IOSR Journal of Pharmacy and Biological Sciences (IOSR-JPBS)
e-ISSN:2278-3008, p-ISSN:2319-7676. Volume 12, Issue 5 Ver. VI (Sep. – Oct. 2017), PP 24-31
www.iosrjournals.org

Classification of Lung Diseases Based on K-mean Clustering

*Alyaa Hussein Ali, Zahra Jaafer Abod Jasim, Maysaa Raba Naeemah
University of Baghdad, College of Science for Women
Corresponding Author: Alyaa Hussein Ali

Abstract: *this paper give indication about the automatic system which can be used in the analysis of the computed tomography(CT) in the detection of the lung diseases. Which are (lung cancer, tuberculosis, pneumonia, lymphatic drainage… etc).The chest computerized tomography (CT) is considered in this search to detecting the lung tumors. The objective of this paper is to support efficient methods to predicated lung diseases. Different proceed are used to produce the best result, the first step is to remove the noise from the CT image this has been done by using the wiener filter. The unsupervised process which is the k-mean classification technique which is used to isolated the **lymphatic and pneumonia** the second order statistical feature which is calculated from the co-occurrence matrix is used to describe the **lymphatic and pneumonia** features.*

Keyword*: winner filter, texture analysis, k-mean, second order statistical features.*

Date of Submission: 02-10-2017 Date of acceptance: 14-10-2017

I. Introduction

The term lung disease can be consider as many disorders affecting the lungs such as asthma, chronic obstructive pulmonary (COPD) disease, infections such as tuberculosis, influenza, lung cancer, pneumonia and other breathing problems. Lung diseases signs can differ by the type of the affected disease. Common signs are trouble in breathing, shortness of breath, feeling, decreased ability to exercise, a cough that won't go away, coughing up blood or mucus, pain or discomfort when breathing in or out. Medical imaging is the technique and process used to create images of the human body for clinical purposes , diagnosis and analysis In this paper, an automated approach for classification of the lung diseases using CT images is presented. The lung CT image is engaged as the input. The original image is transformed to gray scale image. After that, removal of the noises and contrast enhancement is done for obtaining the enhanced images. The winner filter is applied to remove the noises and the preprocessed images are given as input for feature extraction where the useful features of the images are extracted and the extracted features are selected by the using the unsupervised classification, the classifiers are used to classify the datasets in to relevant datasets and the performance measures are evaluated for the datasets [1]. Diseases such as lung cancer, old or new pneumonia, tuberculosis, emphysema, lymphatic and chronic obstructive lung diseases (COPD).The human lungs are subdivided into five lobes that are separated by visceral pleura called pulmonary. There are three lobes in the right lung, namely upper, middle and lower lobe. The right upper and right middle lobe are divided by the right minor fissure whereas the right major fissure delimits the lower lobe from the rest of the lung. In the left lung there are only two lobes, the upper and the lower lobe, that are divided by the left major fissure. Lung lobe segmentation is relevant in clinical applications particularly for treatment planning. The location and distribution of pulmonary diseases are important parameters for the selection of a suitable treatment [2].

Test Images: the images which are used in this search is taken as a CT scan images three case for lymphatic and three case for pneumonia. Figure(1) shows the CT lung image of lymphatic case. while, figure(2) shows the pneumonia case**.**

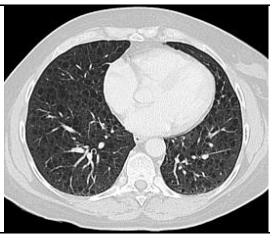
Image No.1	Image No.2	Image No.3
		

Figure (1) shown CT lung image of lymphatic case.

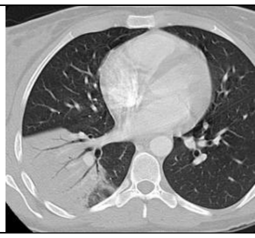
Image No.1	Image No.2	Image No.3
		

Figure (2) shown CT lung image of pneumonia case.

II. Methodology

1- Take the CT image three lymphatic and three pneumonia cases .
2-Using the wiener filter to remove the noise which has corrupted the signal.
4- Convert the CT images from RGB to HSV.
3- Using the unsupervised classification (K-mean).
4- Using second order statistical features (co-occurrence matrices) for statistical analysis.

Wiener Filter: The goal of the wiener filter is to filter out noise that has corrupted a signal. It is based on a statistical approach. However, the design of the wiener filter takes a different approach. One is assumed to have knowledge of the spectral properties of the original signal and the noise, one seeks the linear time-invariant filter whose output would come as close to the original signal[3]. The wiener filter is optimal filter for remove noise from image, wiener filter also known as the least mean square filter, is given by the following expression [3]:

$$\hat{F}_{(u,v)} = \left[\frac{H_{(u,v)^*}}{\left|H_{(u,v)}\right|^2 + \left[{S_{n\,(u,v)}}/{S_{f\,(u,v)}}\right]}\right] G_{(u,v)} \qquad ....1$$

Where $H_{(u,v)}$is the degradation function (* indicates complex conjugate) and $G_{(u,v)}$ is the degraded image. The functions $S_{f\,(u,v)}$ and $S_{n\,(u,v)}$are the power spectra of the original image (prior to degradation) and the noise if the noise is zero.

Color image: After removing noise from the original image using the wiener filter, it is converted gray scale image to color image (HSV). This good result giving for classification k-mean. Whereas HSV varies the hue component of the hue-saturation-value color model. The colors begin with red, pass through yellow, green, cyan, blue, magenta, and return to red [4].

RGB to HSV Conversion: The HSV stands for the Hue, Saturation and Value. The value represents intensity of a color, which is decoupled from the color information in the represented image. The hue and saturation components are intimately related to the way human eye perceives. The transformation equations for RGB to HSV color model conversion is given below [5].

$$V = \max(R, G, B) \dots\dots\dots (2)$$

$$S = \frac{V - \min(R, G, B)}{V} \dots\dots\dots (3)$$

$$H = \frac{G - B}{6S} \qquad , if V = R \dots\dots\dots (4\,)$$

K-means Classification: K-means algorithm is the most popular partitioning based clustering technique. It is an unsupervised algorithm which is used in clustering. k- mean clusters are formed by finding out the data points nearest to the clusters. The steps of the K-means algorithm[6]:

1. Choose k number of points randomly and make them initial centroids.
2. Select a data point from the collection, compare it with each centroid and if the data point is found to be similar with the centroid then assign it into the cluster of that centroid.
3. When each data point has been assigned to one of the clusters, calculate the value of the centroids for each k-number of clusters.
4. Repeat steps 2 and 3 until no data point moves from its previous cluster to some other cluster.

Co-occurrence Matrices: An important and powerful statistical texture analysis algorithm is the co-occurrence matrices. The co-occurrence matrix is a two dimensional histogram, which indicated to number of times that pairs of intensity values occur in a given spatial relationship [7].The co-occurrence matrices are constructed by considering that every pixel have eight neighbors (horizontally, vertically and diagonally at 45 degrees). It also assumed that the matrix of relative frequencies of gray levels co-occurrence can specify the texture-context information. Some of the texture measures can be obtained from these matrices, (like homogeneity and the contrast) [8]. The texture is specified by the matrix of relative frequencies of co-occurrence p(i,j), which indicate the number of times that each two neighboring pixels of an image, separated by a distance (d), will have gray tone (i) for one pixel and (j) gray tone for the other pixel. Such matrices of gray tone spatial dependence frequencies are the functions of the angular relationship between the neighboring pixels, as well as a function of the distance between them[9]. The co-occurrence matrices are based on the repeated occurrence of the gray level configuration in the considered texture. This configuration varies rapidly in fine textures, more slowly in coarse textures.

Second-Order Statistics Features:

- Correlation : It measures how a pixel is correlated to its neighborhood pixels. Its value lies between (-1 and +1).Its value is(-1) for perfectly negatively correlated image and (+1) for positively correlated image[10].

  $Correlation = \sum_{i=0}^{G-1} \frac{i\, p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$ ……(5)

  μ_x, μ_y: mean value in the x and y direction.

  σ_x, σ_y: variance of x,y

- Homogeneity :It gives the distribution value of the closeness of elements of the gray level co-occurrence matrix. It gives the value between the range of 0 and 1[10].

  $Homogeneity = \sum_{i,j=0}^{G-1} \frac{p(i,j)}{1+(i-j)^2}$ …..(6)

- Contrast : It gives the local variations in the gray level co-occurrence matrix. It determines the intensity difference between a pixel and its neighborhood[10].

  $Contrast = \sum_{i,j=0}^{G-1} (i-j)^2\ p(i,j)$….(7)

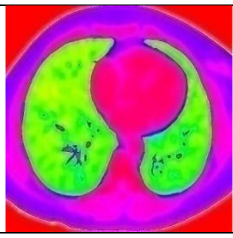
- Energy: it is the sum of squared elements. Its range is from 0 to 1 [10].

$$E = \sum_{i=0}^{G-1} (p(i,j))^2 \quad \ldots\ldots\ldots\ldots\ldots \quad (8)$$

Where p(i) the probability density of occurrence of the intensity, as determined from the histogram with total number of pixels in the image is given as[11].

K-mean: Clustering is a way to separate groups of objects. the K-means clustering treats each object as having a location in space. It finds partitions such that objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. K-means clustering requires that you specify the number of clusters to be partitioned and a distance metric to quantify how close two objects are to each other.

For every object in input, k-means returns an index corresponding to a cluster. The cluster center output from k-means will be used later in the demo. Label every pixel in the image with its cluster index. Using pixel labels, can separate objects by color. Figure(3,4,5) shows the winner filter, converted image to the HSV and the classes of K-mean clustering analysis for the lymphatic case . Figure(6,7,8) shows the winner filter, converted image to the HSV and the classes of K-mean clustering analysis for the pneumonia case .

Wiener filter for image No. 1	**Color image (HSV)**
	

k-mean		
Class.1	**Class.2**	**Class.3**

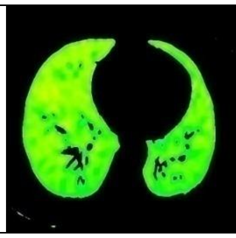
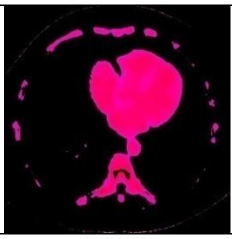
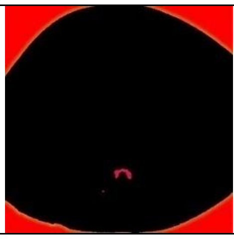
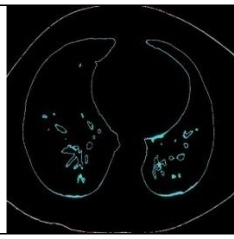
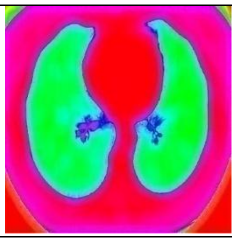
		
Class.4	**Class.5**	**Class.6**
		

Figure (3) shown the k-mean for the lymphatic case image .

Wiener filter for image No.2	**Color image (HSV)**
	

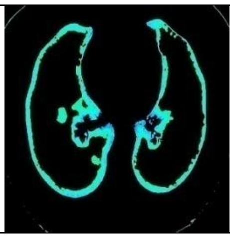
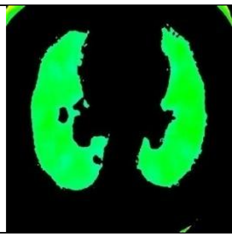
k-mean		
Class.1	**Class.2**	**Class.3**
		
Class.4	**Class.5**	**Class.6**
		

Figure (4) shown the k-mean for the lymphatic case image .

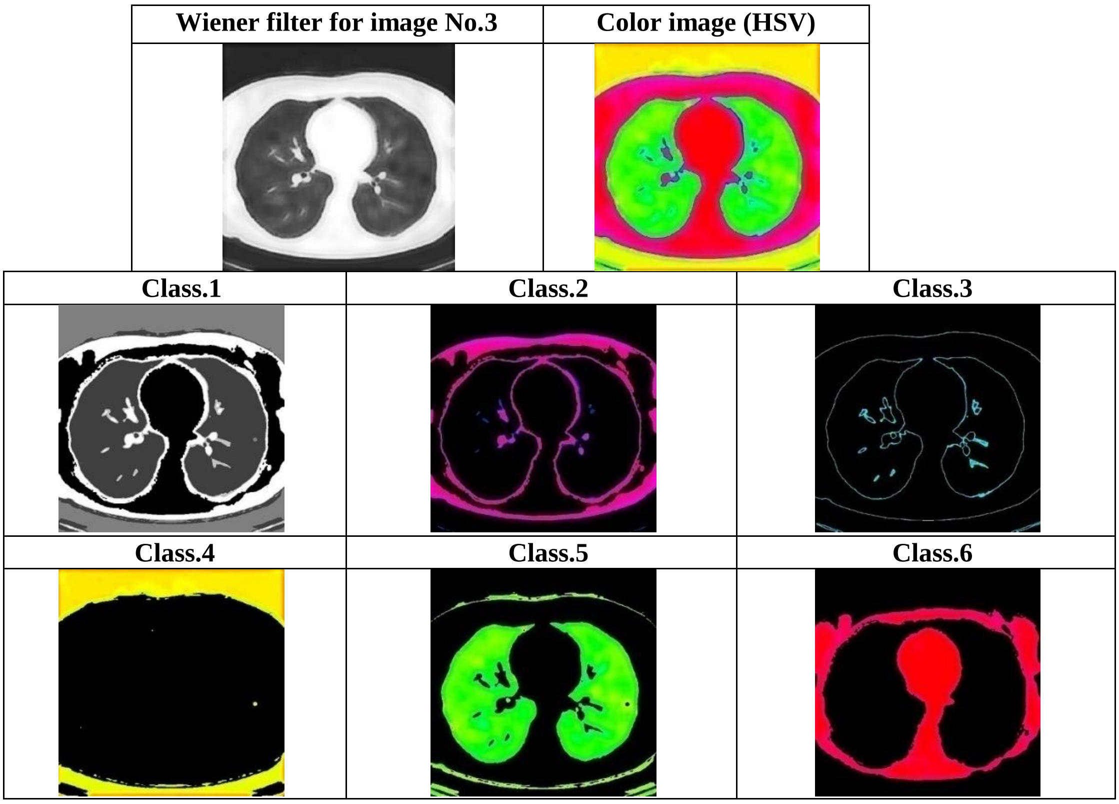


Figure (5) shown the k-mean for the lymphatic case image .

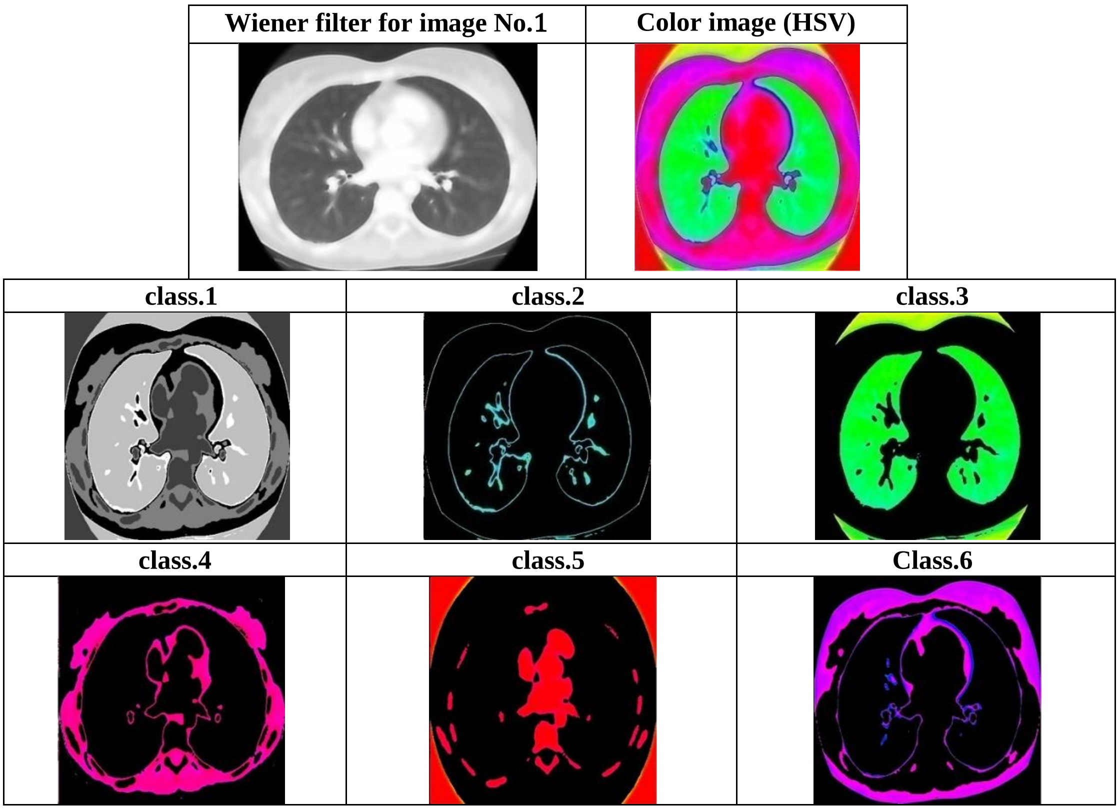


Figure (6) shown the k-mean for the pneumonia case image .

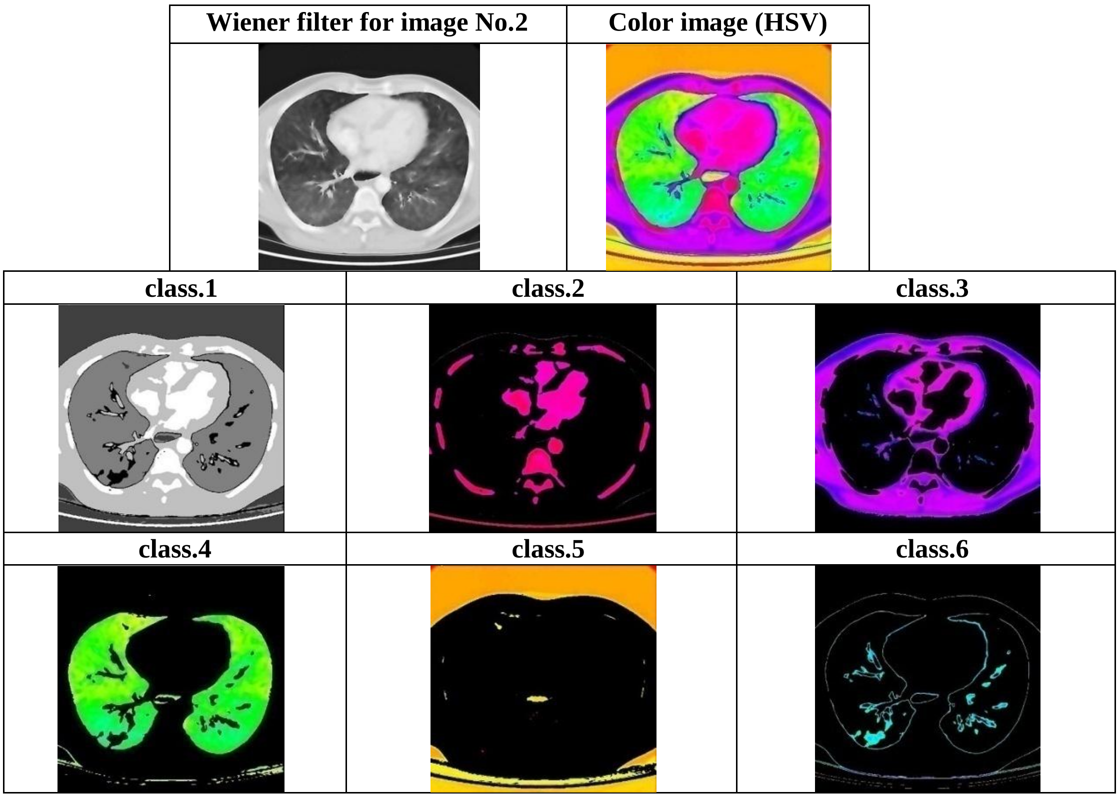


Figure (7) shown the k-mean for the pneumonia case image .

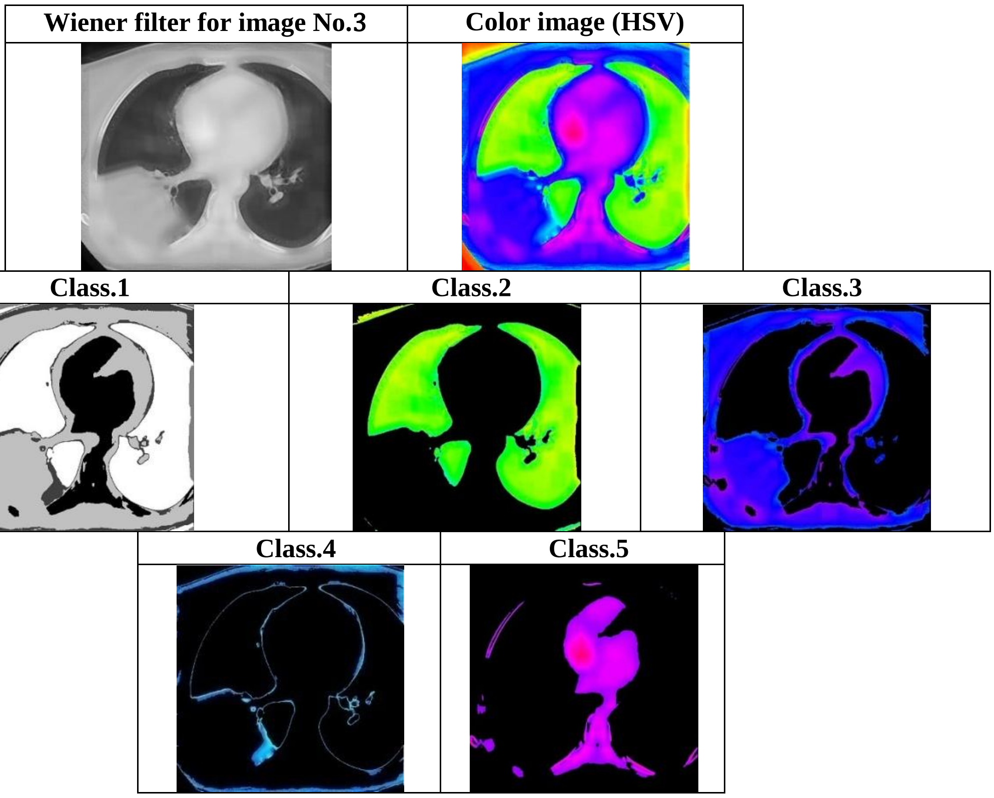
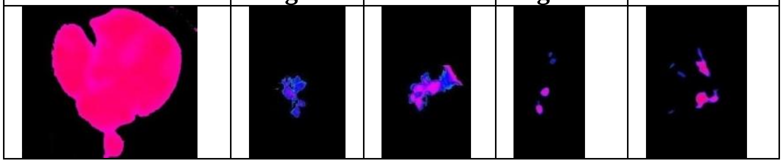


Figure (8) shown the k-mean for the pneumonia case image .

Image No.1	Image No2		Image No.3	
	right	Left	Right	left



Figure(9) shown the lymphatic case.

Table (1) gives the statistical features second order lymphatic case.

Image No.	second order co-occurrance matrix			
	Contrast	**Correlation**	**Energy**	**Homogeity**
Image No.1	0.0508	0.9826	0.3846	0.9783
Image No.2 right	0.0232	0.7978	0.9623	0.9938
Image No.2 left	0.0291	0.8991	0.9364	0.9916
Image No.3right	0.0053	0.8976	0.9882	0.9983
Image No.3 left	0.0120	0.9039	0.9739	0.9961
Average value	0.02408	0.8962	0.84908	0.99162

Image No.1		Image No.2		Image No. 3
Right	left	right	Left	

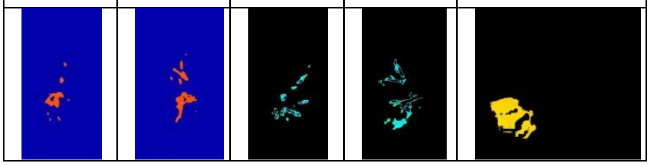


Figure (10) shown the pneumonia case.

Table (3) gives the statistical features second order pneumonia case

Image No.	Second order co-occurrence matrix			
	Contrast	**Correlation**	**energy**	**Homogeneity**
Image No.1 right	0.0298	0.8323	0.9616	0.9936
Image No.1 left	0.3187	0.7101	0.9111	0.9774
Image No.2 right	0.1518	0.7696	0.9446	0.9860
Image No.2 left	0.2380	0.7607	0.9184	0.9799
Image No.3	0.1377	0.9385	0.9202	0.9896
Average value	0.1752	0.80224	0.93118	0.9853

III. Discussion

Figure(9) shows the region of interest and table (1), gives the value of the statistical features for the lymphatic case, while figure(10) shows the region of interest for the pneumonia case, table(2) shows the statistical feature for the second order features. The features which are studies in this paper are the contrast, correlation, energy and homogeneity, the contrast is higher for the pneumonia than that for lymphatic, while the correlation for the lymphatic is higher than that for the pneumonia, so is the homogeneity, this means that lymphatic cells are more pure and correlated than the pneumonia because the homogeneity gives indication about the purity of the texture and the correlation indicated that the pixels in the neighbor have the same texture.

Reference

[1]. C.Bhuvaneswari1 , P.Aruna2 , D.Loganathan3," Classification of Lung Diseases by Image Processing Techniques Using Computed Tomography Images" International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-4 Number-1 Issue-14 March-2014.

[2]. Ganesh Sable1 , Harsha Bodhey2"Lung Segmentation and Tumor Identification from CT Scan Images Using SVM",International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Impact Factor (2012): 3.358.

[3]. 3. Prewitt , J. 1970, "Object Enhancemet and Extraction, Picture Processing and Psychopictorics", (B. Lipkin and A. Rosenfeld, Ed.), NY, Academic Pres.

[4]. Chang, J. D.; Yu, S. S.; Chen, H. H., and Tsai, C. S. 2010. "HSV-based Color Texture Image Classification using Wavelet Transform and Motif Patterns". Journal of Computers. Vol.20, No.4, pp:64-69.

[5]. James, I. S. P. 2013. "Face Image Retrieval with HSV Color Space Using Clustering Techniques". The Slovenska industrija jekla Transaction on Computer Science Engineering & its Application. Vol.1, No.1, pp:17-20.

[6]. Balaji, T.and Sumathi, M. 2013. "Relational Features of Remote Sensing Image Classification using Effective K-Means Clustering ". International Journal of Advancements in Research & Technology. Vol.2, No.8, pp:103-107.

[7]. Arfan. M., Jaffar and Latif. "Classification and Segmentation of Brain Tumor Using Texture Analysis". Recent Advance in Artificial intelligence, Knowledge engineering and data bases.

[8]. Alyaa, H. and Alaa, N. M. and Asraa, R. 2012, "Classification of Breast Tissue for mammograms images using intensity histogram and statistical methods", A Refereed Science Journal issued by the college of science university of Baghdad, vol 53, No.4.

[9]. Matipietikainen, A.Rosenfeld, December 1981, "Image Segmentation by Texture Using Pyramid Node Linking" IEEE Transactions on System, Man, and Cybernetics, Vol. SMC-11, No. 12.

[10]. L. Van Gool ,P. Dewaele ,and A. Oosterlink, 1985, "Texture Analysis". Computer Vision, Graphics and Image Processing ,29:336-357.

[11]. Julesz, B.1981, "A Theory of Preattentive Texture Discrimination Based on First-Order Statistics of Textons", Biological Cybernetics, 41, pp. 131-138.

Alyaa Hussein Ali. "Classification of Lung Diseases Based on K-Mean Clustering." IOSR .Journal of Pharmacy and Biological Sciences (IOSR-JPBS) , vol. 12, no. 5, 2017, pp. 24–31