Research Article **Open Access**

Plant Leaf Disease Detection Using Support Vector Machine

Mohammed A. Hussein*, Amel H. Abbas
Department of Computer Science, College of Science, Mustansiriyah University, IRAQ
*Correspondent author email: mohammed.alsraf@gmail.com

Article Info

Received
30/03/2018

Accepted
26/07/2018

Published
15/08/2019

Abstract
Agriculture has special importance in that it is a major source of food and, clothing and is an important economic source for countries. Agriculture is affected by a variety of factors, biotic such as diseases resulting from bacteria, fungi, and viruses and non-biotic such as: water and, temperature and other environmental factors. Detection of these diseases require people to experts in addition to a set of equipment and it is expensive in terms of time and money Therefore, the development of a computer based system helps the detection of the plants' diseases is very helpful for farmers As well as to specialists in the field of plant protection. the proposed plant disease detection system consists of two phases, in the first phase, the knowledge base is established by introducing a set of training samples in a series of processing that include first use pre-processing techniques such as: cropping, resizing, fuzzy histogram equalization, extracting a set of color and texture features and used to great the knowledge base that used as training data for support vector machine classifier. In the second phase, we use the classifier that was trained using the knowledge base for detection and diagnosis of plant leaf diseases. To create the knowledge base, we used 799 sample images that divided it by 80% training and 20% testing. We have use Three crops each yield three diseases in addition to the proper state of each crop .the accuracy of disease detection was 88.1%.

Keywords: Plant Disease Detection, SVM, GLCM, Texture, Color Feature.

الخلاصة
للزراعة أهمية خاصة لأنها مصدر رئيسي للغذاء واللباس وتعتبر مصدرا اقتصاديا هاما للبلدان. تتأثر الزراعة بمجموعة متنوعة من العوامل الحيوية مثل الأمراض الناتجة عن البكتيريا والفطريات والفيروسات وغير الحيوية مثل: الماء ودرجة الحرارة والعوامل البيئية الأخرى. إن الكشف عن هذه الأمراض يتطلب اشخاص خبراء بالإضافة إلى مجموعة من المعدات وهو مكلف من حيث الوقت والمال ولذلك فإن تطوير نظام قائم على الكمبيوتر يساعد في اكتشاف أمراض النباتات مفيد جدًا للمزارعين وكذلك للمتخصصين في مجال وقاية النبات. يتكون نظام اكتشاف أمراض النبات المقترح من مرحلتين ، في المرحلة الأولى ، يتم تأسيس قاعدة المعرفة عن طريق إدخال مجموعة من عينات التدريب في سلسلة من العمليات التي تشمل تقنيات المعالجة الأولية الأولى مثل: الاقتصاص ، تغيير الحجم ، الرسم البياني الضبابي المعادلة ، واستخراج مجموعة من ميزات اللون والملمس واستخدامها في قاعدة المعرفة التي تستخدم كبيانات التدريب لل SVM. في المرحلة الثانية ، نستخدم المصنف الذي تم تدريبه باستخدام قاعدة المعرفة للكشف عن أمراض أوراق النبات وتشخيصها. لإنشاء قاعدة المعرفة ، استخدمنا ٧٩٩ عينة من الصور التي قسمتها بنسبة ٨٠٪ من التدريب واختبار ٢٠٪. لدينا ثلاث محاصيل تنتج كل منها ثلاثة أمراض بالإضافة إلى الحالة السليمه لكل محصول. كانت دقة الكشف عن المرض ٨٨,١٪.

Introduction
There are many diseases that affect crops and lead to significant production losses, which threaten the issue of food security. Human visual examination with the naked eye is the way most widely used and common. This method gives a large room for error depending on where the farmers trying to detect the disease through visual inspection as a big chance of error in some cases resorting to experts, this latter needs a lot of time, effort and money. Another problem in Iraq, is that the most of the crop Fields are located in rural areas, which require farmers to go long distances to find experts [1]. Image processing gives accuracy, high-speed, do not require large amount of money to and time-consuming as in brought experts [2].

Related Work

There is a lot of research that has worked in the field of identification of plant diseases including: Basavaraj Tigadi1 and Bhavana Sharma [3]. In this research, the researcher used Color Features and Artificial Neural Network For classification a range of diseases affecting bananas. In this work the first step is converting the image from RGB to gray and HSV color space then extracts Histogram of template and color features. The researcher uses the color features including Mean and Standard Deviation.So as to use these features to create a knowledge base that is used later by the classifier for training. use the feed-forward back propagation neural network to classify the banana disease.

One of the important missing things in this system is the lack of segmentation process .the process of segmentation is very important in separating the injured part from the proper part of the leaf image and even if the results of the diagnosis are good, the possibility of more than one disease in the leaf is possible. Therefore, it is necessary to use segmentation method such as clustering to separate the different diseases from the healthy part.

R. N.kadu et al [4] also develop a research to detect the leaf disease using Otsu threshold and Support vector machine. In this research first contrast adjustment is done as a preprocessing after that RGB image is converted to YCbCr . Otsu threshold was applied to separate the injured part from the healthy part. Feature extraction is obtained from the textual feature. For which Gabor filter applied on the segmented image. Finally, the features used as input to SVM classifier.

Leaf rot disease detection of Betel Vine also done by using Color analysis [5]. In the pre-processing, the process of cropping was performed to eliminate the background that containing unnecessary information in the process of disease detection. The color feature is used to distinguish rotted leaf area form healthy leaf area .the image is converted to three type of color space RGB, HSV and YCbCr next by using color analysis the researcher find the hue of HSV give the best result the next step is using hue thresholding for discriminating leaf rotted part of the rest of the background. Finally, convert the affected part to a binary image and calculate the white pixel to find the area of the affected part.

The Proposed System

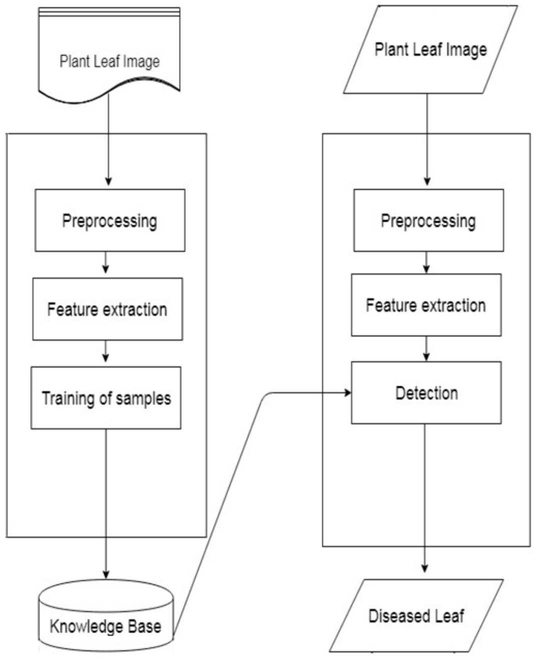


Figure 1: Proposed plant disease detection System Architecture.

The first part of the proposed plant disease detection system is the process of training. At this stage, the images of the plants are obtained using a digital camera. Then the image preprocessing techniques are applied to these images. After preprocessing, useful image features are extracted using feature extraction technique that will be used as training samples for the support vector machine algorithm (the proposed machine learning algorithm in this system). In the detection phase, the images will be obtained first by capturing them with a digital camera. After that, the image processing techniques referred to in the training phase will be applied and finally, the case will be classified as either infected or healthy through support vector machine (SVM).

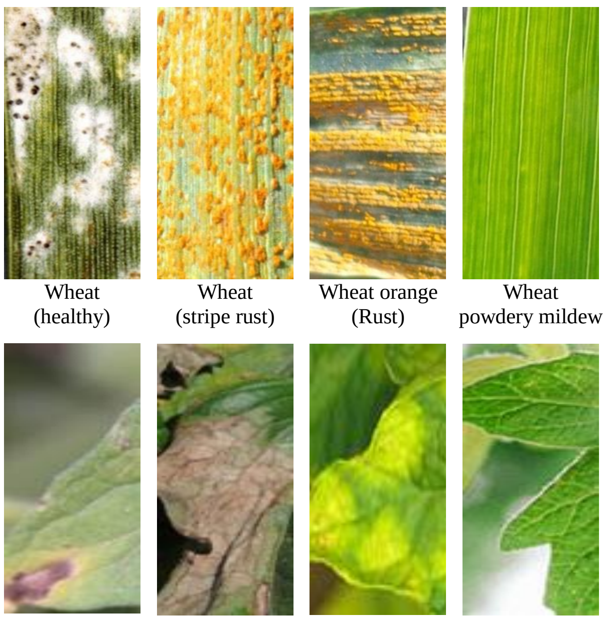
Dataset

In this study, the images were collected from multiple sources, including the field visit to Wasit governorate in Al-Nu'maniyah district, Due to the difficulty of conducting field visits as well because of seasonal conditions, the researcher was forced to take part of the pictures from websites, mostly to a group of international universities.

In this study, images were collected for three crops, wheat, tomatoes and cucumbers, three different diseases for each crop, as well as the health status of each of these crops. Table 1 showing the diseases used for each of these crops and the number of samples. Sample images for the above specified groups are shown in Figure 2.

Table 1: Disease dataset.

Name of crop	Type of Case	Number of Samples
Wheat	powdery mildew	70
Wheat	orange rust	79
Wheat	stripe rust	70
Wheat	healthy	24
Tomato	septoria leaf spot	71
Tomato	late blight	73
Tomato	leaf mold	74
Tomato	healthy	60
Cucumber	Downy mildew	80
Cucumber	powdery mildew	86
Cucumber	Mosaic virus	40
Cucumber	healthy	71



Wheat (healthy) | Wheat (stripe rust) | Wheat orange (Rust) | Wheat powdery mildew

Tomato (healthy) | Tomato (leaf mold) | Tomato (late blight) | Tomato (septoria leaf spot)

Figure 2: Sample images of diseased crops.

Image pre-processing

It is a process aim to improving the image and, configures it for subsequent processes by removing noise and unwanted objects and improving the visual appearance; it also gives a positive effect on both the process of segmentation and features extraction and therefore has an impact on the final outputs of the system and accuracy.

The process of image processing begins with the acquisition of the image from the environment through the digital camera and stored on the hard disk of the computer and then downloaded to the system for the rest of the operations. In this study, three pre-processing operations were applied. The first process crops the image to eliminate the background as much as possible. The second operation was to improve the image by using Fuzzy Histogram Equalization (FHE); the third process was to give a fixed size for all the input images both in the training phase and in the diagnosis phase.

Image Cropping

The image that is captured by a digital camera containing about 30% of the infected plant leaf information and the remaining 70% of the rest of the information is not important because it represents the background. This background is an unnecessary consumption of memory, and also in the treatment time in the CPU during the process of retail segmentation In order to gain efficiency in the storage and speedup the processing time. It is important we deduct the portion of the image through a process of cropping. In this study, we use the command imcrop (I) in Matlab that uses the interactive cropping tool where (I) is the image we need to crop it. We would like to point out that the process of cropping must be precise and careful not to cut any important information from the image because in this type of systems the accuracy is more important than the time.

Image Resize

Resize all images that will be used to a fixed size (300*400). This fixed size was used for all imported images because the accuracy of the feature extraction process is affected if the images are in different sizes.

Fuzzy Histogram Equalization

To enhance the image this study using Fuzzy Histogram Equalization. The FHE contains two periods. First, the fuzzy histogram is computed based on weird set theory to manage the inexactness of grey level values in an improved way in comparison to classical clean histograms. The second stage, the fuzzy histogram is divided into two sub histograms based on the median value of the initial image and then equalizes them independently to preserve image brightness.

Features Extraction

The image contains a lot of information, only some of these information can be used to distinguish between different situations. so much of the information in the image must be converted to reduced representation set called the features process that extracts features from the image called features extraction.

The image has many features such as texture, color, and shape. These features can be used as mini-information representing the useful information in the image which can be used to distinguish between different situations. In this study, a range of texture features and color features were used to separate the affected case from the health case, as well as to diagnose the various diseases. Shape features were not used in this study because the shape of the injury changes continuously during the stages of disease growth.

In order to obtain the color and texture features, the GLCM is considered by set the number of gray level to 8 and the offset be [0,1]. The final output of the process of extracting the texture features and the color features is the Excel file as shown in Figure 3. This includes all the extracted features that are later used as a knowledge base for training the classifier to detect plant disease.

Contrast	Correlation	Energy	Homogeneity	Mean	SD	Entropy	RMS	Variance	Smoothness	Kurtosis	Skewness	IDM
0.234661654	0.920562952	0.846437324	0.987879799	11.157408	41.832	0.9748	3.34	1603.7	0.99999975	17.2634	3.871399	255
0.185530493	0.922665945	0.834470396	0.98415881	9.6375278	38.894	0.9871	3.96	1459.4	0.99999971	22.5359	4.418751	255
0.139172932	0.945107975	0.83940679	0.985648317	8.2423722	39.972	0.8983	3.44	1300.7	0.99999966	32.6338	5.512575	255
0.828170426	0.849303855	0.488288039	0.91210403	30.433392	56.621	2.9123	7.94	2800.3	0.99999991	4.55926	1.691121	255
0.241428571	0.758403164	0.864949097	0.976490562	5.0220833	24.542	0.7394	2.87	565.67	0.99999945	41.5125	5.87787	255
0.320250627	0.712841196	0.885518839	0.975809206	4.9183694	26.231	0.6378	2.33	638.58	0.99999944	43.0493	6.144689	255
1.338437761	0.8328478	0.52752853	0.912245236	34.315261	69.664	2.6536	6.96	3402.9	0.99999992	5.03253	1.856461	255
0.351453634	0.439436726	0.713299318	0.9340037	5.7731194	22.509	1.112	3.48	440.89	0.99999952	27.6369	4.690834	255
0.123650794	0.937744451	0.802146915	0.9853248	9.8948194	35.406	1.2054	4.47	1195.4	0.99999972	17.8039	3.869465	255
0.455438596	0.948200074	0.35232235	0.962558221	57.914689	77.792	4.364	9.63	3848.4	0.99999995	2.45319	0.953109	255
0.308755221	0.451705195	0.742131999	0.940271433	5.1923556	21.639	1.0107	3.21	411.29	0.99999947	30.2258	4.990935	255
0.811771094	0.87848721	0.647325135	0.945213977	26.936194	63.772	2.0576	6.43	3774.9	0.9999999	6.72099	2.253283	255
0.235605681	0.923512111	0.879713112	0.990068843	9.6473944	42.422	0.7517	3.37	1704.8	0.99999971	22.6307	4.527903	255
0.256349206	0.936460948	0.759263307	0.979291522	15.554936	49.28	1.4558	5.04	2311	0.99999982	13.7323	3.384036	255
0.124168755	0.708544235	0.728769732	0.964564148	4.9641972	19.729	1.1426	3.73	343.7	0.99999944	51.471	5.906278	255
0.732272348	0.83590039	0.748502548	0.951584089	17.122289	52.675	1.4524	5.22	2631.1	0.99999984	12.1058	3.175155	255
0.627936508	0.871200479	0.702350062	0.953368093	19.895378	55.232	1.7439	5.83	2919.3	0.99999986	10.0305	2.839358	255
0.258780284	0.84844205	0.682947712	0.956439989	11.341483	31.11	1.7774	4.74	731.94	0.99999976	13.0277	3.130098	255
0.590977444	0.852966669	0.477261367	0.934696404	25.967214	54.948	2.8198	7.34	2600.5	0.99999989	6.32205	2.099296	255
1.71304929	0.759635552	0.563338088	0.892834984	29.405392	68.047	2.3376	6.53	3815.5	0.99999991	6.47685	2.217247	255
0.364302423	0.935911301	0.71921283	0.971955086	22.349342	59.198	1.7045	5.53	3041.1	0.99999988	8.47634	2.591994	255

Figure 3: Texture feature and color feature for a set of the training sample.

Training the Detection System

Identification of Patterns using a machine learning approach has two basic stages. In the first stage, the classifier is trained using the training samples to extract the weights. The system then examines the accuracy of the system using the test samples. Therefore, the total samples were to be divided into training samples and testing samples. In this study, we divided the total samples into 80% training and 20% testing. As it is shown in the Figure 4 from the total dataset (799 images) 80 % were used for training the SVM classifier and 20% used for testing. The following settings were used in SVM to get the best result (Kernel function: Quadratic, Box constraint level: 4.0) Full details for SVM settings that used is shown in the Figure 5. The same ratios were used in the division of disease samples for the diagnosis process

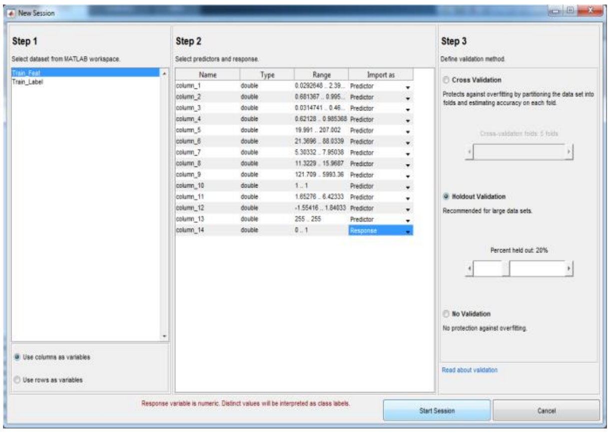
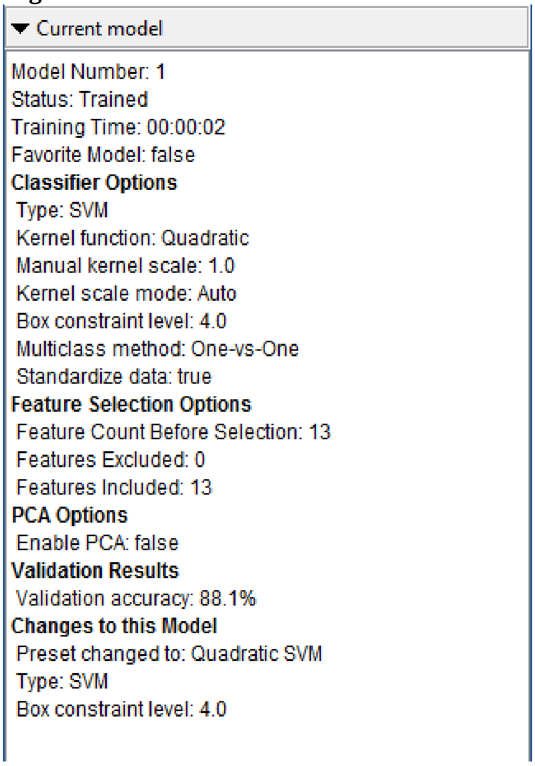


Figure 4: screenshot of the dataset classification.



▼ Current model

Model Number: 1
Status: Trained
Training Time: 00:00:02
Favorite Model: false
Classifier Options
Type: SVM
Kernel function: Quadratic
Manual kernel scale: 1.0
Kernel scale mode: Auto
Box constraint level: 4.0
Multiclass method: One-vs-One
Standardize data: true
Feature Selection Options
Feature Count Before Selection: 13
Features Excluded: 0
Features Included: 13
PCA Options
Enable PCA: false
Validation Results
Validation accuracy: 88.1%
Changes to this Model
Preset changed to: Quadratic SVM
Type: SVM
Box constraint level: 4.0

Figure 5: Full details for SVM settings.

Testing the detection system

After the training is completed, the classifier will use 20% of the total samples to examine the accuracy of the system depend on confusion matrix. Were accurately in detection the disease case 90.6.1% and the accuracy of detecting the healthy case 77.4 with average accuracy for the detection system 88.1% as shown in the figure 6 and figure 7. Accuracy (AC) is defined as the ratio of correct predictions (CP) to the total number of predictions which represents correct predictions (CP) + false predictions (FP). The following equation will be used to calculate accuracy.

$$AC=CP/(CP+FP) \quad (1)$$

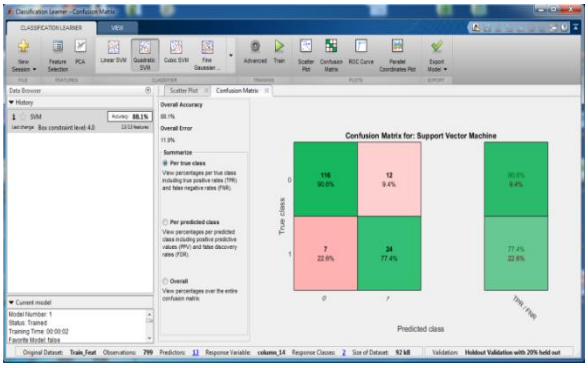


Figure 6: SVM disease detection testing result.

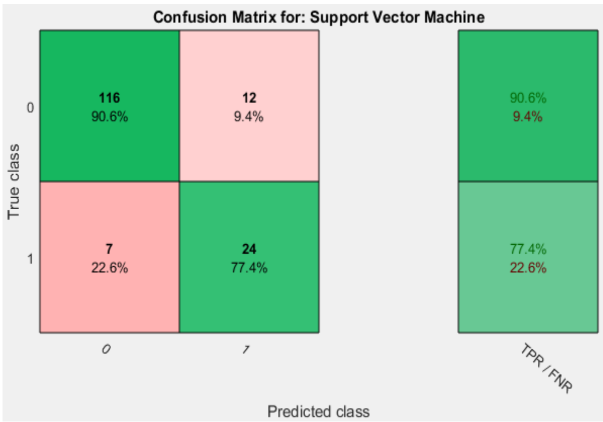


Figure 7:SVM disease detection confusion matrix.

Conclusion

It was concluded that the accuracy of disease detection is increased as the number of training samples increases and that the change in SVM settings also affects accuracy.

Reference

[1] Ayswarya.R1 , Balaji.B1, Balaji.R , Arun.S1 , Ramya.R, " Weed Detection in Agriculture Using Image Processing",International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. ISSN 2278 – 8875

[2] Fenfang Lin, Dongyan Zhang, Yanbo Huang, Xiu Wang and Xinfu Chen," Detection of Corn and Weed Species by the Combination of

Spectral, Shape and Textural Features",MDPI. 9, 1335; doi:10.3390/su9081335

[3] Basavaraj Tigadi, Bhavana Sharma," Banana Plant Disease Detection and Grading Using Image Processing".International Journal of Engineering Science and Computing IJESC,Volume 6 Issue No. 6

[4] R. N.kadu, S. Kangane , S.Vikhe3, R.Pandita, V. Inamke," Leaf Disease Detection Using Arm7 and Image Processing",Int. Journal of Engineering Research and Applications. ISSN: 2248-9622, Vol. 5, Issue 2

[5] Amar Kumar Deya, Manisha Sharmaa, M.R.Meshramb," Image Processing Based Leaf Rot Disease, Detection of Betel Vine", International Conference on Computational Modeling and Security. 748