A Survey on Diseases Detection and Classification of Agriculture Products using Image Processing and Machine Learning

Nilay Ganatra
Smt. Chandaben Mohanbhai Patel Institute of Computer Applications
Charotar University of Science and Technology (CHARUSAT)

Atul Patel, PhD
Smt. Chandaben Mohanbhai Patel Institute of Computer Applications
Charotar University of Science and Technology (CHARUSAT)

ABSTRACT
Quality agriculture production is the essential trait for any nation's economic growth. So, recognition of the deleterious regions of plants can be considered as the solution for saving the reduction of crops and productivity. The past traditional approach for disease detection and classification requires enormous amount of time, extreme amount of work and continues farm monitoring. In the last few years, advancement in the technology and researchers' focus in this area makes it possible to obtain optimized solution for it. To identify and detect the disease on agriculture product various popular methods of the fields like machine learning, image processing and classification approaches have been utilized. This paper presents various existing techniques used to detect the disease of agriculture product. Also, paper surveys the mythologies utilized for disease detection, segmentation of the affected part and classification of the diseases. It also includes the summary of various feature extraction techniques, various segmentation techniques and various classifiers along with benefits and drawbacks.

Keywords
Classification, image processing, machine learning, segmentation, feature extraction, pre-processing

1. INTRODUCTION
In agriculture products, diseases are the main cause for the lessening in both quality and production of the agriculture products. So, plant disease diagnosis in early stage is very essential to cure and control them. Farmers puts their great effort in picking best seeds of plant and also provide proper environment for the growth of the plant, although there are lot of diseases that affects plant result in plant disease. In agriculture it is important to discover the plant diseases batches in the early stage which helps us to minimize the damage, reduce production costs, and rise the income. The human eye alone is not that much effective many a times to identify the correct disease. In past farmers used to follow naked eye observation of experts with samples of affected plants or expert used to visit the farm and based on their suggestions farmers were taking the corrective action to cure the plant diseases. In this method, it is very difficult to find trust worthy expert and the solution does not work properly for the large fields, the method takes long time. Also, this method is expensive because it requires continuous monitoring of experts. Agriculture is the root for the economy of any country and so the correct and timely identification of agriculture products diseases is very important. So, we need some automatic, fast, accurate and less expensive methods to detect diseases. Modern technological advancement in the field of image processing and machine learning will help farmers in the aspect of cost reduction of pesticides. Mainly there are two types of factors which cause the diseases in the agriculture products; living and nonliving agents. Insects, bacteria, fungi and viruses come under the category of living agents while temperature changes, excess moisture, insufficient light, less nutrients and pollution in air come under non-living agents. Agriculture related many applications are developed for leaf identification, lead diseases detection, fruit diseases etc. All these applications require digital images which are captured through digital camera. Then image processing and machine learning techniques are applied on captured images to extract necessary information for the analysis. In this paper we are presenting survey on different plant disease detection and classification techniques using image processing and machine learning which is used for fast, spontaneous, automatic and accurate detection as well as classification of plant leaf diseases. The various stages of disease detection and identification are image acquisition, image pre-processing, image segmentation, feature extraction and classification

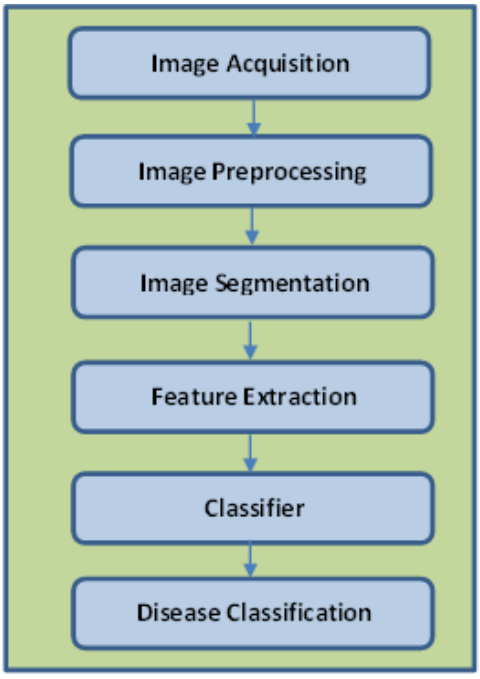

Figure 1. Steps to detect and classify disease from agriculture product.

2. RELATED WORK

Ajay Gurjar & et al (and other) [1]. In this paper three diseases (Red Spots, Leaf Crumple and White Spots) detection have been done using Eigen feature regularization and extraction technique. With the proposed algorithm they are able to achieve 90% accuracy in detection of Red Spot (Fungal Disease).

Yan-Cheng Zhang & et al (and other) [2]. In this paper authors proposed fuzzy selection approach i.e. fuzzy curves (FC) and surfaces (FS) on cotton disease leaves image. They identify the best set of features using fuzzy feature selection approach. This research follows process in two steps. In the first step FC automatically isolate important and required features from the original feature set and eliminate the inaccurate features. Then in the second step by utilizing FS to obtain the feature dependent on the significant feature. This approach is particularly important to reduce the dimensionality of feature space which provides simplified practical implementation for classification applications.

Libo Liu & et al (and other) [3]. In "Extraction of the Rice leaf disease image based on BP neural network" authors' proposed BP Neural network classifier for classifying health and diseased part of rice leaves. In research authors has selected brown spot as research object. They have taken sample images from the northern part of Ningxia Hui autonomous region. In this research coloured features of diseases and health portion passed as input for the BP neural network. This method is also used to identify the other diseases as shown in result.

Sachin D. Khirade & et al (and other) [4]**.** In this paper authors highlighted various stages of disease detection like image acquisition, image pre-processing, image segmentation, feature extraction and classification. Moreover, they have discussed the different methods used for plant diseases detection using images of leaves. Moreover, this paper highlights some feature extraction and segmentation algorithms that can be used for plant disease detection.

Jitesh P Shah & et al (and other) [5]. This paper concisely discusses the important methods of images processing and machine that can be applied to plant disease detection and classification. Survey of 19 papers covering the work on rice plant diseases and other different fruits and plants has been carried out by the authors based on important criteria like size of dataset, no. of classes (diseases), Segmentation and pre-processing techniques, classifiers and its accuracy etc.

Shiv Ram Dubey & et al (and other) [6]. In this paper apple fruit diseases detection and classification solution has been proposed and practically validated. In the proposed solution K-Means clustering technique is used for image segmentation, then feature extraction from the segmented image has been done and Multi-class Support Vector Machine is used for classification. The proposed solution provides accuracy up to 93%.

Monika Jhuria, Ashwani Kumar & et al [7]. In this paper using Image Processing authors monitor the diseases on the fruit right from the plantation to harvesting. Neural network with backpropagation used to train the system. Two diseases of apple and three of grapes are selected for the implementation. Classification and mapping of the images to the respective disease categories is done using colour, texture and morphology feature vectors. Out of these three feature vector, morphology gives 90% correct result over other two.

Bhavini J. Samajpati & et al (and other) [8]. In this paper authors mainly discussed about three diseases which cause major loss in yield and production of good quality apples. Apple scab is gray or brown spots on the apple. Apple rot is small dipped, black or brown which can be occurred by red halo. Apple blotch is fungal disease which appears on the apple surface and makes apple as dark, uneven and lobed edges. This paper survey various methodology utilized for apple disease detection, segmentation and classification using image processing. Also, summary of various texture analysis methods, segmentation techniques and various classifiers along with their pros and cons are discussed.

Godliver Owomugisha & et al (and other) [9]. In this paper authors presents an application of machine learning for the crop disease diagnosis based on the image captured using smartphone. They have represented a classification system with 5 classes to determine the state of disease of a plant. Total 5 classes used to represent health class and 4 for diseases classes. Also, further classification has been done for severity levels of the 4 diseases. The assigned classes for severity are 1-5, where 1 is for healthy plant and 5 is for diseased plant. Also, they have used different feature extraction methods and shows different extraction methods affects the performance of the classifiers. They took four major diseases affecting cassava plant. Farmer can upload the image by smart phone and obtain the disease score from the server.

A. Meunkaewjinda & et al (and other) [10]. Authors of the paper mainly focus on the grape leaf disease diagnosis using multiple artificial intelligent techniques. In the proposed work is mainly divided into three parts: i) pre-processing using grape leaf colour segmentation ii) grape leaf disease segmentation and iii) analysis and classification of diseases. They have used self-organizing feature map together with a back-propagation neural network is used to recognize grape leaf image colour. Also, for segmentation self-organizing feature map with genetic algorithms for optimization is used and a support vector machine is used for classification. Moreover, Gabor wavelet filter is applied to segmented image for better analysis of diseases colour features. Finally, support vector machine is again applied to classify the grape leaf disease from the classes: scab disease, rust disease and no disease.

M. Ravindra Naik & et al (and other) [11]. In this paper authors use SVM to identify whether the leaves are affected by diseases or not. The Genetic algorithm is used for image segmentation, which is crucial aspect for disease identification in agriculture products. And then if it is affected by any diseases than again NN classifier is used to classify the disease of the particular leaf. The proposed algorithm was tested and executed on Banana, beans, jackfruit, lemon, mango, potato, tomato, and sapota.

Suhaili Beeran Kutty & et al (and other) [12]. This paper focuses on the classification of Anthracnose and Downey Mildew, leaf disease of watermelon using neural network analysis. Some of the infected leaf samples were collected using camera with specific calibration procedure under controlled environment. To classify the watermelon's leaf diseases, colour feature extraction from RGB colour model is used where from the identified Regions of Interest (ROI) the RGB pixel colour indices have been extracted. Prosed architecture used Statistical Package for the Social Sciences (SPSS) and Neural Network Pattern Recognition Toolbox in MATLAB.

K. Muthukannan & et al (and other) [13]. In this paper various neural network algorithms are used to classify the detected spot diseases in the leaves to avoid extreme amount of

pesticides for the plant disease treatment. In this paper methodology used to classify the diseased plant are Feed Forward Neural Network (FFNN), Learning Vector Quantization (LVQ) and Radial Basis Function Networks (RBF) by using various shape and texture features from the affected leaf image. The proposed algorithm in the paper is tested on the bean and bitter gourd leaves.

Loyce Selwyn Pinto & et al (and other) [14]. In this paper to detect and classify sunflower crop disease image processing is used. After applying pre-processing on the captured images through the digital camera, the k-means clustering is used to get the diseased part of the leaf. Then various machine learning algorithms are applied on identified part and classified based on colour and texture features. Also, it compares accuracy between various machine learning algorithms like K-Nearest Neighbors, Multi-Class Support Vector Machine, Naive Bayes and Multinomial Logistic Regression to get high accuracy. The feature set consists of Contrast, Energy, Mean, Homogeneity, Standard Deviation and Coarseness.

2.1 Image Segmentation Techniques

The meaning of segmentation is to identify the region of interest from the image. The segmentation is basically used to make simpler and/or change the way of representation of an image into something which is easier and meaningful to analyse [15].

Table 1: Summary of different segmentation techniques [16]

Segmentation Technique	Description	Benefits	Drawbacks
Thresholding Method	It is the simplest method approach of image segmentation by dividing the image pixels based on their intensity level. The threshold value can be computed depending on the peak of the image histogram.	Any prior information about image is not required Fast, simple and computationally inexpensive. Can be easily applicable and suitable for real life applications	It does not work well for image with broad and flat valleys and does not have any peak. Spatial information may be ignored and resultant image cannot guarantee that the segmented regions are contiguous. Threshold selection is very crucial. Extremely noise sensitive.
Region Based Method	In this method construction of segmentation region is based on association and dissociating neighbour pixels. It works on the principle of homogeneity, with the fact the adjacent pixels inside specific region flocks related characteristics and unrelated to the pixel in the other region.	It is flexible enough to choose between interactive and automatic technique for image segmentation. More clear object boundaries by the flow from the inner point to outer region. Gives more accurate result compare to other methods.	Required more computation time and memory and sequential in nature. Noisy seed selection by user leads to faulty segmentation. Because of splitting scheme in region splitting segments seem square.
Clustering Method	In this method pixels having similar characteristics in image are segmented into same clusters. Cluster an image into different parts based on the features of the image. The k-means algorithm is commonly used for this method.	Homogeneous regions can be easily obtained. Computationally faster. K-means works faster for the smaller value of K.	Poor worst-case behaviour. It requires similar size clusters, so the assignment of the adjacent cluster center is the correct assignment.
Edge Based Method	In this method all edges are detected first and then to segment the required region, edges are connected to form the object boundaries. It is based on discontinuity detection in edges.	Works well for the images with better contrast between regions.	Work not well for the image having more edges. Selection of right object edge is difficult.
Partial Differential Equation Based Segmentation Method	These are fast and appropriate for time critical applications. It is based on the differential equation working.	Fastest Method	Computational Complexity is more

2.2 Feature Extraction

When input data to an algorithm is very large and it is supposed to be redundant, than it can be transferred into small set of features. Finding the subset of the initial features is called feature selection. It is expected that selected features contains the required information, so the preferred task can be performed with using reduced representation [17].

Table 2: Summary of different colour techniques

Method	Description	Merits	Demerits
L*a*b [18]	a) This colour space consists one channel for Luminance and two other channels are a and b known as chromaticity layers. b) Space consists of dimension L for lightness and a and b for colour adversary dimensions.	a) In this colour and intensity manage individually. b) It can measure small colour differences.	a) Problem of singularity as other nonlinear transformation.
HSV Histogram [19]	a) HSV can be represented as hexacone in three dimensions in which intensity can be represented as central vertical axis. b) It is Hue, saturation value. c) Colours are described in term of shades and brightness	a) Accuracy is more b) Applicable for real time applications.	a) Sensitivity to lighting variations is less.
RGB [18]	a) It is colour space based on RGB model. b) Consists of three independent image planes, one for each primary colour red, green and blue c) It is an additive model	a) suitable for display	a) It is highly correlative. So, not good for colour image processing
YUV [18]	a) Main channel luminance describes the light intensity like rod cells of the retina b) Chrominance components U and V carry the colour information c) In this black and white colour information is separated from the colour information	a) Overcome the correlation of RGB to some extent and require less computation time	a) Correlation exists but less than RGB

Table 3: Summary of different texture feature extraction techniques [18]

Method	Description	Merits	Demerits
Grey Level Co-occurrence Matrices	a) It is statistical method used to examine the texture which considers the spatial relationship of pixels is the grey level co-occurrences matrix.	a) Feature vector length is small b) Can be applied for the different colour space for colour co-occurrence matrix	a) Many matrices is required to be computed b) It's not invariant with rotation and scaling
Wavelets Transform	a) It works better on the frequency domain rather than the spatial domain	a) Best features with the higher accuracy can be produced	a) It is quite complex and slower
Independent Component Analysis	a) It is computational method for splitting a multivariate signal into additive small subcomponents	a) Higher order statistics can be easily obtained b) It separates mixed signal into a set of independent signals.	a) It is rarely used method.
Gabor filter	a) It is used to analyse specific frequency content in the image in specific directions in a localized region around the region of interest	a) It is multi resolution and multi-scale filter b) It is used for orientation, spectral bandwidth and spatial extent	a) So many filters are used in application so overall computational cost is high.

2.3 Classifiers

Classifiers are used to classify the images according to their features. There are various classifiers like Naive Bayes Classifier, k-Nearest Neighbors (k-NN), Support Vector Machine (SVM), Artificial Neural Network (ANN) and Random Forest Tree Classifier.

Table 4: Summary of different classifiers [18]

Classifier	Description	Merits	Demerits
Naive Bayes Classifier	a) It is Probabilistic classifier b) Strong independence assumption theorem c) value of the particular feature is independent of the value of any other feature	a)Small amount of training data is required for classification	a)Interaction between features can't be learnt because of independency among the feature
K-nearest neighbour	a) It is statistical and non-parametric classifier b) Weight can be assigned to the contributions of the neighbours, so nearer neighbour donate more in the average than the distance neighbour c) Distance metric has been calculated for samples and classify based on this distance d) It uses Euclidean distance to calculate distance	a) Implementation is simple b) Don't required classes to be linearly separable	a) Very Sensitive to noisy or irrelevant data b) More time consuming testing process because requires calculation of distance to all known instances
Support Vector Machine	a) It is based on the decision planes that define decision boundaries. b) There are two stages of its working 1) off-line process 2) online process c) Multi-class support vector machine as a set of binary vector machine is used for training and classification	a) It is effective in high dimensional spaces b) In comparison with other classification techniques classification accuracy is high. c) SVM is robust enough, even though training samples have some distortion.	a) Training time is very high with large data set b) For mapping original data into high dimension data selection of kernel function and kernel parameters is difficult
Decision Tree	a) It repetitively divides the working area into small sub parts by identifying its attributes. b) Leaves present the class labels and branches present features that lead to those classes.	a) Small sized trees can be easily interpreted b) For many simple data sets accuracy is comparable with other classifications	a) For some datasets it is observed to over fit with noisy classification tasks.
Artificial Neural Network	a) It is derived from the concept of the human biological neurons system b) It consist of two datasets one for training and one for testing	a) It is robust and can handle noisy data b) Well suited to analyse complex numbers	a) Requires more training time b) Requires large training samples c) Requires more processing time

3. CONCLUSION

In this paper we have presented survey on different agriculture product diseases detection and classification using various image processing and machine learning techniques. We have highlighted summary of different colour and texture based feature extraction with their advantages and disadvantages. Moreover, we have discussed different segmentation techniques with its benefits and negative marks. Also, summary of different segmentation techniques along with its merits and disadvantages discussed briefly in the paper. In future we are going to use some of the methods discussed in this paper for our research work.

4. REFERENCES

[1] V. A. G. Ajay a. Gurjar, "disease detection on cotton leaves by eigenfeature regularization and extraction technique," internation journal of electronics, communication & soft computing science and engineering(ijecscse), vol. I, no. I.

[2] H.-p. M. B. H. M.-x. L. Yan-cheng zhang, "features selection of cotton disease leaves image based on fuzzy feature selection techniques," in ieee proceedings of the

2007 international conference on wavelet analysis and pattern recognition, beijing, china, 2007.

[3] G. Z. Libo liu, "extraction of the rice leaf disease image based on bp neural network," in ieee, 2009.

[4] P. Sachin d. Khirade, "plant disease detection using image processing," in ieee 2015 international conference on computing communication control and automation, pune, 2015.

[5] H. B. P. V. K. D. Jitesh p. Shah, "a survey on detection and classification of rice plant diseases," in ieee international conference on current trends in advanced computing (icctac), bangalore, 2016.

[6] S. J. Shiv ram dubey, "detection and classification of apple fruit diseases using complete local binary patterns," in ieee 2012 third international conference on computer and communication technology, 2012.

[7] K. B. Monika jhuria, "image processing for smart farming: detection of disease and fruit grading," in proceedings of the 2013 ieee second international conference on image information processing , 2013.

[8] S. D. D. Bhavini j. Samajpati, "a survey on apple fruit diseases detection and classification," international journal of computer applications, vol. 130, no. 13, pp. 25-32, november 2015.

[9] E. M. Godliver owomugisha, "machine learning for plant disease incidence and severity measurements from leaf images," in 15th ieee international conference on machine learning and applications, 2015.

[10] P. K. K. A. A. Meunkaewjinda, "grape leaf disease detection from color imagery using hybrid intelligent system," in ieee 2008 5th international conference on electrical engineering/electronics, computer, telecommunications and information technology, 2008.

[11] M. R. S. M. Ravindra naik, "plant leaf and disease detection by using hsv features and svm classifier," international journal of engineering science and computing, vol. 6, no. 12, pp. 3794-3797, 2016.

[12] N. E. A. H. H. Suhaili beeran kutty, "classification of watermelon leaf diseases using neural network analysis," ieee business engineering and industrial applications colloquium, 2013.

[13] P. L. R. P. S. P. N. K. Muthukannan, "classification of diseased plant leaves using neural network algorithms," arpn journal of engineering and applied sciences , vol. 10, no. 4, 2015.

[14] R. M. U. R. Loyce selwyn pinto, "crop disease classification using texture analysis," in ieee international conference on recent trends in electronics information communication technology, 2016.

[15] F. Encyclopedia, "image segmentation," [online]. Available: https://en.wikipedia.org/wiki/image_segmentation.

[16] X. J. Y. S. J. W. H.d. Cheng, "color image segmentation: advances and prospects," elsevier, 2001.

[17] T. F. E. Wikipedia, "feature extraction," [online]. Available: https://en.wikipedia.org/wiki/feature_extraction.

[18] U. K. J. D. G. T. Uravashi solanki, "a survey on detection of disease and fruit grading," international journal of innovative and emerging, vol. 2, no. 2, 2015.

[19] V. L. M.-f. D. O. Kleynen, "development of a multi-spectral vision system for the detection of defects on apples," elsevier, 2004.