Automatic Citrus Fruit Disease Detection by Phenotyping Using Machine Learning

Conference Paper · September 2019

DOI: 10.23919/IConAC.2019.8895102

Fida Hussain

Ronky Doh
Valley View University

Kwaku Ayepah
Jiangsu University

Proceedings of the 25th International Conference on Automation & Computing, Lancaster University, Lancaster UK, 5-7 September 2019

AUTOMATIC CITRUS FRUIT DISEASE DETECTION BY PHENOTYPING USING MACHINE LEARNING

Benjamin Doh, Duo Zhang, Yue Shen*, Fida Hussain, Ronky Francis Doh, Kwaku Ayepah
School of Electrical and Information Engineering, Jiangsu University
Zhenjiang, China
Corresponding Author: Email: shen@ujs.edu.cn

Abstract - **It is really a frustrating occurrence of how fruit diseases can cause a reduction in production and the economy in the agricultural field all over the world. A growing body of research proves that fruits are critical to promoting good health. In fact, fruits should be the foundation of a healthy diet. This paper presents a better and modern proposed solution for the detection of fruit diseases using their physical attributes. The proposed methods are composed of K-Means clustering technique, ANN and SVM. K-Means is used for the image Segmentation. It has a function of mapping images to their corresponding disease classes based on the phenotypic characteristics such as the texture, color, structure of holes on the fruit and physical make-up. ANN (Artificial Neural Network) is pragmatic in achieving enhanced results in relations to the accuracy of detection and classification have some advantages over the other algorithms. They utilize quite little pre-processing in regards to other image classification methods. It implies the filters were studied by the network that in traditional methods were hand-engineered. This autonomy from earlier information and human exertion in feature design is a major advantage. A Support Vector Machine SVM also is used with the ANN to increase the high rate of classification. The proposed solution can significantly engineer precise detection and automatic classification of fruit diseases.**

Keywords- Support Vector Machine (SVM), K-means clustering, Artificial Neural Network (ANN), Phenotyping.

I. INTRODUCTION

The role Agriculture plays across the world is phenomenal. Nothing in this world beats 'food'. Both humans and animals depend on it on a daily basis, therefore a high premium should be given to agriculture; be it the kind, type, a number of yield or production of any kind of crop produces. A major threat and issue affecting the agriculture sector are diseases. Fruits (citrus) are vital ingredients in agriculture and almost everyone consumes it every day. Citrus fruit diseases are serious issues that are greatly affecting the quality and quantity of yields all over the world. There are different kinds of Diseases affecting citrus fruits. Some of these diseases are canker, greasy spot, and black spot. This paper detects, classifies analyzes some diseases that affect citrus fruits. An analysis was made with the use of Image processing methods so as to analyze the squalor citrus fruits. Techniques related to detection is presented. A proposal has been made in regards to an automated solution for the sign-based recognition of citrus fruit diseases; the proposed solution also tackles the problems present within the citrus fruit images.

Jiangsu University

A dataset is extracted, studied and used in this paper. Citrus fruits are much more important in our lives and the agricultural sectors across the world are always seeking for new ways and techniques to maximize production and as well as minimize reparations caused by diseases in citrus plants.

An (SVM) support vector machine is used to identify plant disease distressing agriculture/horticulture crops. The study has chosen SVM due to its effective applications and performances proved to be ideal for high dimensional issues and small data sets. Viewing training input vector in an n-dimensional space, SVM constructs a hyperplane in the space, that can be utilized for the classification which has the highest distance to the contiguous training data point of any class (functional margin). In computation of the margin, two parallel hyperplanes are created, one on every side of the isolating hyper-plane, which are strapped up in opposition to the two data sets. The aim of this is to decide which class a new data point belongs to based on data factors in relations to one of the two classes. In the case of svms, a data point is computed as a p-dimensional vector (a listings of p numbers) and it is meant to know whether such levels can be pronged by a $(p-1)$ dimensional hyperplane.

Quite a number of techniques have been studied and proposed by researchers within the terrain of machine learning, deep learning, image processing for the detection and classification of citrus disease.

A proposal was made for an early detection of two types of fungi that belong to the Penicillium genus in citrus fruits [1]. The objective is to avoid or if possible, minimize related financial losses. These defects are detectable only by using UV (Ultraviolet) light, which is perilous, but cannot be identified in a visual assessment or in normal RGB (Red, Green, Blue) computer vision framework that uses visible illumination. The proposal has been based on a hyperspectral computer vision framework that permits a more noteworthy capability of separation. Moreover, as the number of features increases to a large degree, they can all not be utilized in a classifier directly, and feature selection methods required.

There is a requirement for high yield in agricultural industries better-quality yield of fruit is important, for this, there is a need for an automated technique which will find disease on fruits [2]. For this artificial neural network methodology is suggested which can be helpful to categories fruit infection. K-Means clustering is applied to find the diseased area on the fruit, but it has a disadvantage of sizable estimation load. It will encourage agronomist to build better products and make the correct time to time judgment. Image processing mechanism in collaboration with machine learning techniques enhances the performance of detection [3]. In addition, work has been extended to include crop like rice. The detection of the problem within rice crop is accomplished using the technique of image processing like preprocessing, segmentation, extraction, etc. Fruit Detection using Improved Multiple Features based Algorithm. It provides an improved solution for locating the fruits on the plant based on multiple features [4]. Multiple feature extortion technique can include steps like extraction of color and intensity feature, extraction of orientation feature, extraction of edge feature, extraction of the area from feature maps. The process is entirely automatic, and it can work without user involvement. To improve output, it considers numerous features. The system classifies ripeness of fruit based on its color or texture [5]. It involves current techniques mainly manual inspection which leads to erroneous classification; it results in economic losses due to an inferior product in the market chain. The shortcomings are several methodologies which require highly expensive setups and complicated procedures; overall accuracy is achieved up to 98%. A fruit grading system based on discrete cosine transform (DCT) [6]. The framework is created to gage quality of fruits, to be specific, citrus fruits. Texture characteristics based on DCT are extricated from the surface of ordinary and influenced fruit sample images and bolstered as input to SVM and PNN (citrus) classifiers. The classification results have appeared that the SVM performed as way better discriminator compared with PNN classifier. Have illustrated a level of harm within the fruits amid collecting, transport, and controlling. The stress created in fruits driving to its harm is analyzed based on general material properties and unit stack utilizing finite element method (FEM) reenactments. Concurring to the model created, it is detailed that the maximum stress is created within the center of the fruits which results in breakage of fruits. [7]

II. Proposed Methodology

A. *The proposed framework*

- Phase1:Read input image/ Contrast Enhancement / transform image from RGB to color space
- Phase 2: Extract disease containing a segment of Orange using K-means clustering.
- Phase 3: Extract feature from the segmented image.
- Phase 4: Classification of the disease using both ANN and multi-class SVM.
- Phase 5: Detection of the various diseases

B. *ANN (Artificial Neural Network)*

The newly improved paradigms of Artificial Neural Networks have greatly contributed to the exploration, understanding and the use of possible functional similarities between human and artificial information processing systems.

Artificial Neural Networks are physically cellular systems which can obtain, store and make use of experimental knowledge. They are brain-inspired systems which are intended to replicate the way we humans learn. In general, neural networks consist of input and output layers, as well as the hidden layer consisting of units that transform input into something that the output layer can utilize. They are great tools for finding patterns which are far too complex or numerous for a human to extract and teach the machine to recognize.

There are multiple types of the neural network, each of which come with their own specific use cases and levels of complexity. The most basic type of neural net is something called a feedforward neural network, in which information travels in only one direction from input to output.

Artificial neural networks can be classified into two categories according to the allowed direction of information flow: feedforward and recurrent neural networks. A more widely used type of the network is the recurrent neural network, in which data can flow in multiple directions. These neural networks possess greater learning abilities and are widely employed for more complex tasks such as learning handwriting or language recognition. The recurrent connections have a time-delay (usually 1-time step when using discrete time), thus making the model aware of its previous inputs.

In feedforward nets, the information flows strictly forward from inputs towards the outputs.

C. *SVM (Support Vector Machine)*

Support Vector Machine (SVM) can be explained as a supervised binary classification algorithm. Presenting a set of factors of two categories in N-dimensional put SVM produces an (N−1) dimensional hyperplane to partitioned those factors into two bunches. Classifying data using a hyperplane with the largest margin between class data point, i.e., the norm of the vector that defines the separating hyperplane should be minimized. It is also generalized to linearly inseparable classes by the introduction of the slack variables. Then the loss function includes the norm of the slack variables.

SVM is still a widely used method nowadays, in particular, linear SVM. In many cases, if you have a lot of data and a lot of features, linear SVM can work very well. They can also be fast to train.

It is the name given to those inputs/samples which are right on the margin of separation. The hyperplane separates the samples as much as possible with a particular margin. Now right on the boundaries of the mar separation margin, there are some samples which are the closest to the hyperplane and these are called support vectors.

They are a type of algorithm that is often used in supervised learning. It allows your model to find a way to separate a labeled dataset, and thus to classify new un-seed data. It is one of the most used algorithms in supervised learning and is often used at the end of deep neural networks, to classify images based on the features extracted by the network

SVMs also need fewer training samples because the separating hyperplane only depends on the vectors of the class borders, the support vectors which usually also leads to lower training time.

Figure 1 below shows the step by step proposed algorithm used in order to detect the citrus fruit diseases by phenotyping. The proposed algorithm consists of Preprocessing, Segmentation, Feature Extraction, Classifiers.

D. The proposed algorithm

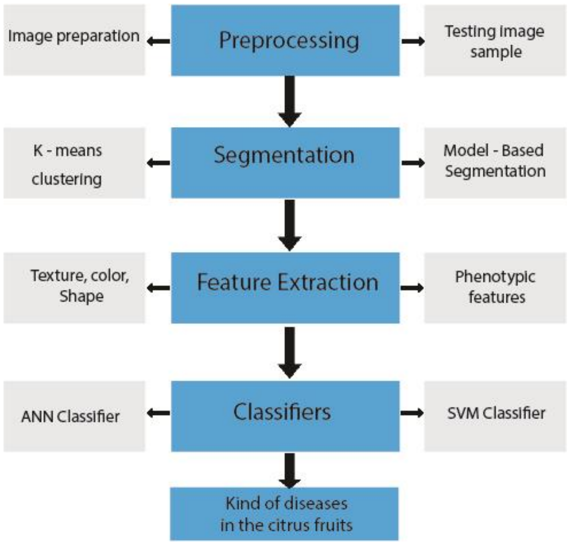


Figure 1. *proposed algorithm*

Image Preparation / Sample Testing - This initial stage is for the workflow series of image processing because as processing is possible only with the help of an image. The image obtained is entirely natural and is the consequence of any hardware which was handled to produce it.

Image Segmentation - It is the method for segregation of digital image into several segments. The primary aim of segmentation is to clarify and/or convert the rendering of an image into something that is further relevant and easier for analyses. Objects and bounding line of images are located by using image segmentation. Pixels with similar label portion share distinguishing features for allocating a label to each pixel in an image. For this, we are using K-Means Clustering methodology.

Feature Extraction - Four feature vectors are considered namely color, texture, morphology and structure of hole of the fruits. For describing huge data set sometimes enormous resources are required: multi-class SVM algorithm is applied for extracting the features.

Let $I(i,j)$ be the original RGB image having dimension $250 \times 250 \times 3$

$$s(x,y)=\sum_{i=0}^{x}\sum_{j=0}^{y} I(i,j) \qquad (1)$$

The RGB color space is affected by light and angle of the image(s) which has been captured so there is a need for conversion into HSI color space.

$$\text{Hue} = \begin{cases} \theta & if\ B < G \\ 360-\theta & if\ B > G \end{cases} \qquad (2)$$

The Hue value needs to be multiplied by 60 to convert it to degrees on the color circle

If Hue becomes negative you need to add 360, because a circle has 360 degrees.

$$\theta = \cos^{-1}\frac{1/2(R-G)+(R-G)}{(R-G)^2+(R-B)(G-B)^{1/2}} \qquad (3)$$

(4)

$$\text{Saturtion} = 1-\frac{3}{(R+G+B)}[\min(R.G.B \qquad (4)$$

$$Intensiy = \frac{1}{3}(R+G+B) \qquad (5)$$

Texture identification is done by modeling textures as the 2-D deviation of a grey level. It is expressed as;

$$(\psi)^{ab}(x) = |a|^{\frac{1}{2}}\psi(x-b/a) \qquad (6)$$

Then,

$$W = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(t)\,\psi \qquad (7)$$

Where ψ (x) is the major wavelet and $(\psi)^{ab}$(x) is the minor wavelet.

III. EXPERIMENTAL ANALYSIS

Figure 2. shows the different types of known citrus fruits diseases. The classifiers are well trained and verified by means of images of citrus fruit diseases.

Types of Citrus diseases

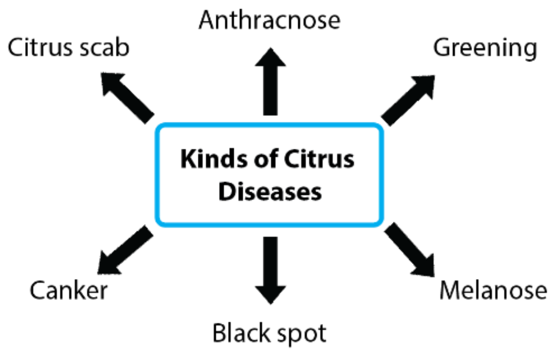


Figure 2. *Different types of citrus diseases*

A. Anthracnose

Anthracnose is an initial colonizer of damaged and elderly tissue. The plants grow on stagnant copse in the cover, and it grows short distance by rain spray, dew, and overwatering. The leaf symptoms of citrus are a more or less broadside, smooth region, and a rich brown in color with a prominent purple perimeter. Moreover, the fruit symptoms are usually damaged by other factors such as sunburn, chemical high temperature, pest damage, and expanding storage time. The color of anthracnose lesion spot is brown having 1.5mm or greater diameter.

B. Black spot

The citrus black spot also abbreviated as CBS is available at the time when the plant is sensitive and the atmosphere is favorable for disease. The symptoms of the citrus leaf and fruits are small, round, and under dangerous spots with gray squares. The color of black spot lesion is a dark brown having spot diameter of 0.12 to 0.4 in.

C. Canker

Citrus canker spot is coated by wind-driven rain. On the citrus leaf, the range of lesion is 2 to 10mm in size and put circles around the leaf. On the other hand, the range of fruit lesion will be 1 to 10mm in size and these lesions differ in size from each other. Mostly, the lesions are enclosed by water soaked and yellow halo. The color of this lesion is dark brown and black.

D. Citrus Scab

In citrus leaf and fruit, the scab acne is a composite of fungal and organism tissue. These acnes are hardly raised and pink to light brown in color. The color of scab lesion is yellow-brown and dirty grey.

E. Melanose

Melanose is a saprophyte in which the severity is defined by the number of inoculums on dead wood in the tree cover. The symptoms of the leaf appear as a small brown spot which becomes filled with a reddish-brown gum. On the other hand, the fruit symptoms depend on the age of the fruit at infection time. The symptoms appear outside on the citrus leaf and fruit. [8]

Figure 3. Shows the images of the various known citrus fruit diseases. They are causing much harm to the citrus fruits and reducing production. These diseases remain dormant while the fruit is still attached to the tree. Once the fruit is removed and put into storage, symptoms will be produced. If allowed to progress unchecked, it's trees will eventually result in dieback, fruit drop, and leaf loss

Images of citrus diseases causing low production massively to citrus fruits.

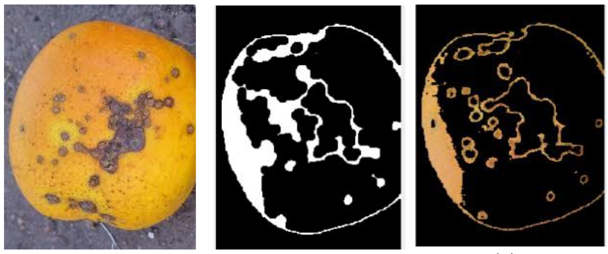
Figure 3. *Sample images of citrus diseases*

Training and testing was made on the classifiers using data set images of fruits (citrus). The images are grouped into two different categories where one group only was trained and tested. Color and texture feature were utilized in order to train and test the Artificial neural network model and the SVM..

F. Datasets used

The feature extraction which is composed of the vectors, color, texture, morphology, structure of hole (phenotypic nature) is used as the dataset images for the extraction. The color of the fruits plays a huge role in the detection, extraction, segmentation and classification process. The datasets were obtained from Kaggle datasets.

Figure 4. shows the infected, binary and clustered images of the citrus fruits. Due to severe infections there may be defoliation, twig and stem may show die-back symptoms.



(a) (b) (c)

Figure 4. *a) Infected Image b) binary image c) clustered image*

IV. Results And Discussions

The classifiers are trained and tested utilizing images of citrus fruit diseases. sample images are categorized into two parts and one part is utilized for training and used for testing. The color and texture highlights are utilized to train and test the Artificial neural network model. The rate of accuracy of recognition and classification is characterized as the proportion of accurately recognized sample images to the total number of sample images chosen.

TABLE 1. THE PERFORMANCE OF BOTH CLASSIFIERS

Detection and classification performance (%)	**Detection / classifiers**					
	ANN			**SVM**		
	C	T	SH	C	T	SH
Minimum Accuracy	73	82	86	82	86	85
Maximum Accuracy	94	97	94	93	95	97
Average Accuracy	89	90	92	87	86	82

classifier performances – C – Color, T- Texture, SH- Structure of Hole on the fruit

Both classifiers were studied and analyzed. The contrast of classification precisions of both the classifiers using reduced color, texture and structure of holes is given in Table 2 below. The Table depicts the high performance of the SVM slightly as compared to the ANN classifier.

Experimental analysis depicts that only five texture features are common in all the total sample images are important. These five features greatly play more role to the classification of plant diseases. Subsequently, five highlights have been considered as first-level feature reduction. The decrease is done based on limit and delta value. Any feature values below the limit are disposed of. The threshold is chosen based on the average of the least feature value and maximum feature value.

TABLE 2. THE PERFORMANCE EVALUATION OF BOTH CLASSIFIERS

Classifier	**TP+ rate**	**FT rate**	**Precision**	**ACA(%)**
SVM	0.922	0.914	0.948	93.12
ANN	0.869	0.942	0.867	88.96

Performance evaluation for both ANN & SVM classifiers

The performance of the two classifiers in regard to true positive rate (TPR), false positive rate (FPR), precision and average classification accuracy (ACA) using SVM and ANN classifiers with combined features is given in Table 2.

V. CONCLUSION

Recognizing the citrus infection is fundamentally the reason of the proposed strategy. The proposed methods are tested on citrus fruits infections. The proposed classifiers utilize color and texture features for classification. In spite of the changing outcomes for the various images of the citrus infections, the grouping of features features highlights has demonstrated to be compelling. The major advantage of the proposed strategy is considered a smaller number of features is sent to attain way better classification precision and to diminish computation time. The experimental results show that the proposed strategy is much way better, which can altogether back a precise detection of citrus infection in a small computational effort. It appears that SVM achieves a critical enhancement within the classification accuracy over ANN. SVM demonstrated to be a powerful tool for automatic classification of plant infections considered in the present work. There's a tremendous scope for upgrade within the classification precision

.ACKNOWLEDGMENT

This work was supported by the National Natural Scie nce Foundation of China [51975260], Jiangsu Internation al Science and Technology Cooperation Project [BZ2017 067], Jiangsu Provincial Key Research and Development Program [BE2018372], Jiangsu Natural Science Foundati on [BK20181443],Jiangsu Provincial Development and P romotion Program on New Agricultural Machinery Equip ment and New Technology（NJ2018 12), Zhenjiang City Key Research and Development Prog ram [NY2018001], Qing Lan project of Jiangsu Province and the Priority Academic Program Development (PAP) of the Jiangsu Higher Education Institutions, China..

REFERENCES

[1] J. Gómez-Sanchis, J. D. Martín-Guerrero, E. Soria-Olivas, M. Martínez-Sober, R. Magdalena-Benedito, and J. Blasco, "Detecting rottenness caused by Penicillium genus fungi in citrus fruits using machine learning techniques ☆," *Expert Systems with Applications,* vol. 39, pp. 780-785, 2012.

[2] M. Jhuria, A. Kumar, and R. Borse, "Image processing for smart farming: Detection of disease and fruit rading," in *IEEE Second International Conference on Image Information Processing*, 2014.

[3] J. P. Shah, H. B. Prajapati, and V. K. Dabhi, "A survey on detection and classification of rice plant diseases," in *IEEE International Conference on Current Trends in Advanced Computing*, 2016.

[4] H. N. Patel, R. Jain, and M. V. Joshi, "Fruit detection using improved multiple features based algorithm," *International Journal of computer applications,* vol. 13, pp. 1-5, 2011.

[5] S. R. Rupanagudi, B. S. Ranjani, P. Nagaraj, and V. G. Bhat, "A cost-effective tomato maturity grading system using image processing for farmers," in *International Conference on Contemporary Computing & Informatics*, 2015.

[6] S. A. Khoje, S. Bodhe, and A. Adsul, "Automated skin defect identification system for fruit grading based on discrete curvelet transform," *International Journal of Engineering and Technology (IJET),* vol. 5, pp. 3251-3256, 2013.

[7] L. Fenyvesi, D. Fenyvesi, and A. Csatár, "Stress analysis in fruits," *Advances in Mechanical Engineering,* vol. 5, p. 874673, 2013.

[8] M. Sharif, M. A. Khan, Z. Iqbal, M. F. Azam, M. I. U. Lali, and M. Y. Javed, "Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection," *Computers and electronics in agriculture,* vol. 150, pp. 220-234, 2018.