

Plant Leaf Disease Identification Supported by Image Segmentation, Feature Extraction and Ensemble Classification

Navneet Kaur¹, V Devendran², Sahil Verma^{3*}
Lovely Professional University.
navneetphul@gmail.com¹, devendran.22735@lpu.co.in²,
sahilkv4010@yahoo.co.in^{3*}
*Corresponding author: sahilkv4010@yahoo.co.in

Abstract

Plant health remains a vital part of agricultural science. Production quality severely falls in such scenarios. Most often, the disease makes the leaves as their first victim as these are one of the delicate parts of the plant. An attempt has been made to create a system that identifies diseases with least human intervention. Computer vision does it all within seconds. This paper offers an insight into how different image processing techniques that comprise segmenting the image, extracting features and classifying can be employed for the aim identification of diseases in plant leaves. Several state-of-the-art techniques have been analyzed and hence problem formulation has been put forward and methodology has been proposed. In methodology, k means segmentation, law's textural measures and ensemble classifiers were focused upon. For extracting better and more features, Laws' textural mask will be applied. The strong classifier method can be used for the detection of plant disease detection. Ensemble based classifiers are the combination of two or more classifier. Plant Village dataset is focused to be considered. The paper will assist the investigators to grasp concepts of numerous techniques for leaf disease identification and an insight into the problem and future opportunities in the stated field.

Keywords –Leaf disease, image processing, segmentation, feature extraction, classification, K-means, ensemble classifier

1. Introduction

Quality is an influential factor in agricultural yield. For many countries, such as India, agricultural production determines the economy. Hence, scientists come forward to keep a check on the yield quality. These diseases need to be detected during their initial stages so as not to adversely affect the productiveness. Earlier, examination with bare eye was much prevalent among the researchers and the scientists. A large number of personnel were required for this purpose that also involved the cost. There is even a likeliness of predicting the disease wrongly. Hence, systems were designed in order to automate the task of identifying diseases in plant leaves. The use of diseased leaves images play the major role. The study consists of different modernized techniques for the intended task. The paper comprises different sections. Section 2 focuses on leaf disease detection system along with its modules. Section 3 presents the related work. Section 4 presents

problem formulation and methodology. Section 5 presents conclusion along with future scope of the work.

2. Leaf disease detection system

It comprises phases if two– (1) training, (2) testing. The first phase of a leaf disease detection system comprises several modules as depicted in Fig. 1. As from the figure, it is very clear that training phase starts with the process of image acquisition, in other words, acquiring or capturing image by employing a digital camera or other sources. Further, pre-processing techniques can be applied such as color space conversions, filters, changing brightness and contrast, noise minimization. Segmentation splits the region of interest (infected areas) from the other parts of image. And ultimately, the features are extracted from the region of interest for the purpose of training the classifier for the identification of disease. The testing phase involves passing of the diseased leaf image through these modules. The classifier (which is already trained) identifies the correct disease. Accuracy serves a very important performance measure to determine the appositeness of these modules. Accuracy is dependent on the number of diseases correctly identified. Fig. 1 depicts different modules of leaf disease detection system.

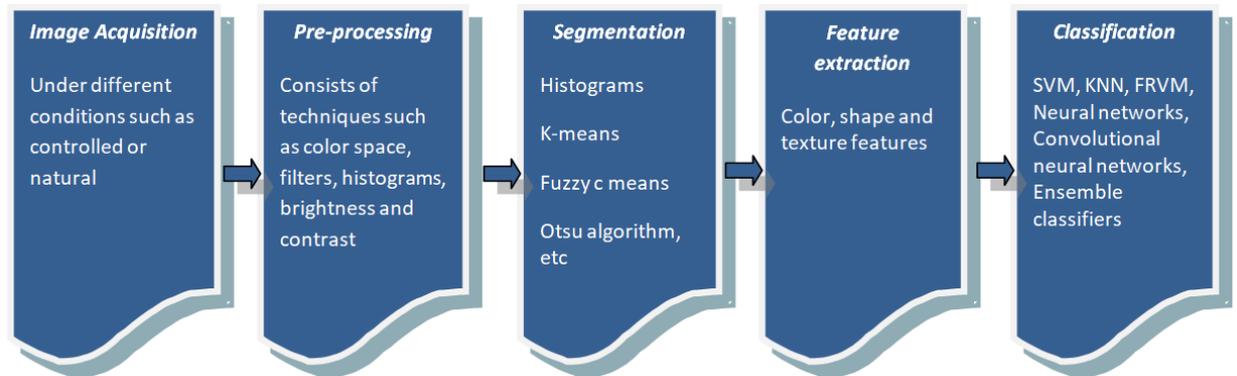


Fig. 1 Different modules of leaf disease detection system

2.1 Image Acquisition

The images used for training plays an important role as the accuracy largely depends on the kinds of image samples obtained. Many of the researchers have opted for online datasets such as Different datasets include IPM images[15], PlantVillage dataset[16],[22],[23],[32],[35] and forestry images[32]. Also the quality depends on the camera type and its position with respect to the leaf for which the image is intended to be captured. Digital cameras[4],[12],[26] and mobile phones[4],[5],[6],[13],[26],[32] have been used in various papers. Also whether the image has been captured under controlled environment or natural environment also determines the classification accuracy of the system. Hyper

spectral images were used for detecting apple leaf diseases [33]. Digital imaging and spectroscopy give rise to hyper spectral imaging.

2.2 Pre-processing

Pre-processing is recommended in order to make further processing on the images as easy as possible. Mainstream pre-processing techniques comprise color space conversions, filters, changing brightness and contrast, cropping and noise minimization, etc. Color spaces include HueSaturationIntensity(HSI), RedGreenBlue(RGB), and $L^*a^*b^*$ (CIELAB). HSI[7] splits chroma from luma. CIELAB [35] is near to human sensation. After the conversion is performed, filters and other enhancements such as brightness and contrast can be applied. Examples of filters comprise Gabor Wavelet transform[9] as the band pass filters and discrete wavelet transform[35], etc. Subsequently, cropping can be applied in order to exclude the complicated backgrounds and noise reduction can also be applied to deal with varying brightness and color information.

2.3 Segmentation

Segmentation, as already discussed splits the ROI from the rest of the image, needs to be very effective so as to carry out the feature extraction process over the ROI to achieve the correct result which will finally improve the classification accuracy. Popular methods are threshold based segmentation, K-means, Fuzzy c means(FCM), Otsu algorithm, etc. Threshold based segmentation[21] has been used to detect crops based on vein patterns. For illustrating color information, color histograms[13] are used. K means clustering[34] reduces the objective function. FCM[28] is based on fuzzy association of the pixel. Otsu's segmentation[8] is based on transformation of a grayscale image to binary image

2.4 Feature Extraction

Color, texture and shape constitute the broad categories of mainstream features. PHOG (Pyramid Histogram of Gradients) [34] and Color co-occurrence method [27] are used as methods for extracting color and texture features. Color features comprise color histograms. Depicting color distributions is the primary purpose of color histograms. Comprehensive color feature (CCF) [24], [32], [4] has also been used in multiple papers. A CCF consists of multiple color components. Features such as correlation, homogeneity, entropy can be considered as that of the texture[6]. Grey level color co-occurrence method (GLCM) [15], [4] is for spatial association among the pixel combination. Law's textural Mask focuses on edge, level, ripple and waves and an attempt has been made to improve the mask descriptor [25]. A descriptor known as Local Binary Pattern(LBP) [37] has been used to identify diseases in various crop types. The denseness of the symptoms was able to be incorporated by LBP histograms in order to present texture details.

2.5 Classification

The ultimate module in leaf disease detection system is the classifier or classification module. Discriminating the diseased and non-diseased parts is a must task for the classifier. SVM, KNN, FRVM, DCNN, ensemble classifiers are some commonly used classifiers. SVM[3], [27], [30] splits the data points using a hyperplane. K nearest neighbors are determined in KNN[9]. For automatic learning, convolutional neural networks[29], [36]are required. Ensemble based classifiers[20],[30],[31] are the combination of two or more classifier. It is a meta-classifier which is based on majority or plurality voting so as to predict the final class label.

3. Related Work

Pertot I, Kuflik T, Gordon I, Freeman S, Elad Y [2] The paper is based on a case study on a web based tool to identify disease visually for strawberry. The tool is for those users who don't have the specialized knowledge of plant diseases. It's based on description on the basis of text or images. The observer can compare the symptoms of the plant with those suggested by the tool. The tool then reverts with the information about the most possible disease. The objective is to support non-experts for disease identification.

Grinblat GL, Uzal LC, Larese MG, Granitto PM [11] adduced that using deep CNNs, recognition of plants can be done from their vein patterns. It improves the accuracy of the pipeline. Training the neural network models that consist of various processing layers refers to deep learning. The basic methodology in this papers includes steps as image acquisition, vein segmentation, central patch extraction(cropping the image), vein measures(feature extraction), use of SVM, PDA, RF for final classification. The models were trained from 2 to 6 layers. The total accuracy was offered with 5 CNN layers(white bean, red bean, soybean)

Zhang S, Wang Z. [18] adduced a method to identify four cucumber leaf diseases on the basis of global-local singular values. Segmentation of leaf spot is done with the help of watershed algorithm. Then, extraction of global-local singular values and construction of key-point vectors is performed. Ultimately, SVM model is used to classify. The recognition rate was found to be the highest in case of global-local SVD as compared to IP and LS+NN. The main drawback of the proposed method is its computational complexity. As a future work, large database can be considered.

Qin F, Liu D, Sun B, Ruan L, Ma Z, Wan H [10] adduced a technique to identify four types of Alfalfa disease that uses K-median clustering and Linear Discriminant Analysis(LDA) to segment lesion. Extraction of 129 features was

done with the help of ReliefF, 1R. Then RF, SVM, and K-NN were used to build the recognition mode. The highest accuracy was offered by SVM model that was 97.64 with training set and 94.74% with the testing set. Top 45 features were selected for building the model based on SVM

Zhang M, Meng Q [1] proposed an approach to uncover dangerous disease (in citrus plants) known as citrus canker. The paper is concentrated on the mentioned disease uncovering from images of citrus leaves which are captured in real time from the fields. Segmentation of citrus lesions was performed with the help of AdaBoost algorithm so as to achieve the significant features to further achieve a global feature vector. Then, division of the leaf image was done into various zones in order to achieve the lesion descriptor. The proposed technique achieved higher classification accuracy (just as a human expert) than other techniques.

Schor N, Bechar A, Ignat T, Dombrovsky A, Elad Y, Berman S [14] adduced a robotic mechanism to identify Powdery Mildew and Tomato Spotted Wilt Virus in greenhouses. The mechanism is based on numerous identification poses. The dataset was taken from a nursery (Hishtil, Ashkelon, Israel). Out of 36 plants, 24 were used for PM identification and the rest were used for TSWV identification). PCA and Coefficient of variation(CV) were used. Best classification accuracy was offered by PCA.

Dey AK, Sharma M, Meshram MR [8] proposed an approach has been adduced to detect betel vine leaf rot. The approach is based on Otsu thresholding for segmentation. The approach was implemented on twelve leaf images. The methodology is very helpful for precision agriculture. The methodology comprise image acquisition, image preprocessing, segmentation, and finally rotten leaf area calculation. The transformation of the images was done from RGB to HSV and hence Otsu method was applied on 'H'.

Zhou R, Kaneko S, Tanaka F, Kayamori M, Shimizu M [3] adduced a technique that uses template matching OCM with the help of which disease can be identified in sugar beet and hence efficient site specific observations based on diseases can be done. In OCM, orientation codes are matched between two images which are computed with quantized gradient angle. OCM is robust to changes in illumination as there is constant phase angle in gradient operations. Finally, to classify the disease according to pixels, SVM classifier (trained using 2D xy color histogram) has been used. Results prove that the proposed technique is robust.

Zhang S, Wu X, You Z, Zhang L [26] adduced a technique to identify cucumber disease. The technique uses three pipelines, k-means segmentation (lesion segmentation) of images of leaves with disease, color and shape feature extraction, and finally classifying images with the help of SR(Sparse Representation). Color,

shape, color histogram and lesion shape form the basis of feature extraction. The dataset images were collected from Northwest A&F University. Comparison was done with four features extraction techniques – K-means based segmentation followed neural network(KMSSN), texture feature(TF), plant leaf image based classification(PLI). Overall comparison rate was more than 85.7%.

Ma J, Du K, Zheng F, Zhang L, Gong Z, Sun Z [32] adduced a technique to identify cucumber diseases on the basis of convolution neural network. The symptom images that were considered were of anthracnose, downy mildew, powdery mildew, and target leaf spots. Data augmentation methods such as rotation and flip were applied. Segmentation of the diseases leaf images by combining CCF. CCF consists of three color components, Excess Red Index (ExR), H component of HSV and b^* component of $L^*a^*b^*$ that differentiates disordered background and spots of the disease. The technique was compared with existing classifiers such as RF(Random Forests), SVM(Support Vector Machines), and AlexNet. Best performance was obtained with downy mildew. 96.7 % were correct out of 1100 predictions of downy mildew, 98.2 % were correct out of 1006 predictions of powdery mildew.

Park K, Hong Y, Kim G, Lee J [33] adduced a technique using which to minimize the number of bands and choose only the most important ones from the hyper spectral images. Principle Component Analysis (PCA) is not that useful for this purpose. The technique proposed is named as mRMR(minimum redundancy maximum relevance). Further, to classify the obtained hyper spectral data after implementing mRMR, a deep neural network has been used which comprises CNN and FCN. For feature extraction, CNN is used and for feature extraction, FCN is used. From the experimental results, it was found that the most important five spectral bands provided more accurate results as compared to classification of an RGB image. Hence, it reduces the complexity to classify hyper spectral images. It implies DNN gives a very good performance in terms of classification of leaf conditions both with RGB and hyper spectral images(having only essential bands).

Abobakr A, Hossny M, Nahavandi S [30] adduced a method to detect free-fall based on recognizing posture. It uses an ensemble of classifiers-Random Decision Forest(RDF) and Support Vector Machines(SVM). RDF is for the purpose of posture reorganization and SVM is for detecting falls occurrence. The sensitivity and specificity rates were found to be more than 95% in cases of both synthetic and live datasets.

Wang H, Zheng B, Yoon SW, Ko HS [31] adduced an ensemble algorithm based breast cancer examination. Twelve distinct SVMs were combined for this task.

The classification rate was found to excellent as compared to state-of-the-art classifiers.

A comparison of several approaches has been presented in Table 1 and Table 2. Fig. 2 depicts evolution of state-of-the-art techniques in the recent few years

Table 1 Comparison among different state-of-the-art techniques used in leaf disease identification

Paper Title	Technique	Year	Dataset Source	Findings
Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning [22]	DCNN	2017	Apple black rot images from PlantVillage	VGG16 model was found to be the best offering an accuracy of 90.4%
A Deep Learning-based Approach for Banana Leaf Diseases Classification [23]	LeNet architecture, HIS and L*a*b*, OSTU and K means	2017	PlantVillage	The accuracy using a color image was found to be 98.61% while in grayscale it was found to be 94.44%
A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network [32]	CCF feature extraction and CNN to classify	2018	PlantVillage, https://plantvillage.org/ Forestry Images, https://www.forestryimages.org/	Best performance was obtained with downy mildew - 96.7 %
Semi-automatic leaf disease detection and	GLCM with K means as the classifier	2018	PlantVillage, https://plantvillage.org/	The average accuracy for classification is ~90%

**classification
 system for
 soybean
 culture [35]**

Table 2 Comparison among different state-of-the-art techniques used in leaf disease identification

Paper Title	Technique	Year	Dataset Source	Findings
Multi-resolution Laws' Masks based texture classification [25]	Combination of dyadic wavelet transform and Laws' masks, K-NN classifier	2017	Brodatz and VisTex	Comparison was done on the basis of Mean, Absolute mean and standard deviation.
An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification [20]	Ensemble of CNN architectures(AI exNet and GoogleNet)	2017	ImageCLEF 2016	Results were found to be better as compared to existing techniques in terms of precision, sensitivity, specificity and F-score
A Skeleton-Free Fall Detection System From Depth Images Using Random Decision Forest [30]	SVM, RDF (ensemble of decision tree predictors)	2018	URFD dataset	Sensitivity – 99% (synthetic and live datasets) Specificity – 99% (synthetic datasets) and 96% (live datasets)
A support vector machine-based	Combined twelve different SVMs for breast cancer diagnosis	2018	WBC, WDBS, and SEER datasets	97.68% accuracy rate as compared to existing classifiers

**ensemble
algorithm for
breast cancer
diagnosis [31]**

4. Problem formulation and proposed methodology

For uncovering diseases found in input plant leaf, the properties of the image can be examined in the form of texture and color. In the form of red, green and blue color, the color strength of the input can be represented by the color properties of the image. The color features of the pixels bunch were characterized by the textual characteristics of the image. For examining the texture characteristics, GLCM algorithm[35] was used while for the classification of diseases SVM classifier was used. A limitation was present in the basic GLCM algorithm that it did not create the co-occurrence matrix in a well-organized manner because of which the texture characteristics were not examined precisely. The GLCM algorithm extracts various features like energy, entropy, correlation, homogeneity, etc. The plant disease classify can be improved when the more number of features can be extracted for the detected and also strong classifier can be used for the classification. For extracting more number of features Laws' textural mask [25] will be applied. The strong classifier method can be used for the detection of plant disease detection. Ensemble based classifiers are the combination of two or more classifier. It is a meta-classifier which is based on majority or plurality voting so as to predict the final class label. According to the state-of-the arts investigation, ensemble classifiers have only been used for the purpose of medical diagnosis. An attempt will be made in order to use a combination of best classifiers as an ensemble classifier for leaf disease identification. Fig. 2 refers to the proposed methodology.

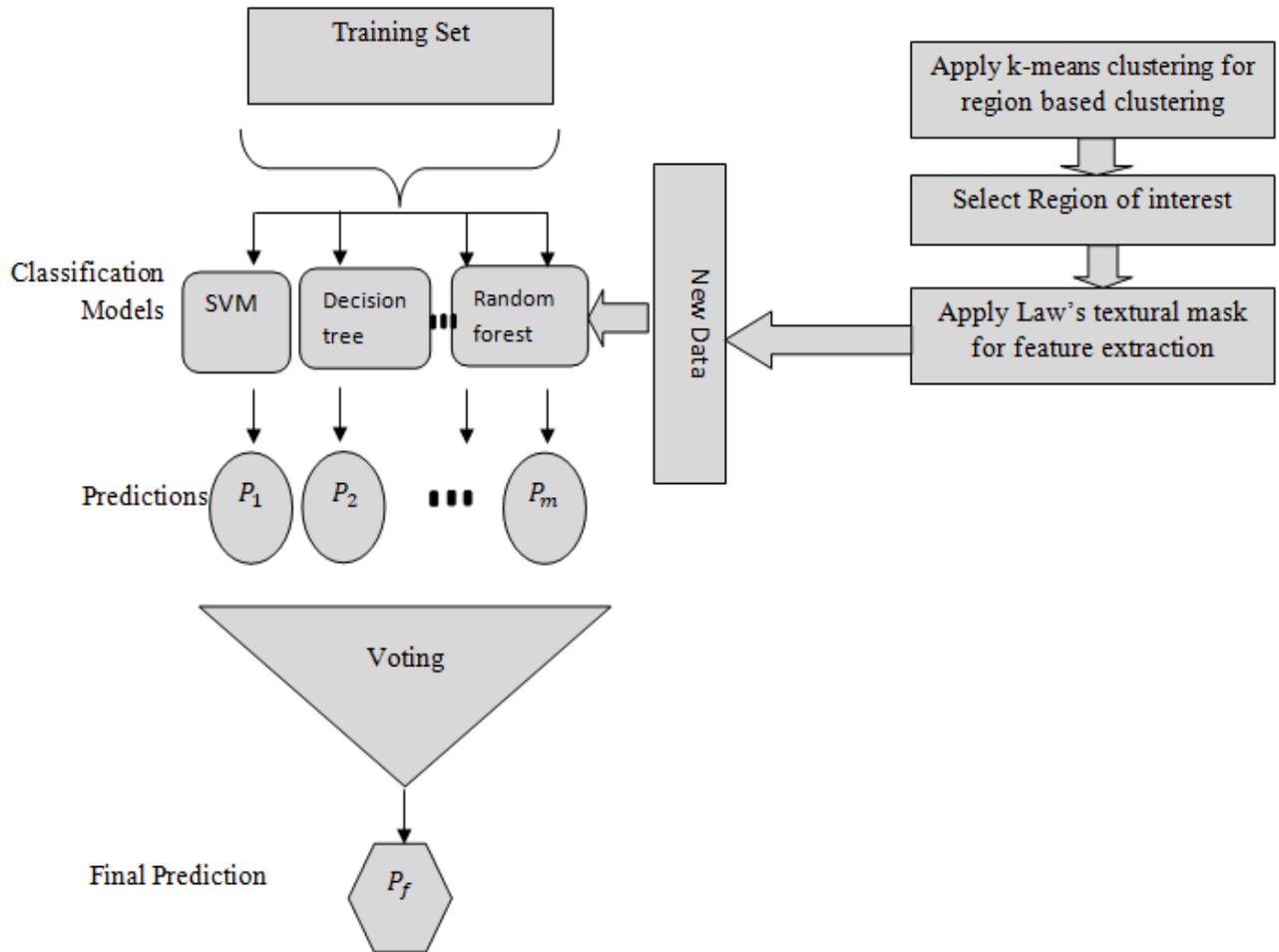


Fig. 2 Methodology

5. Conclusion

Summarization of different avant-garde techniques has been presented along with problem formulation and proposed methodology. Since, an analysis has been presented, focus was more on extracting more number of features from the diseased leaves images. Law's textural mask will be used for this purpose. A good collection of extracted features will lead to better classification results. As of now, Law's textural mask feature extraction approach has been chosen to extract feature set. This approach has never been used for feature extraction in leaf disease diagnosis. Better classifiers will be chosen to be combined into an ensemble classifier to obtain maximum classification accuracy. It will be interesting to know the combination of classifiers. As a matter of future opportunities in this work, the feature set that will be extracted can be extended. More feature extraction approaches need to be explored. And also, more classifiers will need to be explored to make an ensemble of it. Classifiers such as SVM, random forests and decision trees can be take into consideration. Hence, ensemble learning is going to be one of the most important aspects of the proposed methodology for implementation.

REFERENCES

1. Zhang M, Meng Q. Automatic citrus canker detection from leaf images captured in field. 2011; 32(15): 2036-2046.
2. Pertot I, Kuflik T, Gordon I, Freeman S, Elad Y. Identifier A web-based tool for visual plant disease identification a proof of concept with a case study on strawberry. *Computers and Electronics in Agriculture*. 2012; 84: 144-154.
3. Zhou R, Kaneko S, Tanaka F, Kayamori M, Shimizu M. Disease detection of Cercospora Leaf Spot in sugar beet by robust template matching. *Computers and Electronics in Agriculture*. 2014; 108: 58-70.
4. Prasad S, Peddoju SK, Ghosh D. Multi-resolution mobile vision system for plant leaf disease diagnosis. *Signal Image and Video Processing*. 2015.
5. Shrivastava S, Singh SK, Hooda DS. Color sensing and image processing-based automatic soybean plant foliar disease severity detection and estimation. *Multimedia Tools and Applications*. 2015; 74(24): 11467-11484.
6. Zhang SW, Shang YJ, Wang L. Plant disease recognition based on plant leaf image. *Journal of Animal and Plant Sciences*. 2015; 25: 42-45.
7. Kazmi W, Garcia-Ruiz FJ, Nielsen J, Rasmussen J, Andersen HJ. Detecting creeping thistle in sugar beet fields using vegetation indices. *Computers and Electronics in Agriculture*. 2015; 112: 10-19.
8. Dey AK, Sharma M, Meshram MR. Image Processing Based Leaf Rot Disease. Detection of Betel Vine (*Piper BetleL.*). *Procedia Computer Science*. Elsevier. 2016; 85: 748-754.
9. VijayaLakshmi B, Mohan V. Kernel-based PSO and FRVM An automatic plant leaf type detection using texture shape and color features. *Computers and Electronics in Agriculture*. 2016; 125: 99-112.
10. Qin F, Liu D, Sun B, Ruan L, Ma Z, Wan H. Identification of Alfalfa Leaf Diseases Using Image Recognition Technology. *Plos One*. 2016.
11. Grinblat GL, Uzal LC, Larese MG, Granitto PM. Deep learning for plant identification using vein morphological patterns. *Computers and Electronics in Agriculture*. 2016; 127: 418-424.
12. Barbedo JGA. A novel algorithm for semi-automatic segmentation of plant leaf disease symptoms using digital image processing. *Tropical Plant Pathology*. 2016; 41(4): 210-224.
13. Barbedo JGA, Koenigkan LV, Santos TT. Identifying multiple plant diseases using digital image processing. *Biosystems Engineering*. 2016; 147: 104-116.
14. Schor N, Bechar A, Ignat T, Dombrovsky A, Elad Y, Berman S. Robotic Disease Detection in Greenhouses: Combined Detection of Powdery Mildew and Tomato Spotted Wilt Virus. *IEEE Robotics and Automation Letters*. 2016; 1(1): 354-360.
15. Gharge S, Singh P. Image Processing for Soybean Disease Classification and Severity Estimation. 2016: 493-500.

16. Mohanty SP, Hughes DP, Salathe M. Using Deep Learning for Image-Based Plant Disease Detection. *Frontiers in Plant Science*. 2016.
17. Sladojevic S, Arsenovic M, Anderla A, Culibrk D, Stefanovic D. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. *Computational Intelligence and Neuroscience*. 2016.
18. Zhang S, Wang Z. Cucumber Disease Recognition based on Global-Local Singular Value Decomposition. *Neurocomputing*. 2016.
19. Johannes A, Picon A, Gila AA, Echazarra J, Vaamonde SR, Navajas AD, Barredo AO, et al. Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. *Computers and Electronics in Agriculture*. 2017; 138: 200-209.
20. Kumar A, Kim J, Lyndon D, Fulham M, Feng D. An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification. *IEEE Journal of Biomedical and Health Informatics*. 2017; 21(1): 31-40.
21. Hamuda E, Ginley BM, Glavin M, Jones E. Automatic crop detection under field conditions using the HSV colour space and morphological operations. *Computers and Electronics in Agriculture*. 2017; 133: 97-107.
22. Wang G, Sun Y, Wang J. Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning. *Computational Intelligence and Neuroscience*. 2017.
23. Amara J, Bouaziz B, Algergawy A. A Deep Learning-based Approach for Banana Leaf Diseases Classification. pdfs.semanticscholar.org. 2017.
24. Ma J, Du K, Zhang L, Zheng F, Chu J, Sun Z. A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing. *Computers and Electronics in Agriculture*. 2017; 142:110-117.
25. Dash S, Jena UR. Multi-resolution Laws' Masks based texture classification. *Journal of Applied Research and Technology*. 2017; 15(6): 571-582.
26. Zhang S, Wu X, You Z, Zhang L. Leaf image based cucumber disease recognition using sparse representation classification. *Computers and Electronics in Agriculture*. 2017; 134: 135-141.
27. Singh V, Misra AK. Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture*. 2017; 4(1): 41-49.
28. Bai X, Li X, Fu Z, Lv X, Zhang L. A fuzzy clustering segmentation method based on neighborhood grayscale information for defining cucumber leaf spot disease images. *Computers and Electronics in Agriculture*. 2017; 136: 157-165.
29. Lu Y, Yi S, Zeng N, Liu Y, Zhang Y. Identification of Rice Diseases using Deep Convolutional Neural Networks. *Neurocomputing*. 2017.

30. Abobakr A, Hossny M, Nahavandi S. A Skeleton-Free Fall Detection System From Depth Images Using Random Decision Forest. *IEEE Systems Journal*. 2018; 12(3): 2994-3005.
31. Wang H, Zheng B, Yoon SW, Ko HS. A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European Journal of Operational Research*. 2018; 267(2): 687-699.
32. Ma J, Du K, Zheng F, Zhang L, Gong Z, Sun Z. A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network. *Computers and Electronics in Agriculture*. 2018; 154: 18-24.
33. Park K, Hong Y, Kim G, Lee J. Classification of apple leaf conditions in hyperspectral images for diagnosis of Marssonina blotch using mRMR and deep neural network. *Computers and Electronics in Agriculture*. 2018; 148: 179-187.
34. Zhang S, Wang H, Huang W, You Z. Plant diseased leaf segmentation and recognition by fusion of superpixel, K-means and PHOG. *Optik*. 2018; 157: 866-872.
35. Kaur S, Pandey S, Goel S. Semi-automatic leaf disease detection and classification system for soybean culture. *IET Image Processing*. 2018; 12(6): 1038-1048.
36. Garcia J, Barbedo A. Plant disease identification from individual lesions and spots using deep learning. *Biosystems Engineering*. 2019; 180: 96-107.
37. Pantazi XE, Moshou D, Tamouridou AA. Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers. *Computers and Electronics in Agriculture*. 2019; 156: 96-104.