Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7733327

PLANT DISEASES DIAGNOSIS AND TREATMENT

Wamidh K. Mutlag[1], Zahoor M. Aydam[2] and Biadaa M. Rashed[3]
[1]Al shatrah Technical Institute, Southern Technical University, Iraq.
[2,3] Computer sciences and mathematics college, University of Thi_Qar, Thi_Qar Iraq.
Emails: [1]wamid.almuhaysen@stu.edu.iq , [2]zahoor.mosad@utq.edu.iq ,
[3]baidaaalsafy@utq.edu.iq

Abstract
In this article, two are the two main characteristics that the machine learning method of plant disease detection must achieve, pace and precision. In this research, an automatic discovery and classification of leaf diseases have be propose then personate the way of treatment, this method is based on K-means as a clustering procedure and KNN as a classifier tool using texture feature set, entropy, contrast, RMS, and mean. As a test phase, we utlize a collection of leaves that are possessed from the Al- Ghor area in Jordan. In our research, eight types of diseases that affect plants were identified; they are Alternaria Alternata , Anthracnose , Bacterial Blight , Cercospora Leaf Spot , Healthy Leaf, cucumber mosaic virus, Graphiola phoenius and Diplocarpon rosae. The propose frame could successfully expose detection and classification of diseases with a precision of 100% on meduim with more than 20% speeding up over the offered path in the training stage and 95% in the testing stage.
Keywords: KNN, entropy, contrast, RMS, mean.

1. Introduction

Plant diseases have turned into a big problem due to the decline in the quality and quantity of agricultural products . The percentage of losses of plant diseases for the year 2007 in Georgia is about 539.74 million dollars , and of this amount, about 185 million US dollars have been spent on combating diseases and the rest is the amount of damage caused by diseases .These numbers are listed in Table 1.

The method of visual control by experts is the prime method in pursuit for the process of discovering and identifying plant disease (Weizheng, Yachun, Zhanliang, & Hongda, 2008). However, it require constant observation by means of experts , Which may be expensive in great farms . Moreover ,in developing countries , farmers have to travel long distances for the purpose of communicating with experts , which makes the consultation process very time consuming and costly (Babu & Rao, 2007; Camargo & Smith, 2009).

The method of detecting plant diseases is one of the main research topics, as it is useful in monitoring large fields , and thus automatic detection of disease symptoms that appear on plant leaves(Hillnhuetter & Mahlein, 2008) (Al Bashish, Braik, & Bani-Ahmad, 2011). Therefore, this research is a fast, accurate and less expensive method for detecting cases of plant diseases and has great practical importance(Rumpf et al., 2010).

Crop	Value of damage ($ millions)	Cost of control
Apple	0.073	0.267
Blueberry	0.14	2.555
Bunch Grape	0.112	0.27
Corn	12.4	0.5
Cotton	81.7	12.2
Muscadine Grape	0.026	0.096
Ornamental	41.22	21.2
Peach	0.177	3.19
Peanut	58.7	41.2
Pecan	0.64	17.4
Soybean	5.3	1.9
Strawberry	0.32	0.683
Turf	126.6	61.2
Vegetable	18.1	20.6
Wheat	0.99	1.8
Totals	**346.49**	185.06

Table 1 :Summary of total losses due to disease damage and cost of control in Georgia ,USA in 2007

2. The ?Proposed ?Algorithm

The proposed algorithm consists of four steps (images acquisition , preprocessing step Segmentation features extraction , classify (in training) or (testing) to one of the eight types of plant diseases then diagnose the method of treatment as shown in Fiqure 1.

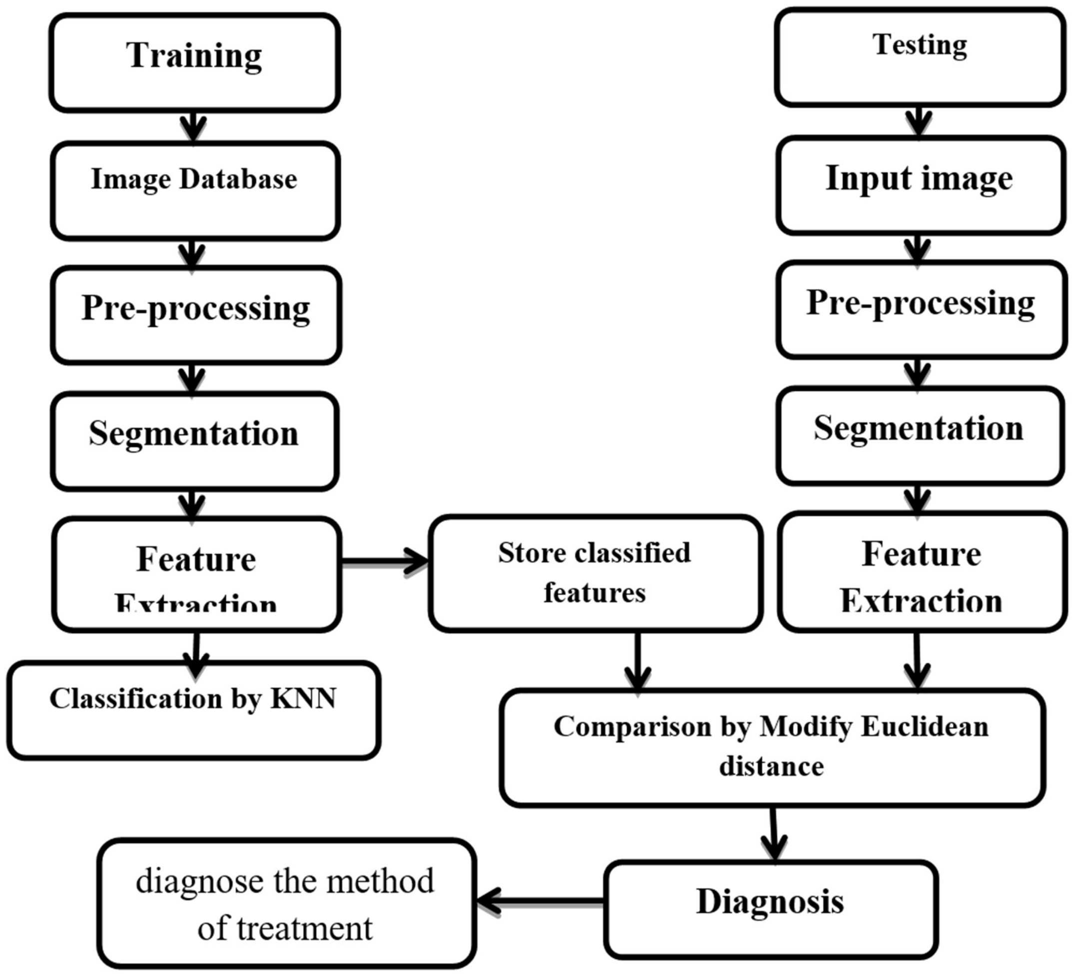

Figure 1. General Block Diagram of Proposed System

2.1.Image acquisition

The images are collected from the plant leaf are taken by the camera, there are 280 from Plant Disease
different used (200 images training) and(80 images testing) with format jpg.

2.2.Pre-processing

In this step, the image Enhancement is done by removing noise into the image using a smoothing filter (median Filter).

2.3.Segmentation

Segmentation of the image into segments each part has the same features. There are numerical methods of segmentation and this research use the clustering k-mean algorithm for segmentation:

k-mean clustering

The clustering of the objects depends on a set of characteristics N in a number of categories by decreasing the sum of the squared distance(Neelamegam & Ramaraj, 2013) .

k-mean algorithm (Ali & Aydam, 2019)

1.limite the number of groups , who is a preparatory initialization stride in this action , the number of N=3. .

2. Determine the coordinates of centers of the Centroid randomly for the first time and calculated (the average points belonging to the center) for the rest of the time.
3. Calculate the square Euclidean distance between each center and each pixel in the image using the equation shown below

$$DE = \sum_{i=1}^{n}(q_i - p_i)^2 \dots\dots\dots\dots\dots\dots\dots\dots (1)$$

4. Gather the data together with the nearest center.
5. Repeat 3 times steps 2 to 4.

2.4.Feature Extraction

2.4.1 Tamura

Tamura is the approach to devising texture features based on human visual perception . It defined six features for textural (coarseness , contrast , directionally, regularity , roughness and line-likeness). The first three achieved are very effective results and are used in our estimation, both separately and as combined values (Howarth & Rüger, 2004; Umamaheswari & Bhavani, 2018).

2.4.1.1. Coarseness

Its objective is to determine the larger volume of textured tissue, and in the case of fine micro-tissue, we first take the arithmetic on average at each point on the neighborhoods, the linear magnitude of that, which is powers of 2 average A to the neighbor of size 2k x 2k at the point (x, y), where k = 0, 1, . . , 5

$$A_k(x,y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} f(i,j)/2^{2k} \dots\dots\dots \quad (2)$$

Then, at each pixel, the difference (Ek(x, y)) between pairs of non-overlapping moving averages in the horizontal and vertical direction is calculated, and at each pixel, the k value that increases Ek (x, y) in any direction is used to determine the size Best (Sbest): Sbest (x, y) = 2k. The Roughness Scale (FCRS) is then calculated by averaging Sbest (x, y) over the entire image .

2.4.1.2. Contrast

It serves to capture the dynamic range of gray levels in an image, in conjunction with the polarization of the black and white distribution. The first is measured using the standard deviation of gray levels and the second is α4 kurtosis. So the measure of variance is defined as .

$$F_{con} = \sigma / (\alpha_4)^n \quad \text{Where} \quad \alpha_4 = \mu_4 / \sigma^4 \dots\dots\dots \quad (3)$$

μ4 is the fourth moment of the mean and σ is the variance. Experimentally, Tamura found n = 1/4 to give the closest agreement to anthropometrics. This is the value we used in our experiments .

2.4.1.3. Directionality

It is a universal fabric feature. Patterns can be non-directional or highly directional. The degree of direction, measured on a scale from 1 to 0, power is used as a descriptor.
In this paper, the Tamura method is applied to extract texture attributes of gray-level images.

2.4.2. GLCM (Gray Level Co-occurrence Matrix)

GLCM is a mixture of different levels of gray level in the image .The features used in this research are anisotropy, correlation, entropy, and homogeneity Below is how to calculate the extracted features (Albregtsen, 2008; Mutlag, Ali, Aydam, & Taher, 2020):

2.4.2.1Contrast

It is a measure of the difference in gray pixels
The image level is equivalent. 1 is a formula for calculating the variance.

$$Contrast = \sum_{i=1}^{L}\sum_{j=1}^{L}|i-j|^2 \times GLCM(i,j) \qquad \text{......................(4)}$$

2.4.3. Color Feature Extraction

The color is the most essential characterizes of the image . Most color images are in the RGB color space . The image includes three different components : red , green, and blue (Kodituwakku & Selvarajah, 2004). The color characterizes in an image can be known based on the intensity of the pixels . In this study, the required color features were the mean color and the standard deviation of insensitivity for each color component.

2.4.3.1 Mean Color (μ)

we are using to know the range of color depth in an image . Eq. 5 is the formula to have got the mean of color (Kodituwakku & Selvarajah, 2004).

$$\mu = \frac{1}{LL}\sum_{i=1}^{L}\sum_{j=1}^{L}X_{i,j} \qquad \text{......................(5)}$$

where,

L = number of rows and columns

X = amount of color depth

i, j = values of rows and columns

2.4.3.2 Color distribution entropy (Alamdar & Keyvanpour, 2012)

Depend on the toroidal chromaticity graph, the NSDH (Normal Spatial Distribution graph) can be defined as

$$\boldsymbol{P}_i \text{ where } \boldsymbol{P}_i = (P_{i1}, P_{i2}, \hbar, P_{iN}) \text{ and } P_{ij} = |\boldsymbol{A}_{ij}|/|\boldsymbol{A}_i|.$$

John (2000) suggested using entropy developed by Shannon (1948) to represent color information for an image and to retrieve images in CBIR. Depend on the NSDH and entropy definition , we propose a new descriptor, CDE (Color Distribution Entropy), that describes the spatial information of an image. The CDE of a color container i can be defined as

$$E_i(\boldsymbol{P}_i) = -\sum_{j=1}^{N} P_{ij}\log_2(P_{ij}) \qquad \text{......................(6)}$$

Which gives the degree of dispersion of pixel spots for the color container in the image. A large Ei means that the distribution of pixels is scattered, otherwise the distribution will be compressed. Then the CDE index of the image can be written as (h1 E1 , _h, hi, Ei, _h, hn En), where hi is the graph of the color container i, Ei is the CDE of the color container i and n is the number of bins.

Due to the lower dimensional indices, SCH , geostatic , and CDE are more efficient than toroidal chromaticity graphs. In addition , the toroidal chromaticity graphs mentioned in (Rao et al., 1999) are scalable because they are related to the number of pixels in the color container in the toroidal. The parameter r described in (Cinque et al., 1999) is also a variable of size because it is related to the number of pixels and the density of those pixels in the color bin. Normalized by the number of pixels per container, L mentioned in (Lim and Lu, 2003) is the constant size . In this paper , NSDH is the constant size because the toroidal chromaticity graphs are normalized to the number of pixels in the color bins, thus, the CDE is also a constant size.

2.4.3.3 Root Mean Squar*e* Error (Chai & Draxler, 2014)

The root means square error (RMSE) is commonly used to evaluate the differences between values predicted by a model or an estimator and the values observed from the data being modeled or estimated. The RMSE is a good measure of precision, whereas in this paper this analysis is used to measure the difference between an image and its encrypted version. In common applications of RMSE, the algorithms generally use the results and try to minimize the error. In the case of image encryption, greater error depicts better encryption. We call the individual differences between two images residuals, and the RMSE serves to aggregate them into a single measure. Mathematically we can represent RMSE as

$$\mathrm{RMSE} = \left[n^{-1} \sum_{i=1}^{n} |e_i|^2 \right]^{\frac{1}{2}}, \qquad \text{......................(7)}$$

Where ei is the difference between the intensity of the pixel of the plain image and the corresponding pixel of the encrypted image. N is the number of pixels in the image being encrypted. The stated rationale for squaring each ei is usually 'to remove the sign' so that the magnitudes of the errors influence the average error measure .

3.KNN algorithm

The K-NN algorithm is a method that uses a supervised algorithm. The goal of the k-NN algorithm is to classify novel objects based on features and training samples. The result of the new test samples was classified based on the k-NN category. In the classification method, that algorithm does not use a model to match and is based solely on memory. The K-NN algorithm uses biology classification as a predictive value from a novel test sample. The distances used are Euclidean distance, cosine distance, correlation distance, and city block distance (Ünay, Çataltepe, & Aksoy, 2010) .

Pseudocode for k-Nearest Neighbor algorithm (Preece, Goulermas, Kenney, & Howard, 2008)

Classify (**P**, **Q**, p)
P: training data,
$\boldsymbol{Q}$: class labels of $\boldsymbol{P}$
p: unknown sample
for j= 1 **to** K **do**
Compute distance (P_j, p)
end for
Compute set J containing indices for the k smallest distances (P_j, P).
return majority label for {Q_j where j ∈ J}

4.Result:

The proposed system uses 280 plant diseases images, which are divided into two groups (200for training) and 80 for testing. The process consists of four stages, the first stage is the initial treatment of the image, as shown in Fig .2, and then segmentation as shown in Fig .3 and then extracting features from each image of plant diseases as shown in Table 2. The features extracted from each image of plant diseases will be classified using the KNN algorithm, Fig. 4 shows the features extracted from the plant diseases image in the training stage, where the results of the classification are given (100%) accuracy in both KNN algorithm as shown in Table 3.

In the testing stage the features will classify where the input image will be in the class of 8 types of plant diseases by comparing the feature vector with the 200 vectors stored in the dataset training by using Equation(8)[15] where the results by using Modify Equation distance of the classification are given (95%) accuracy as shown in Table 4.

a)Original image

b)Image Enhancement

Figure 2. Pre-processing image (a.orginal image ,b.image Enhancement)

Figure 3. Segmentation image Enhancement

Table 2: Feature Extraction

Type of Disease	Contrast	Mean	Entropy	RMS	Tamura Texture Feature		
					coarseness	contrast	direction
AlternariaAlternata	0.456	82.947	6.795	14.394	31.828	43.752	0.781
Anthracnose	0.83	27.195	2.904	7.984	17.405	37.487	0.779
Bacterial Blight	0.61	48.279	3.982	10.312	30.09	60.479	0.79
Cercospora Leaf Spot	0.768	32.293	4.514	10.336	17.158	33.565	0.762
Healthy Leaf	0.683	38.392	3.589	8.804	25.663	58.331	0.762

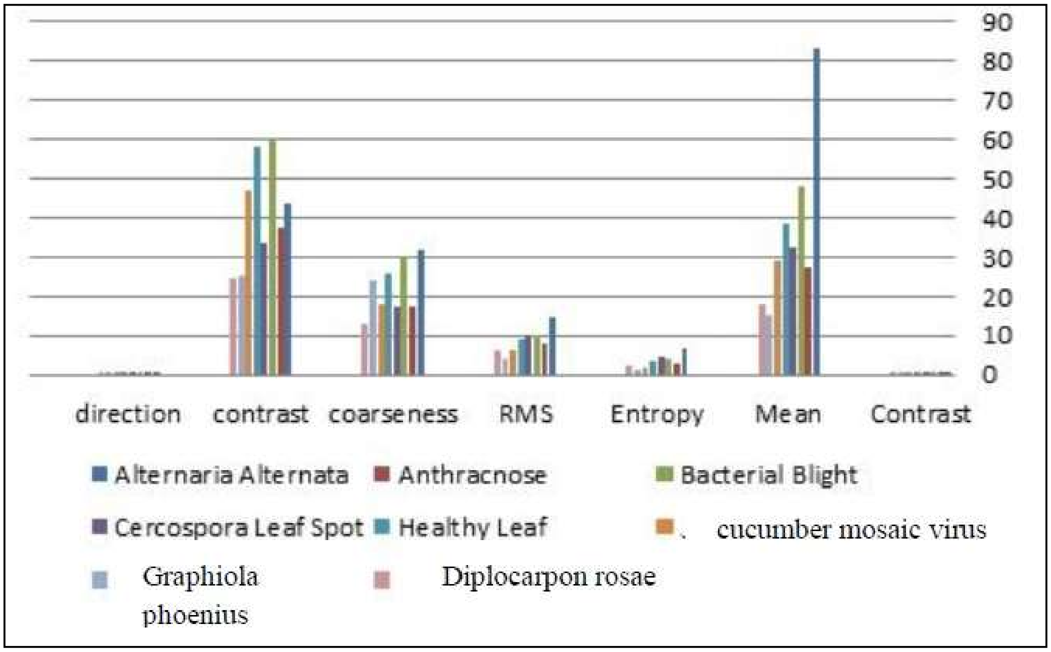

Figure 4: Feature Extraction

Table3: the accuracy rate of the phase Training

No of images	Type of Disease	Accuracy by using KNN algorithm K=2			
		Euclidean distance	Cosine Distance	Correlation distance	Cityblock distance
25	AlternariaAlternata	100%	100%	100%	100%
25	Anthracnose				
25	Bacterial Blight				
25	Cercospora Leaf Spot				
25	Healthy Leaf				
25	cucumber mosaic virus				
25	Graphiola phoenius				
25	Diplocarpon rosae				

$$MED=\frac{\sqrt{\sum_{i=1}^{m}(p_i-q_i)^2}}{\left(\frac{1}{2}+(\frac{1}{m}\sum_{i=1}^{m}(p_i-\bar{p})^2)^{\frac{1}{2}}+(\frac{1}{m}\sum_{i=1}^{m}(q_i-\bar{q})^2)^{\frac{1}{2}}\right)}\ldots\ldots(8)$$

p_i: is the i^{th} valuse of first vector value.

q_i: is the i^{th} valuse of second vector value.

m: is the number of elements in vector.

$\overline{p}$: is the mean value of first vector

$\overline{q}$: is the mean value of second vector

Table4 : the accuracy rate of the phase Testing

No of images	Type of Disease	treatment method	Accuracy
10	AlternariaAlternata	The treatment method has the following method (Fayad & Mania, 2008): spray date palm leaves with Mancozebfungichde, the treatment should be repeated after7-10 days.	95%
10	Anthracnose	Method of treatment We follow the following steps(Mehrotra, 2013): 1-Use(cultivation) of Resistant or tolerant varieties 2-control of vector insects with insecticides 3-Disinfection of seeds with fungicide 4-Spray the plants with appropriate fungicide after month of plantation. 5-collection and destruction of plant residue 6-Avoidence of susceptible varieties 7-use of mono or di potassium phosphate.	
10	Bacterial Blight	Method of treatment We follow the following steps (Agrios, 2010): 1-use of crop rotation 2-use of seeds and seedling free from infection	

		3-Destruction of infected plants out of field 4-Spray the plants with copper compounds 5-use of resistance varieties.	
10	Cercospora Leaf Spot	Method of treatment We follow the following steps (Kumar, Srivastava, Roy, Verma, & Saini, 2019): 1-use of seeds free from pathogens 2-Seeds treatments with fungicides 3-Spray infected plants with fungicides such as Azole fungicides	
10	Healthy Leaf	free from disease	
10	cucumber mosaic virus	Method of treatment We should follow the following steps (Matthews, 2012) : 1-Get rid of the harmful weeds on which aphids feed, especially the weeds that spread at the beginning of sprin. 2-Choosing separate areas on the farm to grow the crop between early and late may limit the spread of the disease in late cultivations. 3-Viruses do not multiply on several plants on their own. They are committed parasites, so when the infection appears in small spots, they should be completely disposed of with complete caution and burned away. 4-Beware of transmitting the virus mechanically through machines such as axes and mowers, and worker's feet and hands	
10	Graphiola phoenius	Method of treatment We follow the following steps(Zhuang, Chen, Shim, & Bai, 2007): 1-Cut and burn infected fronds 2-Spray the infected palm with fungicides	

10	Diplocarpon rosae	The treatment method should follow the following method(Mehrotra, 2013): Control begins by removing the affected leaves and branches and cutting them a few centimeters below the infected site, collecting and burning them, while burying any remaining leaves in the soil. A program of preventative spraying with fungicides is applied starting from the summer and before the spots appear so that it is sprayed twice a week if the disease is rapidly spreading in the region or every 7-10 days if the disease is slow to spread Captan, Copper, sulphur compound,Mancozeb and chlorothalonil compounds were effective in control such disease	

5.Conclution

The proposed hybrid method is applied to plant disease images to classify plant disease diagnosis and treatment. The automatic plant disease diagnosis and detection method reduces the manual marking time and avoids human error. This approach is using a K-means as a clustering procedure to segmentation then features extraction, entropy, contrast, RMS, and mean used for and then KNN algorithm for classification of Plant Diseases Diagnosis Images.

KNN algorithm is the best one in classify with the percentage of (100%) in training and (95%) in testing . In the testing stage, the results found that the modify of the Euclidean distance is the best metric for calculating the corresponding feature vector of the tested image to know the type of which 8 types of Plant Diseases Diagnosis and Treatment .

Reference

- Agrios, G. N. (2010). *Introduction to plant pathology*: Elsevier.
- Al Bashish, D., Braik, M., & Bani-Ahmad, S. (2011). Detection and classification of leaf diseases using K-means-based segmentation and. *Information technology journal, 10*(2), 267-275 .
- Alamdar, F., & Keyvanpour, M. (2012). A new color feature extraction method based on dynamic color distribution entropy of neighborhoods. *arXiv preprint arXiv:1201.3337* .
- Albregtsen, F. (2008). Statistical texture measures computed from gray level coocurrence matrices. *Image processing laboratory, department of informatics, university of oslo, 5* .(5)
- Ali, S. K., & Aydam, Z. M. (2019). Convert Gestures of Arabic Numbers into Voice. *Journal of Computational and Theoretical Nanoscience, 16*(3), 874-879 .
- Babu, M .P., & Rao, B. S. (2007). Leaves recognition using back propagation neural network-advice for pest and disease control on crops. *IndiaKisan. Net: Expert Advisory System*, 607-626 .

- Camargo, A., & Smith, J. (2009). An image-processing based algorithm to automatically identify plant disease visual symptoms. *Biosystems engineering, 102*(1), 9-21 .
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific model development, 7*(3), 1247-1250 .
- Fayad, M., & Mania, A. (2008). Study of date palm leaf spots disease in Basrah and effect of some factors (age of palm, wax content) on infection. *Arab Journal of Plant Protection, 26*(2), 81-88 .
- Hillnhuetter, C., & Mahlein ,A.-K. (2008). Early detection and localisation of sugar beet diseases: new approaches. *Gesunde Pflanzen, 60*(4), 143-149 .
- Howarth, P., & Rüger, S. (2004). *Evaluation of texture features for content-based image retrieval.* Paper presented at the International conference on image and video retrieval.
- Kodituwakku, S. R., & Selvarajah, S. (2004). Comparison of color features for image retrieval. *Indian Journal of Computer Science and Engineering, 1*(3), 207-211 .
- Kumar, A., Srivastava, A., Roy, N., Verma, A & ,.Saini, H. K. (2019). Studies on Cercospora leaf spots disease of groundnut under Bundelkhand region. *IJCS, 7*(4), 2709-2714 .
- Matthews, R. E. F. (2012). *Plant virology*: Elsevier.
- Mehrotra, R. (2013). *Fundamentals of plant pathology*: Tata McGraw-Hill Education.
- Mutlag, W. K., Ali, S. K., Aydam, Z. M., & Taher, B. H. (2020). *Feature Extraction Methods: A Review.* Paper presented at the Journal of Physics: Conference Series.
- Neelamegam, S., & Ramaraj, E. (2013). Classification algorithm in data mining: An overview. *International Journal of P2P Network Trends and Technology (IJPTT), 4*(8), 369-374 .
- Preece, S. J., Goulermas, J. Y., Kenney, L. P., & Howard, D. (2008). A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Transactions on Biomedical Engineering, 56*(3), 871-879 .
- Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Dehne, H.-W., & Plümer, L. (2010). Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Computers and electronics in agriculture, 74*(1), 91-99 .
- Umamaheswari, C., & Bhavani, R. (2018). Texture and Color Feature Extraction Form Ceramic Tiles for Various Flaws Detection Classification. *International Journal on Future Revolution in Computer Science & Communication Engineering, 4*(1), 169-179 .
- Ünay, D., Çataltepe, Z., & Aksoy, S. (2010). Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics): Preface (Vol. 6388.(
- Weizheng, S., Yachun, W., Zhanliang, C., & Hongda, W. (2008). *Grading method of leaf spot disease based on image processing.* Paper presented at the 2008 international conference on computer science and software engineering.

- Zhuang, X., Chen, J., Shim, H., & Bai, Z. (2007). New advances in plant growth-promoting rhizobacteria for bioremediation. *Environment international, 33*(3), 406-413 .