Mobile Application for Plant Disease Classification Based on Symptom Signatures

N. Petrellis
Computer Science and Engineering Dept.
TEI of Thessaly, Larissa, 41110
Greece
npetrellis@teilar.gr

ABSTRACT

Several intelligent systems have already been proposed for the diagnosis of plant diseases. In this way, the plants and crops can be monitored in order to prevent the spread of diseases that can ruin the whole harvest. The symptoms of a disease include lesions or spots in various parts of the plant. The color, area and the number of these spots can determine to a great extent the disease of the plant or serve as a first stage diagnosis. In this paper, a mobile phone application for plant disease diagnosis is presented. It is based on the detection of the disease signature that is expressed as a number of rules that concern the color, the shape of the spots, historical weather data, etc. The disease signature format allows an agriculturist that acts as an end user of the developed application, to extend or customize the supported set of plant diseases. The developed application has been tested with similar performance on various plants including citrus and grapevines. In this paper, experimental results are presented on grape diseases with the accuracy in the plant disease classification exceeding 90%.[1]

CCS CONCEPTS

• **Applied Computing → Computers in other domains**; *Agriculture* • **Human Centered Computing → Ubiquitous and Mobile Computing**; *Ubiquitous and mobile computing theory, concepts and paradigms; Mobile computing*; • **Computing Methodologies → Computer graphics**; *Image Processing*;

KEYWORDS

Plant disease, smart phone application, image processing, grape disease

ACM Reference format:

N. Petrellis. 2017. SIG Proceedings Paper in word Format. In *Proceedings of ACM Pan-Hellenic Conference on Informatics, Larissa, Greece, Sept. 2017 (PCI 2017)*, 6 pages.

PCI 2017, September 28–30, 2017, Larissa, Greece

ACM ISBN 978-1-4503-5355-7/17/09 $15.00
https://doi.org/10.1145/3139367.3139368

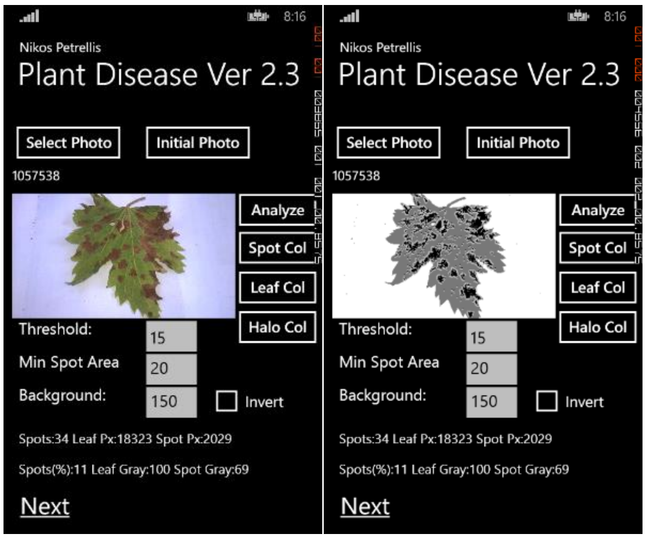
1 INTRODUCTION

The early detection of plant diseases is very important in order to avoid extended seasonal damage of the fruits, the harvest or the destruction of the whole plants and trees. The production cost can be significantly increased if plant diseases are not treated in time by detecting the first symptoms of a disease before it propagates to all the field. The continuous monitor of the plants by agriculturists is not always possible or financially efficient especially for crops or trees in isolated rural areas. Remote monitoring through sensors, drones, etc, can offer a low cost and high precision alternative option. Molecular analyses have higher cost but may be performed at a later stage if necessary in order to confirm that a plant has indeed been affected by a specific disease before a more toxic treatment is applied.

Various symptoms that are used for plant disease diagnosis are described in [1]. Biotic agents can affect the progression of the symptoms that can be classified as primary or secondary. The symptoms of a pathogen can be expressed as fungal or bacterial leaf spots, vein banding, mosaic appearance, the leaves can be distorted or a powdery mildew can appear. Air pollution or chemicals can also make harm to the plants. Molecular test like the PCR technique can be applied as described in [2] and their sensitivity depends on the amount of the microorganism that is tested.

In [3] some image processing techniques and molecular tests are reviewed. Non-destructive techniques like spectroscopic and image processing can be used for plant disease diagnosis based on its symptoms. Reviews of image processing techniques in visible light for plant disease detection can be found in [4]. Image segmentation is used along with a Gabor filter in [5] achieving disease recognition with 91% accuracy. An expert system based either on graphical representation or a step-by-step user-guided method is described in [6]. Optimal solutions for specific plants have been presented for example in [7] for corn diseases. In most of the referenced approaches expensive equipment and software packages of high complexity are required.

The use of mobile applications in plant disease diagnosis has not been widely addressed in the literature. However, some applications similar to the one presented here have been recently made available like the application Purdue Plant Doctor [8]. It is offered in different versions for each plant type for iOS or Android platforms. It is based on questionnaire and decision trees in order to diagnose a disease. More similar to our approach is the Plant Doctor offered by Plantix [9] that is based on an expert system. This expert system is trained to recognize a large number of diseases and can communicate with a remote database for more accurate decisions. Similarly in [10], two deep convolutional neural networks are evaluated (AlexNet, GoogLeNet) using large training sets (in the order of thousands of images). The accuracy in this approach is higher than 98%.



(a) (b)

Figure 1: The main screen of the Plant Disease application after selecting a grape leaf with Downy Mildew (a) and after analyzing it in order to locate the spots (b).

The plant disease classification that is implemented as a mobile phone application in this paper, is based on comparing various features extracted by the photograph of a plant part with the supported disease signatures. The plant part can be a leaf, a fruit, a branch, the root, etc. The features include the number of spots, their area, the spot color, weather conditions, etc. The signature of a disease is a set of strict and loose limits on each of these features. If the corresponding feature falls within these limits a rank is given and the weighted sum of these ranks can be used to sort the diseases and find the one that the plant is likely to have been affected by. Although this classification method is not as sophisticated as an expert system based on a neural network it offers some certain advantages: a) it requires a small training set to define credible signatures, b) it is easily extendible by the end user concerning the supported set of diseases, c) it is fast, d) it does not consume a large number of mobile phone resources since its implementation is not complicated and e) the information not taken into consideration due to the small training set is compensated by the large number of features used in the disease signatures as will be described in the next sections.

The approach presented in this paper is an extension of the application described in [11]. This previous work concerned the recognition of the plant lesion features with high precision. Here, we focus on the new set of rules that is used to define disease signatures in order to demonstrate how plant diseases are defined, examined and how the supported set of diseases can be extended. The developed application has been tested with various plants including citrus and grapevines with similar success rate in the disease recognition. Grapevine diseases are used here as a case study both to explain the proposed classification method and to measure the success rate of the disease recognition which exceeds 90% in most of the cases.

The proposed image processing technique, its implementation as a smart phone application and the disease signatures used, are described in Section 2. The experimental results are discussed in Section 3.

2 MOBILE APPLICATION FOR PLANT DISEASE CLASSIFICATION

2.1 The Mobile Application Environment of the Plant Disease Classification

The proposed plant disease classification application has been currently implemented for Windows Phone platform as shown in Fig 1a. The user selects a leaf and analyzes it trying to locate the pixels of the leaf belonging to lesions. This is currently achieved by examining the grey level of each pixel and assuming that the spot pixels have a grey level exceeding the average leaf pixel greyness by an offset *Th*. If the spots are darker than the normal leaf, their grey level should be lower than G_{av}-Th, where G_{av} is the average grey level of all the leaf pixels. If the spots are brighter than the normal leaf, it is assumed that they have a grey level higher than G_{av}+Th (this condition is examined if the "Invert" box of Fig. 1 is checked). Although in most of the cases this method can successfully isolate the spot areas, more advanced methods can easily be empoyed based on specific color features of the spots (e.g., checking the level of the red, green or blue color of the pixels instead of their grey level). In the spot isolation process, spots consisting of a very small number of pixels (lower than the number defined in the field: Min Spot Area) are assumed to be noise and can be ignored. In the specific realization, the background is assumed to be much brighter than the leaf pixels in order to avoid a complicated background separation method. More specifically, the background pixel grey level should be higher than the value B_g defined in the field "Background" of Fig. 1(B_g>G_{av}+Th).

The Red-Green-Blue (RGB) pixel matrix (size: N×M) of the original color image is used to construct a new matrix called Background-Normal-Spot (BNS) with the same dimensions (0≤i<N, 0≤j<M) as follows:

$$\mathrm{BNS}(i,j) = \begin{cases} 0, RGB(i,j) \geq B_g \\ 1, G_{av} + Th < RGB(i,j) < B_g, Invert = 1 \\ 1, G_{av} - Th > RGB(i,j), Invert = 0 \\ 2, G_{av} + Th > RGB(i,j), Invert = 1 \\ 2, G_{av} - Th < RGB(i,j) < B_g, Invert = 0 \end{cases} \quad (1)$$

As defined by equation(1), BNS(i,j) is 0, 1 and 2 for background, normal leaf and spot pixels, respectively. Fig. 1b, displays the BNS matrix: white for background, grey for normal leaf and black for the spots. The BNS matrix can be used to extract some initial useful features that are shown at the bottom of the application page of Fig. 1b. These features are: the number of spots, their area (expressed as number of pixels), the proportion of the spot area in the leaf, the average grey level of the normal leaf and the spots. This is accomplished by assigning an index to adjacent pixels that are marked as "spot" by equation(1) since they are assumed to belong to the same spot. The higher index used, indicates the total number of spots.

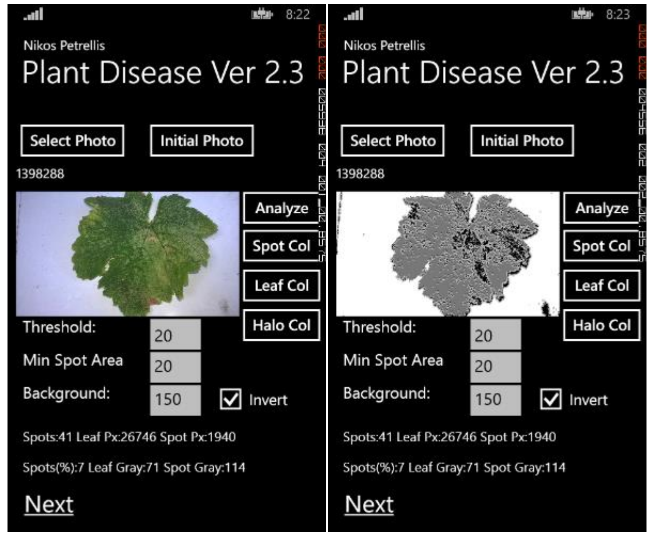
(a) (b)

Figure 2: The histograms extracted for the spot (a) and the normal leaf area (b) of the leaf shown in Fig. 1.

The BNS matrix can also be used as a map to access the pixels of the original color image that belong to the spot or the normal leaf in order to create color histograms like the ones shown in Fig. 2. The histograms for the spots or the normal leaf display how many pixels in these areas have a specific level of red, green or blue color. Each position k in the histogram for the color c (where c can be red, green or blue), is a value $H_c(k)$ formally defined as:

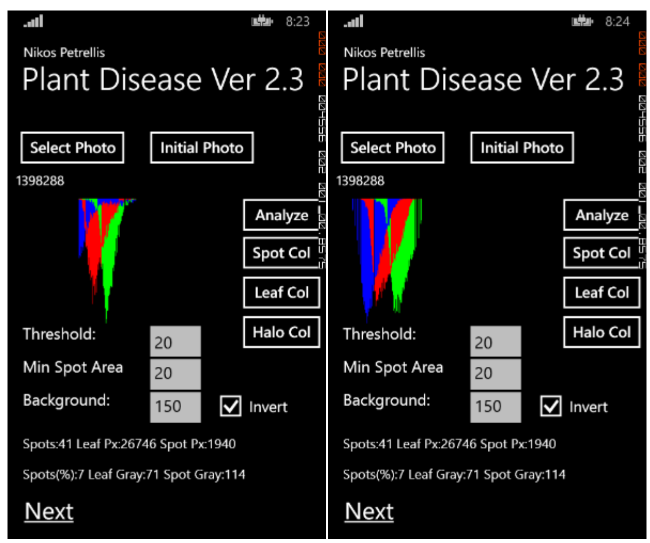
$$H_c(\mathrm{k}) = |\{RGB(i,j,c)|RGB(i,j,c) = k\}| \qquad (2)$$

RGB(i,j,c) is the level of the color c in the RGB(i,j) pixel and k is between 0 and 255 (the level of each color in RGB format). Six histograms are used, 3 for the colors of the spot area and 3 for the normal leaf. Each of the histograms has its information usually concentrated in one lobe that is shifted on the horizontal axis according to the brightness of the leaf or spot area and the lighting conditions of the environment where the photo was taken. It is also possible to have information spread outside of the lobe in a histogram, but this usually happens when the area mapped to spots is excessively wide and also covers normal leaf of different color. The start/end position of the main lobe and the position of its peak are used to define rules that are part of a disease signature. Although these positions depend on environmental lighting conditions they primarily depend on the specific disease that has affected the plant.



(a) (b)

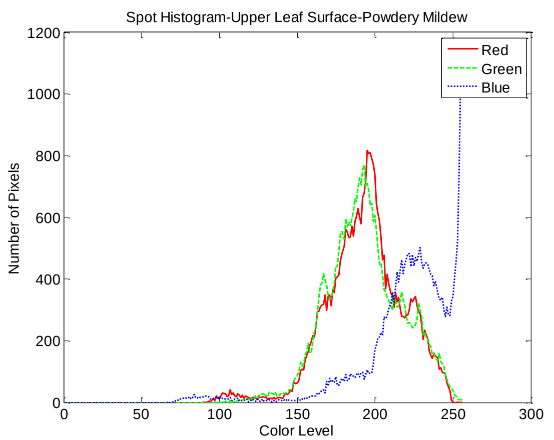
Figure 3: The main screen of the Plant Disease application after selecting a grape leaf with Powdery Mildew (a) and after analyzing it (b).



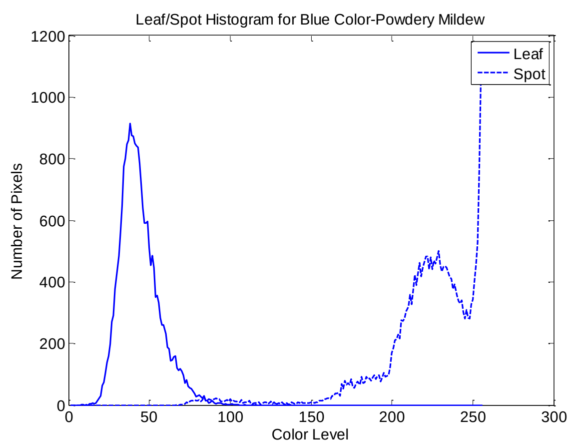
(a) (b)

Figure 4: The histograms extracted by the spot (a) and the normal leaf areas (b) of the leaf shown in Fig. 3.

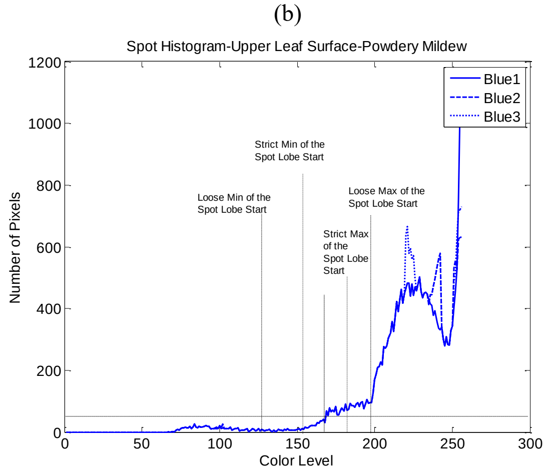
2.2 The Plant Disease Classification Method



(a)



(b)

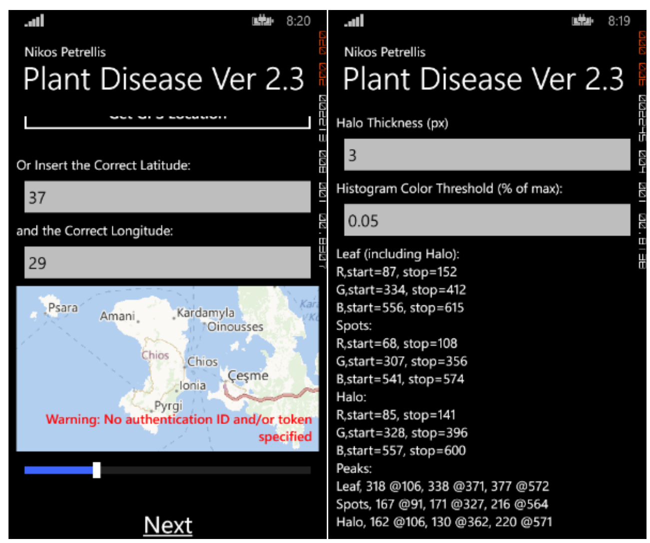


(c)

Figure 5: The histograms of the three colors of a leaf affected by Powdery Mildew (a), the blue histograms for the spots and the normal leaf area (b), and three blue histograms from different leaves (c).

The specific leaf displayed in Fig. 1 and its histograms shown in Fig. 2 is affected by Downy Mildew. Had the leaf been affected by a different disease like Powdery Mildew, different histograms would have been produced as shown in Fig. 3 and Fig. 4. The obvious visual differences shown especially by the histograms of the spot areas are used to classify the plan diseases.

We will examine in more detail the features that can be extracted by the color histograms like the ones shown in Fig. 2 and Fig. 4. For this reason, in Fig. 5, similar histograms have been plotted in MATLAB in order to view more clearly the details of the classification method. In Fig. 5a, the three color histograms of the spots have been drawn while in Fig 5b, the blue histogram of both the normal leaf and the spots area are drawn together showing the different position of the lobes. In Fig. 5c, three blue histograms are shown that are extracted from different leaves and all suffer from Powdery Mildew.



(a) (b)

Figure 6: GPS localization for weather historical data retrieval (a) and the Feature extraction from a specific photo like the ones shown in Fig. 1a and Fig. 3a.

As can be seen from Fig. 5c, the lobes of the three blue histograms are quite similar concerning their start point, their peaks and their ending point. These three positions are features are determined using this simple statistical processing based on just three leaves with the same disease. Using as a start point of the lobe e.g., the 5% of the lobe peak in the histogram it can be seen that it crosses this threshold at approximately the position 165 of the horizontal axis. The observer of these lobes can arbitrarily assume that this starting point may reside in the relatively narrow region: (150..180) (strict limits) or in the worst case in a relatively wider region: (130..200) (loose limits). The loose limits are delimited by the dash-dot vertical lines and the strict limits are delimited by the dot vertical lines in Fig. 5c. In a similar manner the strict and loose limits of the lobe peak or its ending can be defined. If a measured feature is within the strict limits of a disease, this disease gets an additive rank higher than the rank it would get if the feature was only within the loose limits. No rank is given for this feature if the measured value is outside even the loose limits.

The specific features F(l) used to rank a plant disease are the following: number of spots (l=0), proportion of spot area (l=1), spots' grey level (l=2), leaf grey level (l=3), lobe start of the spot color historgram (l=4 (red), 5 (green), 6 (blue)), lobe peak of the spot color histogram (l=7..9 for the three RGB colors), lobe end of the spot color histogram (l=10..12), normal leaf color histogram - lobe start (l=12..15), normal leaf color histogram - lobe peak (l=16..18), normal leaf color histogram - lobe end (l=19..21). For each one of these F(l) features four numbers are defined that correspond to the loose and the strict limits for a disease *d,* as was shown for F(6) in Fig. 5c. These limits are symbolically denoted as $F_{loose,min}(l,d)$, $F_{loose,max}(l,d)$, $F_{strict,min}(l,d)$, $F_{strict,max}(l,d)$. Two more features are used regarding the average temperature (l=22 for minimum, l=23 for maximum daily temperature) and humidity conditions (l=24) in the region where the plant resides. The mobile phone application extracts these features by averaging the corresponding temperatures and humidity values from a number of representative past dates selected by the user. These data are retrieved by a remote weather webpage based on the Global Positioning System (GPS) localization incorporated in the developed mobile phone application (Fig. 6a).

If the rank of a plant disease *d* is R(d), it is estimated as:

$$\mathrm{R(d)} = \sum_{l=0}^{24} b_{ls,l}(d) W_{ls,l}(d) + \sum_{l=0}^{24} b_{st,l}(d) W_{st,l}(d) \tag{3}$$

The weights $W_{ls,l}(d)$ and $W_{st,l}(d)$ are the ranks for the feature F(l) if the comparison with the loose and the strict limits, respectively, is successful. The binary flags $b_{ls,l}(d)$ and $b_{st,l}(d)$ are true (1) if the measured feature F(l) is within the loose and strict limits defined for the specific plant disease *d*, respectively:

$$b_{ls,l}(d) = \begin{cases} 1, F_{loose,min}(l,d) \le F(l) \le F_{loose,max}(l,d) \\ 0, otherwise \end{cases} \tag{4}$$

$$b_{st,l}(d) = \begin{cases} 1, F_{strict,min}(l,d) \le F(l) \le F_{strict,max}(l,d) \\ 0, otherwise \end{cases} \tag{5}$$

All the measured F(l) features described earlier are shown to the user by the application. For example, in Fig. 6b the F(l) features that concern the start/peak/end of all the color histogram lobes are listed (l=4..21). Using these indications from a small number of plant photos, an end user like an agriculturist can define his own $F_{loose,min}(l,d)$, $F_{loose,max}(l,d)$, $F_{strict,min}(l,d)$, $F_{strict,max}(l,d)$ limits defining a new disease signature and extending the supported set of recognizable diseases. These limits can be stored in a standard format like eXtensible Markup Language (XML) or JavaScript Object Notation (JSON) ensuring the expandability of the developed application. Of course, additional information has also to be set for a new disease like its name and actions that have to be taken by the user to protect his harvest and plants if this disease is recognized.

3 RESULTS AND DISCUSSION

Grape leaves were used for the evaluation of the developed application. The leaves of the test bench were affected by Downy Mildew, Powdery Mildew, Esca, Pierce's Disease. Testing healthy leaves was meaningless because they have no lesions or spots to focus on. A small training set of 5 photographs was used for the definition of the *F* feature limit values described in the previous section (disease signature). Then, a benchmark of 80 photographs were tested, 20 of each disease. The success rate of the disease classification is shown in Table 1. The first column lists the diseases tested and the second column lists the success rate in the disease recognition with a *Th* spot threshold value being equal to the one used during the training. The third column is the success rate achieved when the *Th* is 10% higher or 10% lower than the one used during the training.

As can be seen from Table 1 the success rate is higher than 90% if the threshold *Th* used is the same with the one used during the training period. However, the success rate does not get much worse if a threshold is used that deviates by up to ±10% the optimal value. Even if the correct disease is not recognized as the dominant one by the ranking of equation(3), it is sorted as the second possible disease in most of the cases. The tolerance in the selection of an appropriate value for *Th* is important since the user will arbitrary set this value. Nevertheless, the developed application offers an interactive way to help the user in selecting an appropriate value for *Th* since he can press the button "Analyze" in Fig. 1 or Fig. 3 and observe by the grey images of Fig. 1b or Fig. 3b if the spots are precisely mapped using a specific *Th* value.

Table 1: Grape Disease Classification Results

Disease	Success Rate	Success Rate (±10% different Th)
Downy Mildew	18/20	14/20
Powdery Mildew	19/20	14/20
Pierce	18/20	17/20
Esca	19/20	18/20

Similar success rate results have also been achieved by testing citrus leaves for diseases like Alternaria, Anthracnose, Melanose, Nutrient Deficiency, etc. More experiments are underway with different plants, plant parts like fruits and diseases. A difficulty in performing these tests is the fact that the diseases appear in specific seasons and stages of plant growth, thus the tests cannot be performed any time. However, the experiments already carried out show promising results since the disease recognition success rate in referenced approaches like [5] are similar (90%).

4 CONCLUSIONS

A smart phone application for plant disease recognition has been developed and presented in this paper. Photographs of the plants are used to extract specific features like number of spots, their area, position of the histogram lobes, weather data etc. The loose and strict limits of these features are used to define disease signatures. The supported set of diseases can easily be extended since these signatures can also be defined by end users like agriculturists if the information provided by the application for each plant photo is exploited.

Future work will focus on the evaluation of the disease signatures for a large number of plants and diseases, and the porting of the application to other mobile platforms like Android, iOS, etc.

ACKNOWLEDGMENTS

This work is protected by the provisional patent 1008484, published by the Greek Patent Office, May 12, 2015.

REFERENCES

[1] Riley, M.B., Williamson, M.R., and Maloy, O. 2002. Plant disease diagnosis. The Plant Health Instructor. doi= 10.1094/PHI-I-2002-1021-01

[2] Schaad N.W. and Frederick, R.D. 2002. Real time PCR and its application for rapid plant disease diagnostics. *Canadian Journal of Plant Pathology* 24, 3 (July 2002), 250-258.

[3] Sankaran, S., Mishra, A., Eshani, R. and Davis, C. 2010. A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture* 72, 1 (June 2010), 1-13.

[4] Barbedo, G.C.A. 2013. Digital image processing techniques for detecting quantifying and classifying plant diseases. *Springer Plus*, 2:660 (Dec. 2013).

[5] Kulkarni, A. and Patil, A. 2012. Applying image processing technique to detect plant diseases. *International Journal of Modern Engineering Research* 2, 5 (Oct 2012), 3361-3364

[6] Abu-Naser, S.S., Kashkash, K.A. and Fayyad, M. 2008. Developing an Expert System for Plant Disease Diagnosis. Asian Network for Scientific Information. *Journal of Artificial Intelligence* 1, 2 (2008) 78-85.

[7] Lai, J.C., Ming, B., Li, S.K., Wang, K.R., Xie, R.Z. and Gao, S.J. 2010. An Image-Based Diagnostic Expert System for Corn Diseases. *Elsevier Agricultural Sciences in China* 9, 8 (Aug. 2010) 1221-1229.

[8] https://www.purdueplantdoctor.com/

[9] http://plantix.net/

[10] Mohanty, S., Hughes, D., and Salathé, M. 2016. Using deep learning for image-based plant disease detection. Frontiers in plant science 7. doi= 10.3389/fpls.2016.01419

[11] Petrellis, N. 2016. Plant Lesion Characterization for Disease Recognition: A Windows Phone Application. Proc of the 2nd International Conference on Frontiers of Signal Processing, Oct. 15-17, 2016, Warsaw, Poland, 10-14.