8/8_3.html
10. Hashmi and T. Ahmad Jamia Millia IslamiaBig Data Mining: Tools & Algorithmshttp://dx.doi.org/10.3991/ijes.v4i1.5350 , Delhi, India

Retrieval Number:F10980476S519/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication