

Identify and Classify Normal and Defects of Prunus_armeniaca Using Imaging Techniques

Amel H.Abbas

AL-Mustansiriyah University,
college of science, computer science
Baghdad, Iraq

dr.amelhussein2017@uomustansiriyah.edu.iq

Marwa A.Shamel

AL-Mustansiriyah University,
college of science, Computer science
Baghdad, Iraq

mi9437911@gmail.com

Abstract: *The Prunus_armeniaca fruit is classified manually in wholesale markets, supermarkets and food processing plants on a normal or defects basis. The aim of this research is to replace the manual sorting techniques using computer vision techniques and applications by proposing techniques for identify and recognitions patterns through the use of 150 fruits of Prunus_armeniaca, 10 for the testing stage in fresh and 10 for testing stage in case of defects. The fruits Prunus_armeniaca collected from growing trees in the large fields of Salah al-Din province\Iraq. The system designed for classification based on the color image taken inside a black box used camera pixel resolution of (13 mega) with a constant intensity of light. . Used K-mean in phase segmentations and only computed 13 features derive statistics from GLCM .classification phase used SVM classify fruit into two class, either (normal or defects) .Results the system success rate reach 100%.The work done using MATLAB R2016a.*

Keywords: Prunus_armeniaca , GLCM ,SVM, K-mean

1. INTRODUCTION

Fruit production plays a vital role in any country Economically. It is part of agricultural production and industrial production (juices, dried fruit, and jam). Fruits are the main source of health because they contain minerals, vitamins, fiber and other nutrients, which form the main part of the balanced diet because it is considered the main factor of health and physical activity. . In general, fruits are of different sizes, shapes and colors, so the process of determining the quality of fruits is very important in export and packaging of foodstuffs, need a vision machine to check the quality, and efficiency of fruit. use of photography in the process of obtaining images because the images are a major source and important in the agricultural sciences through which data and information is recorded as well as the reproduction of this information and reporting use of image processing has prominent effects in the analysis of images and digital treatment works to improve the image of the microscope to the optical range The analysis of images and digital treatments in agricultural sciences has been adapted to distinguish between fruits in terms quality . The analysis and processing of images is also important in terms of time and material costs, where they

are considered more economical than working hands, development of agriculture and crops agricultural of importance and terms of prices and quality all these factors linked to fruit .Prunus_armeniaca from summer fruits that can be eaten fresh or dried, a fruit of high nutritional value and has many benefits of them, maintains the health of the heart and eye and removes toxins body and resists cancer Anemia, stroke. It is thought to be a low-fat fruit with high sugar content. The importance of this type of fruit plays an important role in many of the food processing industry, such as juices, breeder and dryer, so the process of manual sorting of The large quantities is very expensive and waste of time building an incubation based on computer vision is considered more economical in terms of time and effort required[1]. Image segmentation is a process that partition a input Image 'I' into m sub regions. Strong literature on image segmentation is Permissible [2]. Clustering is the process splting in the sense of a particular pattern partition into groups with specific patterns where each section contains a similar pattern .Studied clustering problems on different scales especially in the field of applications including neural networks, AI, and statistics [3 4 5]. K-Means algorithm Is an unsupervised aggregation algorithm that classifies input data points into multiple categories based on the inherent distance from each other. The algorithm assumes that the characteristics of the data constitute a vector area and try to find natural clusters in it [6 7 8]. Machine learning is about the general structure of learning from data use a class of algorithm called (SVM) the most active scan where focus on distinguishing patterns [9 10] Support Vector Machines (SVMs) is distinctive and evolving style of classification compared to other methods such as neural networks, artificial and decision trees because they rely on the theory of automatic learning. They have great features including high precision, athletic and engineering capability, do not need a large number of training samples [11 12]. Pattern recognition system essentially include the following steps: first step, Data acquisition and preprocessing: Here the data from the surrounding environment is taken as an input and given to the pattern recognition system. The raw data is then preprocessed by extracting object of interest from the background. Tow step, Feature extraction: Then the relevant features from the processed data are extracted. These relevant features collectively form entity of object to be classified. Hired, Decision making: Here the

desired operation of classification or on the description of the extracted features.

2. METHODS

This describes the process of analysis and design, This project is divided in few processes which are imaging, pre-processing, feature extraction, database, classification, training and testing. The details of each element are described in figure(1).

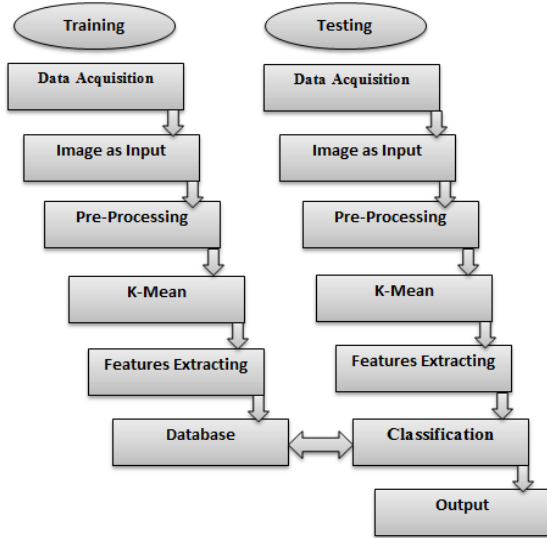


Figure 1 Block diagram of the system

- **Data Acquisition:** The fruit *Prunus armeniaca* was collected from trees in large farms in Salah al-Din/ Iraq collected 150 samples, the colored images of *Prunus armeniaca* is input to the designer system, a small room was designed for photography in order to maintain the conditions for all of the samples used as a stable light source, The camera is a black paper box with a black rectangular base to reduce shadow and reflections. The light is only used on the camera's flash Captured using a phone camera with a pixel of (13 mega). resolution .The camera is 20 cm away from all the samples taken. Capture 55 images for 55 samples are considered normal and 75 images of 75 samples are considered defective from *Prunus armeniaca* . Show figure (2).
- **Preprocessing:** For all input image resize (256 * 256).Followed by image segmentation techniques to remove the background area since our paper only focuses on the fruits. Selsct a K-Mean clustering algorithm for segmentation, RGB image segmentation uses color as homogeneity criteria for splitting. The varieties of K-Mines have been explored for the color image segmentation by paper researchers for decades. The nonlinear conversion of the color components is three layer luminosity layer,Color classification in Space using K meanschromaticity-layer is repersent tow layers ,Color classification in Space using K-mean clustering. Show figure (3).

- **Features Extracting:** Structural method using gray-level co-occurrence matrices (GLCM) and defined statistical measures for texture. Co-occurrence Matrix provides information on the structure and pattern of the texture, and it co-occurrence used to Computational characteristic are calculated as these charactrestic are highly sensitive to differences in light but are nevertheless popular in various areas of texture analysis [13 14 15]. After segment the interest object ,convert RGB segment to gray level and drive features from GLCM computed for each sample either(normal and affected). Statistics features descriptors which are (Contrast, Correlation, Energy, Homogeneity, Mean, Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, and IDM). As $g_{ij(i,j)^{th}}$ entry in GLCM Based on the definition where g is the number of possible gray levels. \sum is sigma (used for the summation) operator, $N_g = L = \text{Number of distinct gray levels in the images}$. μ is mean, σ is variance, $P(I)$ is defined as: [13 14 15 16 17 18 19].

$$p(I) = \frac{\text{number of pixels with gray level } I}{\text{total number of pixels in region}} \quad (1)$$

Expression of GLCM descriptors are:

$$\text{Contrast} = \sum_{ij=0}^{l-1} g_{ij} (i - j)^2 \quad (2)$$

$$\text{Correlation} = \frac{\sum_i \sum_l(i,j) g_{ij} \mu_x \mu_y}{\sigma_x \sigma_y} \quad (3)$$

$$\text{Energy} = \sum_i \sum_j (g_{ij})^2 \quad (4)$$

$$\text{Entropy} = - \sum_i \sum_j g_{ij} \log_2 g_{ij} \quad (5)$$

$$\text{Homogeneity} = \sum_{ij=0}^{l-1} \frac{g_{ij}}{1 + (i-j)^2} \quad (6)$$

$$\text{Variance} = \sum_i \sum_j (i - \mu)^2 \quad (7)$$

$$\mu = \frac{\sum_{i=0}^{g-1} ip(i)}{\sum_{i=0}^{g-1} p(i)} = \frac{\sum_{i=0}^{g-1} ip(i)}{n} = \sum_{i=0}^{g-1} ip(i) \quad (8)$$

$$\text{Skewness} = \sum_{i=0}^{g-1} (i - \mu)^3 p(i) \quad (9)$$

$$\text{kurtosis} = \sum_{i=0}^{g-1} (i - \mu)^4 p(i) \quad (10)$$

$$\text{Standard deviation} = \sqrt{\mu_2 g_{ij}} \quad (11)$$

$$\text{IDM} = \sum_{i=0}^{l-1} \sum_{j=0}^{l-i} \frac{1}{1 + (i-j)^2} p(i - j) \quad (12)$$

$$\text{Smoothness} = 1 - \frac{1}{1 + \sigma^2} \quad (13)$$

$$\text{Root mean square} = \sqrt{\text{mean}(i^2)} \quad (14)$$

Show figure (4).

- **Classification:** In the classification, texture features are derived as that of the training and compared with corresponding feature values stored in the database. use methods SVM (support vector machine) supervised binary classification algorithm works on a set of points based of the spatial dimension to generate $(N-1)$ the separation of those points into two class of features SVM considered the best in finding the best separation line where looking for the nearest point that said by the vector supporting this points such as vector and the best line A break the one on which the work depends. SVM find the points near and draw a line to connect those points by vector subtracting from two classes then declare SVM the best line of separation is a vertical line on the line connection. Show figure (2).

- **Training :** The database was built on the basis of the features extracted for each sample used in the training phase 130 Prunus_armeniaca sample 55 sample normal and 75 defects samples the results were calculated and saved in the database which will be used later in the classification process.
- **Testing:** The test phase used 20 Prunus_armeniaca samples 10 of which were normal and 10 defects were input 20 samples of the system was classified 10 normal and 10 defective. Classification pattern in machine learning are evaluated for their performance by :

$$accuracy = (correctly\ predicted\ class / total\ testing\ class) \times 100\% . \quad (15)$$

The accuracy can be defined as the percentage of correctly classified instances :-

$$(TP + TN)/(TP + TN + FP + FN). \quad (16)$$

standard performance measures:

$$True-PositiveRate=TP/TP+FN \quad (17)$$

$$False-PositiveRate=FP/FP+TN \quad (18)$$

$$True-NegativeRate=TN/TN+FP \quad (19)$$

$$False-Negative\ Rate = FN / FN + TP \quad (20)$$

where FP, TN, TP and FN represent the number of false positives, true negatives, true positives and false negatives, respectively.

For good classifiers, TNR and TPR both should be nearer to 100%. Similar is the case with precision and accuracy parameters. On the contrary, FNR(False-Negative Rate) and FPR(True-Positive Rate) both should be as close to 0% [20] . Show Table(1).

Table 1: Describes the details of the test with accuracy.

Samples Normal	Result Test	Result Test Classify	Samples Defects	Result Test	Result Test Classify
Sample 1	Normal	√	Sample 1	Defects	√
Sample 2	Normal	√	Sample 2	Defects	√
Sample 3	Normal	√	Sample 3	Defects	√
Sample 4	Normal	√	Sample 4	Defects	√
Sample 5	Normal	√	Sample 5	Defects	√
Sample 6	Normal	√	Sample 6	Defects	√
Sample 7	Normal	√	Sample 7	Defects	√
Sample 8	Normal	√	Sample 8	Defects	√
Sample 9	Normal	√	Sample 9	Defects	√
Sample 10	Normal	√	Sample 10	Defects	√

3. RESULTS AND DISCUSSION

It observed that using the algorithm K-Mean to separate the object concerned from the background and by manual input by selecting the class after the separation to calculate the features, which consider the best features that are reliable in the classification of the change is very clear of the values of those features, It (SVM) considered one of the best works and has been given

accurate results and the system was 100%. Show figure (5).

4. CONCLUSION AND FUTURE WORK

The purpose of this method of classification using SVM and training feature of the classification between Prunus_armeniaca (normal and defects) , the results were very satisfactory and accurate classification 100% future business development work to automatic classification .

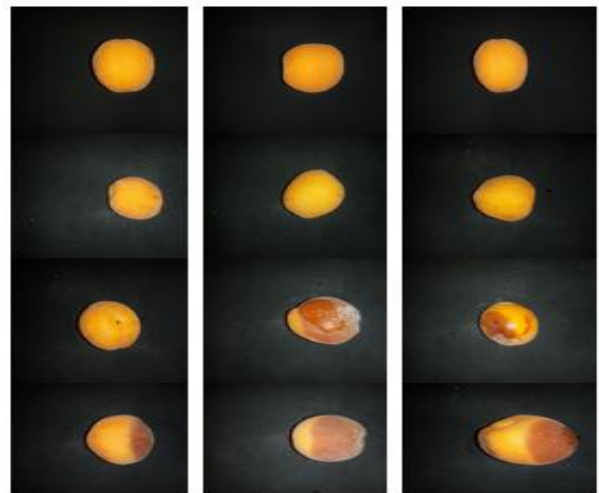


Figure2 Pattern sample normal and defect Prunus_armeniaca

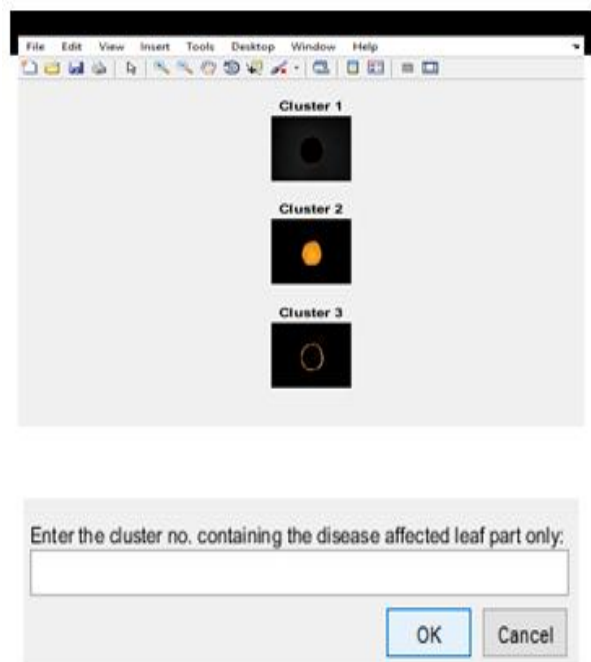


Figure3 Preprocessing to separate the object from the background by using K-Mean, and selecting the class to calculate the features.

0.12604	0.947876	0.827655	0.989935	11.41855	40.4819	1.083686	2.634032	1228.346	1	16.96693	3.793635	255
0.06959	0.913119	0.906341	0.992291	4.726405	23.95485	0.594281	1.925133	508.0299	0.999999	36.12103	5.638411	255
0.088032	0.941128	0.851739	0.991187	8.325509	32.75356	0.929513	2.467487	798.9453	0.999999	21.87019	4.329176	255
0.103261	0.90138	0.910184	0.99186	5.137888	26.23197	0.565908	1.896065	621.0962	0.999999	34.12187	5.514556	255
0.059984	0.878962	0.945389	0.993559	2.582667	17.86064	0.353575	1.197169	287.332	0.999998	65.66459	7.712652	255
0.096739	0.882889	0.932039	0.992629	3.769195	23.04224	0.439951	1.446153	479.5546	0.999999	48.74161	6.658708	255
0.093237	0.918982	0.905809	0.992413	5.312693	28.10578	0.612286	1.571023	642.4881	0.999999	37.65949	5.837407	255
0.113831	0.922973	0.903309	0.990164	6.284005	30.42049	0.603031	1.71935	759.9693	0.999999	28.70644	5.081346	255
0.129365	0.908077	0.901806	0.991024	6.007968	29.98724	0.637041	1.745446	735.7892	0.999999	32.55335	5.405962	255
0.081735	0.910758	0.894082	0.991936	5.172379	26.41942	0.688083	2.01234	613.0498	0.999999	36.44295	5.705545	255
0.096321	0.900519	0.898825	0.990858	5.134608	25.56743	0.645641	1.778606	525.8557	0.999999	34.17503	5.501641	255
0.0471	0.970312	0.926593	0.996299	5.3708	30.96173	0.493991	1.2108	729.6207	0.999999	42.03441	6.228905	255
0.09253	0.924119	0.910325	0.992689	5.576955	27.9893	0.562441	1.584639	612.1716	0.999999	32.29989	5.369594	255
0.109237	0.906805	0.922282	0.993557	5.016997	27.23289	0.49454	1.398394	583.1657	0.999999	37.54822	5.814019	255
0.098892	0.912281	0.90814	0.992289	5.491141	27.39523	0.575357	1.591811	573.611	0.999999	32.52044	5.376906	255
0.097285	0.935576	0.896199	0.992845	6.751493	30.45995	0.652302	1.812361	745.0796	0.999999	24.56794	4.679092	255
0.126715	0.939044	0.877231	0.993243	8.939872	36.2074	0.769818	2.072937	1033.026	0.999999	18.87894	4.091118	255
0.085912	0.943561	0.893098	0.992956	6.972987	31.19781	0.6778	1.811779	766.5351	0.999999	25.02332	4.704057	255
0.096546	0.943271	0.892963	0.993484	7.35112	32.66622	0.680993	1.815483	839.9716	0.999999	24.0722	4.624653	255
0.08045	0.864307	0.935432	0.992014	3.120464	20.10162	0.405683	1.608934	378.3294	0.999998	57.02666	7.186665	255
0.055261	0.874604	0.940966	0.993116	2.564987	17.73806	0.370694	1.437159	290.8015	0.999998	66.10129	7.763438	255
0.049574	0.983397	0.834164	0.994982	11.55695	44.23133	1.065909	2.590324	1459.502	1	18.36632	4.038818	255
0.134747	0.940941	0.891031	0.990364	8.493008	36.73668	0.672682	1.872403	1121.889	0.999999	20.11941	4.282279	255
0.092241	0.882601	0.919387	0.991587	4.202048	22.45351	0.503343	1.539236	427.8993	0.999999	36.63861	5.71579	255
0.092627	0.883734	0.919833	0.992011	4.297595	22.88895	0.500165	1.50788	433.8289	0.999999	35.44012	5.639091	255
0.083759	0.896787	0.938693	0.993034	3.512107	22.2897	0.399516	1.343394	452.4032	0.999998	52.93371	6.928291	255

Figure 4 Show the features computed for each Prunus_armeniaca (Normal and Defects).

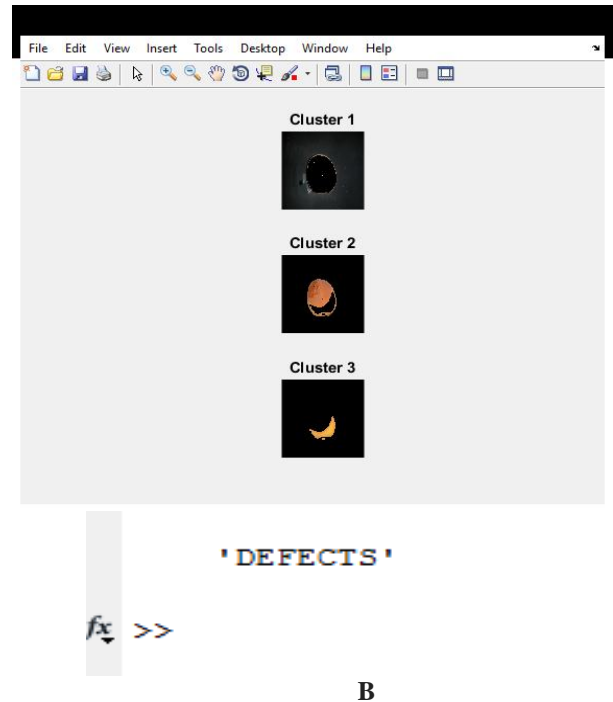
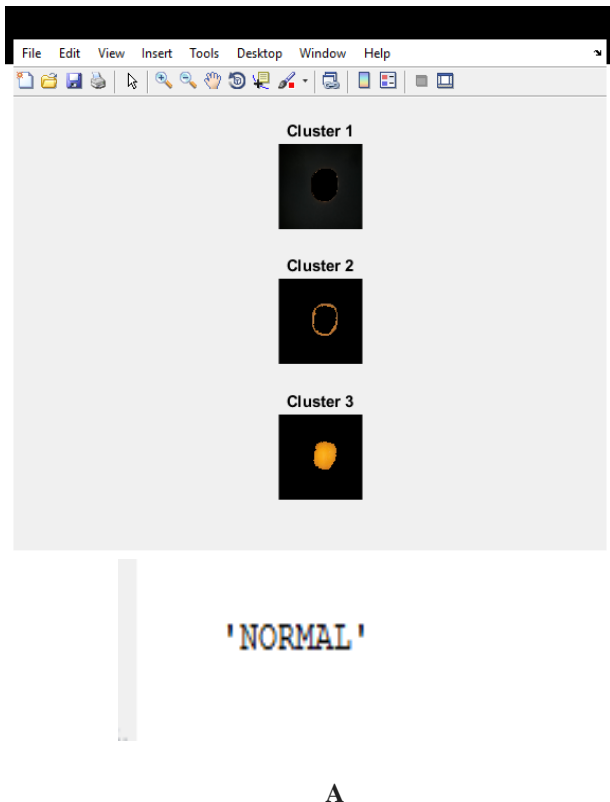


Figure 5 Show the result testing (A) Normal (B) defect

5.REFERENCE

- [1] Apricot , http://www.10_Impressive_Apricot_Benefits_|Organic_Facts_, (2017).
- [2] H.Tariq and M.Aqil Burney, "K-Means Cluster Analysis for Image Segmentation", International Journal of Computer Application, vol. 4,no.4,2014.
- [3] A.Khaled , R.Sanjay , and S.Vineet ,” An Efficient K-Means Clustering Algorithm”,1997.
- [4] T. Kanungo , M. David Mount,S. Nathan Netanyahu,D. Christine Piatko,S. Ruth ,Y.Angela Wu, “An Efficient k-Means Clustering Algorithm: Analysis and Implementation”, IEEE Transaction on Pattern Analysis and Machine Intelligence,vol.24, pp 881-892 ,2002.
- [5] V.Deepali,T. Shweta,M.Geetika, “ Normalization based K means Clustering Algorithm”,International Journal of Advanced Engineering Research and Science (IJAERS), vol.2,2015.
- [6] T.Suman ,M. Avi, “Image Segmentation using k-means clustering, EM and Normalized Cuts”, University Of California Irvine ,2008.
- [7] L.Jinhua,S. Shiji,Z. Yuli, and Z. Zhen, “ Robust K-Median and K-Means Clustering Algorithms for Incomplete Data”, Mathematical Problems in Engineering ,2016.
- [8] M.Umamaheswari, P. Isakki , “ Myocardial Infarction Prediction Using KMeans Clustering Algorithm ” ,International Journal of Innovative Research in Computer and Communication Engineering, vol. 5, 2017.
- [9] L.Martin , “ A Simple Introduction to Support Vector Machines”,CSE 802,2011.

- [10] M.David, " Support Vector Machines: The Interface to libsvm in Package e1071", David.Meyer@R-Project.org, 2017.
- [11] Z.Yudong and W. Lenan, "Classification of Fruits Using Computer Vision and a Multiclass Support Vector Machine", sensors,vol.12,2012.
- [12] D.Matthew Sacchet, P.Gautam, C.Lara Foland-Ross, M. Paul Thompsonm, And H. Ian Gotlib, "Support Vector Machine Classification Of Major Depressive Disorder Using Diffusion-Weighted Neuroimaging And Graph Theory", Free PMC Article , vol. 6,2015.
- [13] S.elvarajah and S.Kodituwakku, " Analysis and Comparison of Texture Features for Content Based Image Retrieval", International Journal of Latest Trends in Computing , vol. 2,2012.
- [14] G.Shantala, P. Jagadeesh , S. Shivanand , " Role of GLCM Features in Identifying Abnormalities in the Retinal Images",MECS,vol.6, pp 45-51,2015.
- [15] S.Kanchan, K. Aditi,S. Kulbeer , " GLCM and its Features",International Journal of Advanced Research in Electronics and Communication Engineering, vol. 4, 2015.
- [16] M.Musa Mokij and H.Pui Ngian, "Vegetable Recognition Based Texture Features",Online,2011.
- [17] A.Goshtasby, " Advances in Computer Vision and Pattern Recognition", Springer London Dordrecht Heidelberg New York, 2012.
- [18] N. Aggarwal, K. Agrawal, "First and Second Order Statistics Features for Classification of Magnetic Resonance Brain Images", Journal of Signal and Information Processing, Published Online(<http://www.SciRP.org/journal/jsip>),pp, vol.3,2012.
- [19] W.Nicholas, " An Investigation into Texture Features for Image Retrieval", University of Bath ,2007.
- [20] A.Baratloo, M.Hosseini,A. Negida,G. Ashal, " Simple Definition And Calculation Of Accuracy Sensitivity And Specificity", Emergency , vol. 3, 2015.