Unsupervised Image Classification Technique for Lung Diseases Diagnosis using Fuzzy C-mean Clustering algorithm and Q-Matrix

Alyaa Hussein Ali

College of Science for Women - Baghdad university - Physics Department

Received : 7 December 2015 ; Accepted : 22 March 2016

Abstract

The study aim to classify three x-ray lung images depending on the fuzzy c- mean classifier process and on the Q-matrix. The fuzzy c-mean (FCM) is unsupervised process helps to gather the pixel which carry the same features in one or in two classes. The Q-matrix is gray level matrix based on some features which may be used to help the user in identifying the identity of the texture. Also, Gabor texture features has been used to calculate the textural statistical features in this search five cases have been used which are (Cancer, Echinococcosis, Lymphatic, Pneumonia, Tuberculosis).

keyword: lung dieses, Q-matrix, Gabor texture features , FCM, X-ray Images .

تقنية تصنيف الصور غير الموجهه لتشخيص امراض الرئه بأستخدام خوارزمية العنقدة المضبب ومصفوفة Q

علياء حسين علي

كلية العلوم للبنات - جامعة بغداد - قسم الفيزياء

المستخلص

تهدف هذه الدراسه الى تصنيف اورام الرئه من الصور الماخوذه بواسطة اشعة السينية بالاعتماد على طريقة التصنيف الغير الموجه ومصفوفة Q وطريقة تحليل النسيج المعروفه باسم كابور التي يمكن من خلالها حساب الخصائص الاحصائيه .ان طريقة العنقده الضبابيه هي واحده من طرق التصنيف الغير موجه والتي تساعد على تجميع البكسلات التي تحمل نفس الخصائص بصنف واحد او اكثر. ان مصفوفة Q هي مصفوفة تدرج لوني تعتمد على الخصائص الاحصائيه التي تساعد المستخدم لتحديد هوية النسيج. في هذا البحث خمس حالات مرضيه تم التعامل معها وهي السرطان ، اكياس مائيه ،عقد لمفاويه، الالتهاب الرئوي ،السل الرئوي.

الكلمات المفتاحية : امراض الرئه، مصفوفة Q ، الخصائص الاحصائيه لكابور، عنقدة C الضبابيه، صور اشعة السينية.

Introduction

The texture is a composition of micro patterns and it an important line in image processing . The texture gives indication about the changes which may occur in the composition of the material , the early detection of the defect which may happened in the tissue can protect the other healthy part. in this search the x-ray images has been used to classify the lung disease , the x-ray images is the most known oldest images which is used in the medical process . The obtained digital x-ray images from the radiographs is a common practice to find the information and reduce the number of rejected x-ray [1], with the help of x-ray the digital image processing play important part in detecting the tumors which may attacked the lung. in this search the unsupervised classification process which is the fuzzy C- mean has been used to separated the gray and white color in the images . The clusters can be used to measure the affected tissue or the defect which occur in texture of an image [2]. The FCM is a method belong to pattern recognition which helps one part of data to belong to one or two cluster in the image with the help of FCM one can separated the tumor part from the healthy part . Applying the Q- matrix to the tumor part to calculate the texture features which is helpful in classifying the lung tumor . Since the Q-matrix depend on the occurrence of gray value it measure the similarity between the pixel in one image, some

features which can be used to obtain information about the classified part can be calculated from the Q-matrix. The early study of medical images were the x- ray also, the Gabor filter has been used to obtain statistical features. The x-ray gives clear information about the lung tumor. The test images including five cases three cancer images, three Echinococcosis, three Lymphatic, three Pneumonia and three Tuberculosis taken from Digestive system diseases hospital. In 2012 M.B. Kuchanur discover a fast and effective method to detect the lung diseases , it separated the cancer images and distinguish it from other tumor [3], in 2013 S.Sivakumar and C.Chandrasekar used a technique based on weighted fuzzy which based on clustering out for lung cancer image [4].

Unsupervised Classification: The Fuzzy C-mean is one of the pattern recognition tools used to separate the image classes. It can be used to detect the defected region in an image depending on measuring the similarity between pixels, it allows the image data to belong into class or more. The algorithm can be display as [5,6].

$$d_m = \sum_{i=1}^{N} \sum_{j=1}^{C} K_{ij}^{a} \left\| x_i - y_j \right\|^2 \qquad 1 \le a < \infty \qquad (1)$$

Where d_m based on minimizing the function.

N= number of vectors in pattern.

C=Number of classes.

a= has a value between 1 and infinity.

K_{ij} =measure the member ship degree of x_i in class j, and it value is between zero or equal to zero and one.

x_i =is the i^{th} data set vector.

V_j =is the center of the cluster j.

$$K_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\|x_i - V_j\|}{\|x_i - V_k\|}\right)^{\frac{2}{m-1}}} \qquad (2)$$

$$V_j = \frac{\sum_{i=1}^{N} K_{ij} \times x_i}{\sum_{i=1}^{N} K_{ij}} \qquad (3)$$

Q-Matrix: The Q- matrix used to measure the similarity in the repeated pixels which have the same gray level value. It is a matrix to perform the texture feature. The operation of this matrix based on using a window such as (3×3) mask or (5× 5) mask. If the window is 3×3 then the sub image elements are 9, if the mask are 5× 5 then the sub image elements are 25 and the range is from (0-24). There are some features which can be obtained from this matrix [7].

1- Small number emphasis (SNE) which can be obtained as follows:

$$SNE = \frac{\sum_{i=0}^{8} \sum_{j=1}^{9} Q(i,j)/j^2}{\sum_{i=0}^{8} \sum_{j=1}^{9} Q(i,j)} \qquad (4)$$

if the mask is 3×3. where (i) is the row and (j)is the column.

2- Large number emphasis (LNE) can be obtained by:

$$LNE = \frac{\sum_{i=0}^{8} \sum_{j=1}^{9} Q(i,j) \times j^2}{\sum_{i=0}^{8} \sum_{j=1}^{9} Q(i,j)} \qquad (5)$$

3- Number non uniformity (NNU) can be obtained as :

$$NNU = \frac{\sum_{i=0}^{8} \left[\sum_{j=1}^{9} Q(i,j)\right]^2}{\sum_{i=0}^{8} \sum_{j=1}^{9} Q(i,j)} \qquad (6)$$

4- Secondary moment(SM)

$$SM = \frac{\sum_{i=0}^{8} \sum_{j=1}^{9} Q(i,j)^2}{\sum_{i=0}^{8} \sum_{j=1}^{9} Q(i,j)} \qquad (7)$$

Gabor texture features: The Gabor texture filter used for different type of images. It main use is to the so eliminated and reflected the orientation and also the spatial effect of the texture. The Gabor texture feature can represented as [8].

$$F(x,y) = EXP\left(-\frac{1}{2}\left(\frac{X^2}{\sigma_x^2}+\frac{X^2}{\sigma_y^2}\right)\right) COS(2\pi f + \emptyset) \qquad (8)$$

f= is the frequency of orientation.

$\emptyset$= equal to zero in order to get symmetrical filter and equal to $\frac{-\pi}{2}$ to get anti-symmetrical.

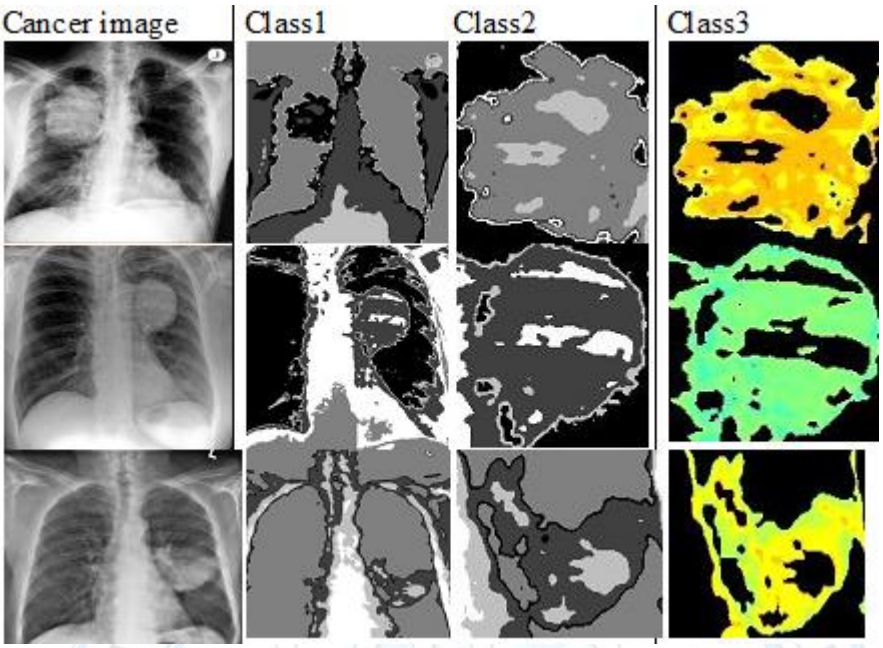
x,y are coordinates

$$\sigma_x = \frac{\sqrt{ln2}\,(2^{Bf}+1)}{\sqrt{2}\pi f(2^{Bf}+1)}, \qquad \sigma_y = \frac{\sqrt{ln2}}{\sqrt{2}\pi f\ \tan(\frac{B_\theta}{2})}, \qquad (9)$$

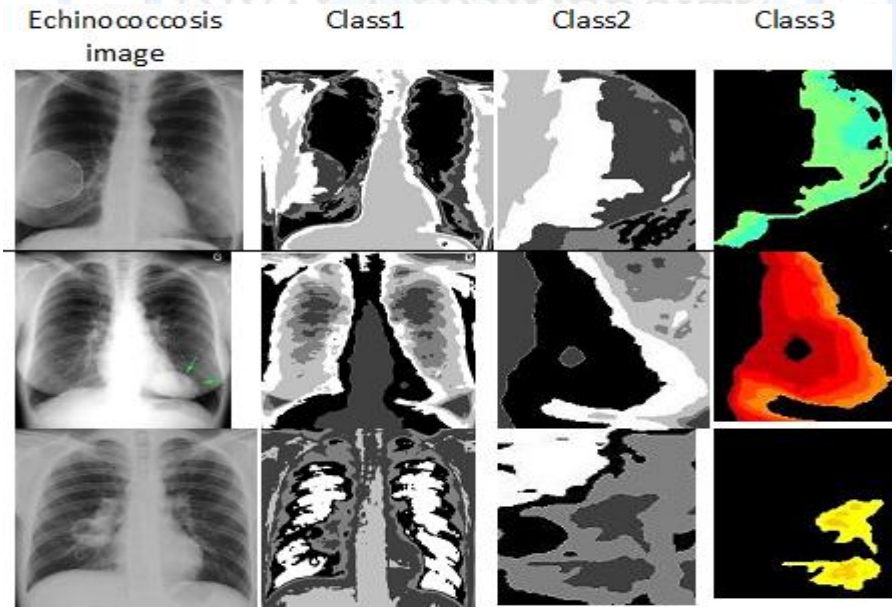
$B_\theta = 30^\circ$ which is the angular band width.

Bf=1 which is the frequency band width.

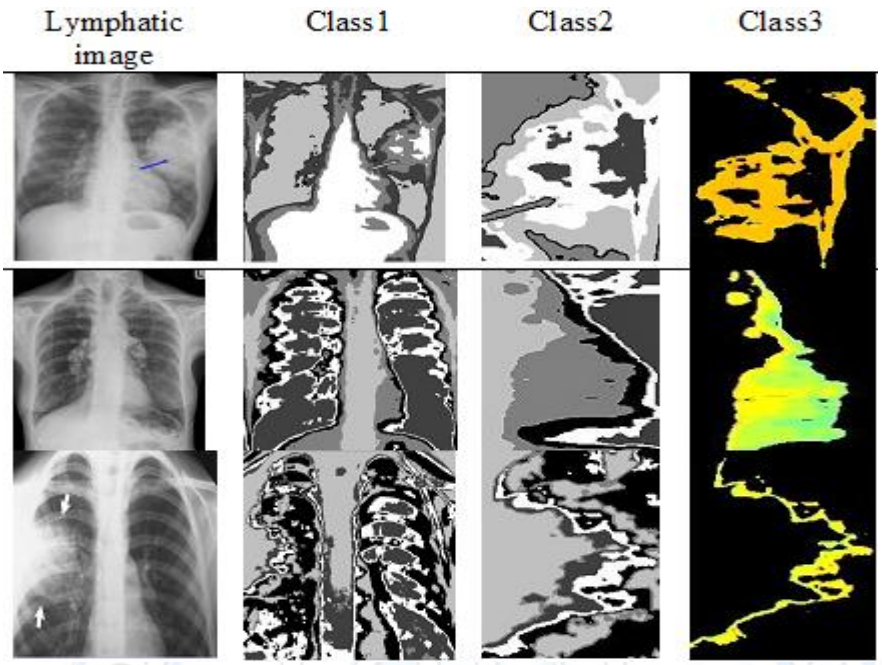
Methodology: The first step is to separate the tumors which are (cancer, Echinococcosis, Lymphatic, Pneumonia and Tuberculosis) from the other healthy tissue in the lung. The lung is a complex tissue since it conation the heart and the rips appear in the x-ray image which make the processing of the image difficult, using the (FCM) helps to separated the tumor from healthy tissue. After the tumor has been spared from the other part of the lung, the rest will be implemented in Q-matrix to measure the emphasis for the tumor, also measuring the non-uniformity and the secondary moment also, the Gabor filter is used to study the texture of lung tumors. Figures (1,2,3,4,5) shows the FCM for lung tumor which appear clearly, it is easy to separated the healthy tissue from the defected one. Tables (1,2,3,4,5) shows the statistical features obtained from Q- matrix, table(6) shows the average value for the three samples of images for each case, table(7) shows the Gabor statistical features value for each diseases case.



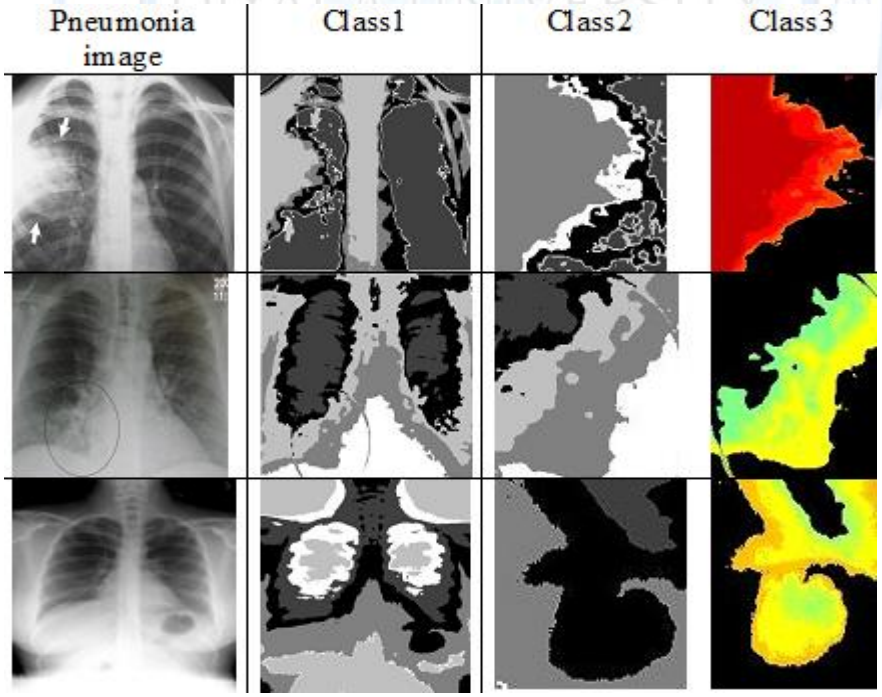
Figure(1) shows the fuzzy c-mean for the Cancer lung images.



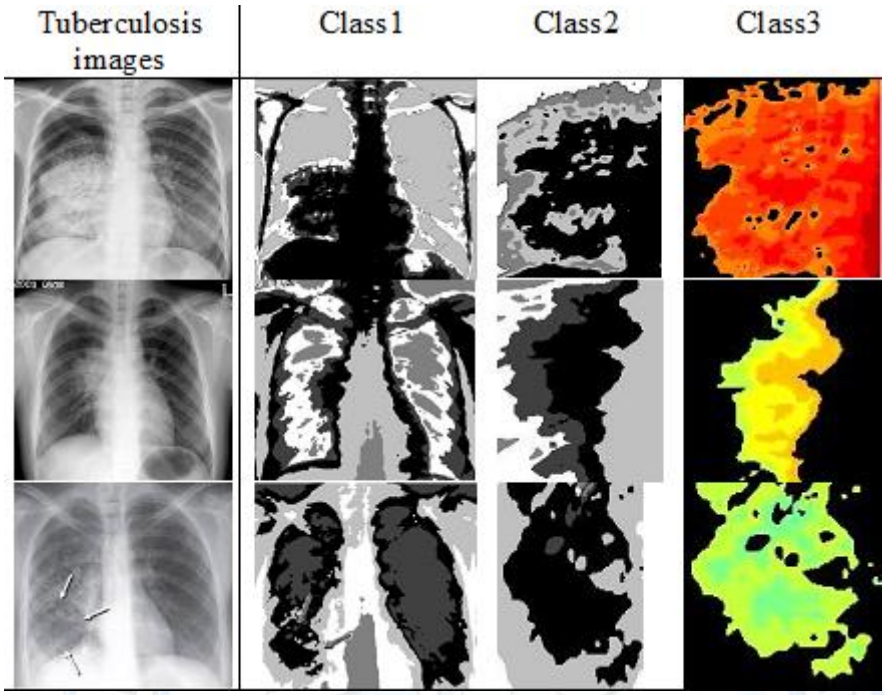
Figure(2) shows the fuzzy c-mean for the Echinococcosis lung images.



Figure(3) shows the fuzzy c-mean for the Lymphatic lung images.



Figure(4) shows the fuzzy c-mean for the Pneumonia lung images.



Figure(5) shows the fuzzy c-mean for the Tuberculosis lung images.

Table (1) shows some features which obtained from Q- matrix for the Cancer case.

Imag NO.	SNE	LNE	NNU	SM
Imag1	0.33	5.91	0.95	0.59
Imag2	0.49	6.01	1.71	0.51
Imag3	0.53	3.11	0.95	0.89
Average	0.45	5.01	1.2033	0.6633

Table (2) shows some features which obtained from Q- matrix for the Echinococcosis case.

Imag NO.	SNE	LNE	NNU	SM
Imag1	0.72	8.01	1.30	1.20
Imag2	0.79	7.92	2.10	1.19
Imag3	0.69	8.90	1.42	1.13
Average	0.7333	8.2767	1.6067	1.1733

Table (3) shows some features which obtained from Q- matrix for the Lymphatic case.

Imag NO.	SNE	LNE	NNU	SM
Imag1	0.12	11.1	1.9	2.15
Imag2	0.23	7.5	2.5	2.01
Imag3	0.19	9.1	2.9	2.13
Average	0.18	9.2333	2.4333	2.0967

Table(4) shows some features which obtained from Q- matrix for the Pneumonia case.

Imag NO.	SNE	LNE	NNU	SM
Imag1	0.81	3.53	2.97	3.00
Imag2	0.85	2.91	3.18	3.31
Imag3	0.79	3.34	3.22	2.87
Average	0.8167	3.26	3.1233	3.06

Table (5) shows some features which obtained from Q- matrix for the Tuberculosis case

Imag NO.	SNE	LNE	NNU	SM
Imag1	0.42	5.53	3.63	2.20
Imag2	0.51	3.91	4.52	4.33
Imag3	0.38	4.34	4.15	3.98
Average	0.4367	4.5933	4.1	3.5033

Table (6) shows features which obtained from Q- matrix for the all cases

case.	SNE	LNE	NNU	SM
The Cancer	0.45	5.01	1.2033	0.6633
Echinococcosis	0.7333	8.2767	1.6067	1.1733
Lymphatic	0.18	9.2333	2.4333	2.0967
Pneumonia	0.8167	3.26	3.1233	3.06
Tuberculosis	0.4367	4.5933	4.1	3.5033

Table (7) shows Gabor Texture features for lung.

Imag NO.	Cancer	Echinococcosis	Lymphatic	Pneumonia	Tuberculosis
Imag1	0.991	1.12	1.02	1.23	1.102
Imag2	1.023	1.20	1.04	1.35	1.24
Imag3	1.10	1.19	1.07	1.43	1.33
Average	1.038	1.17	1.04333	1.33667	1.224

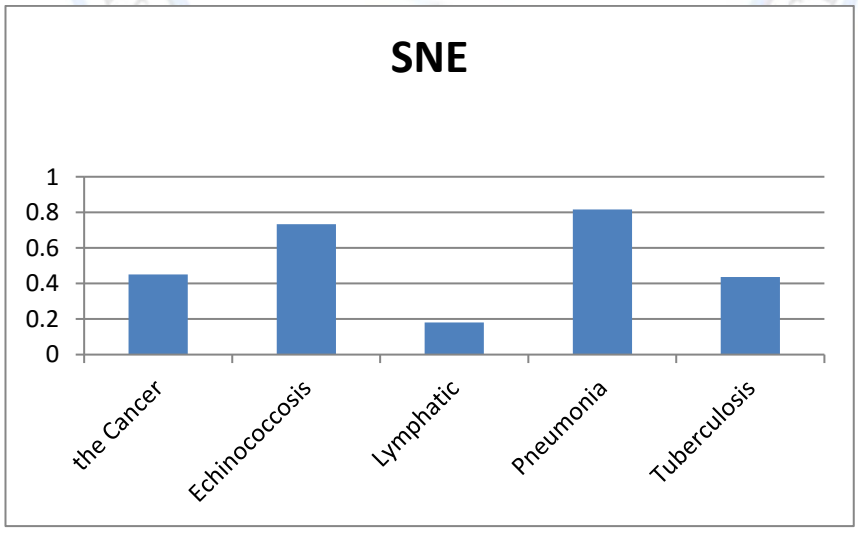


Figure (6)Shows SNE features which obtained from Q- matrix.

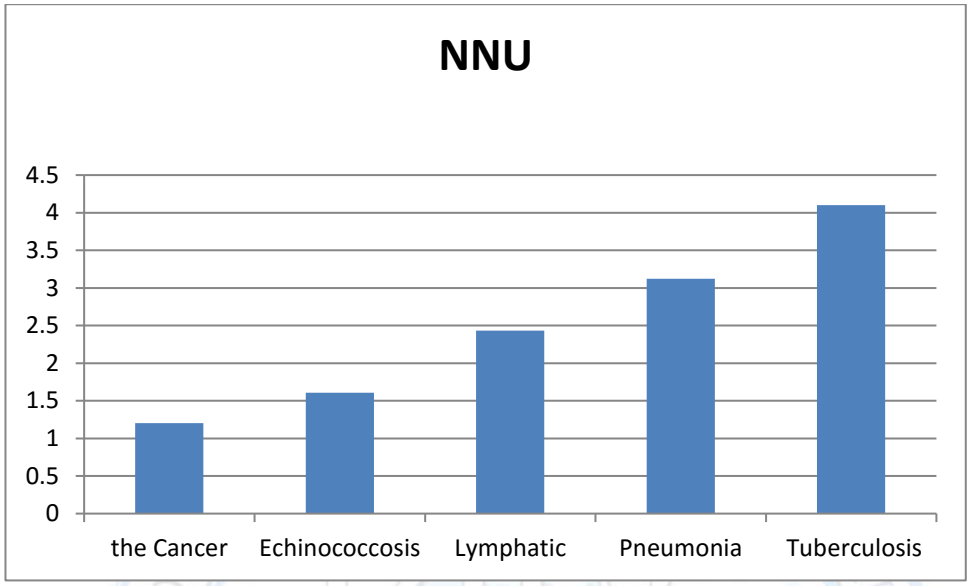


Figure (7)Shows NNU features which obtained from Q- matrix.

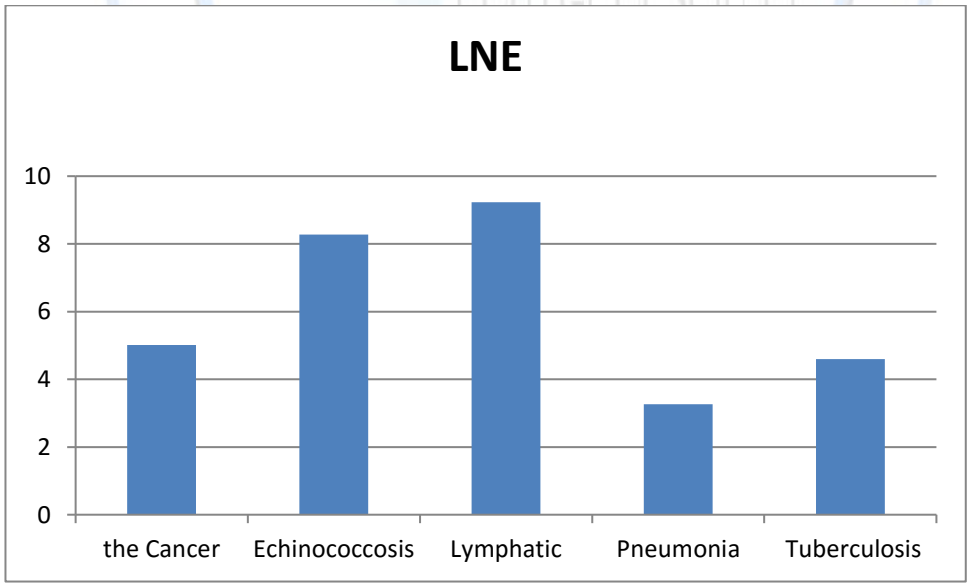


Figure (8) Shows LNE features which obtained from Q- matrix.

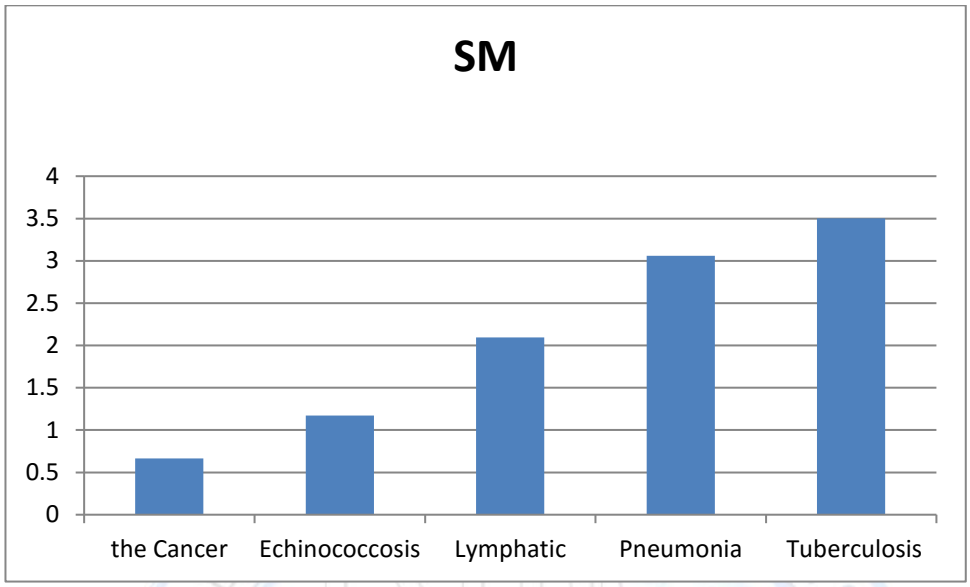


Figure (9) Shows SM features which obtained from Q- matrix.

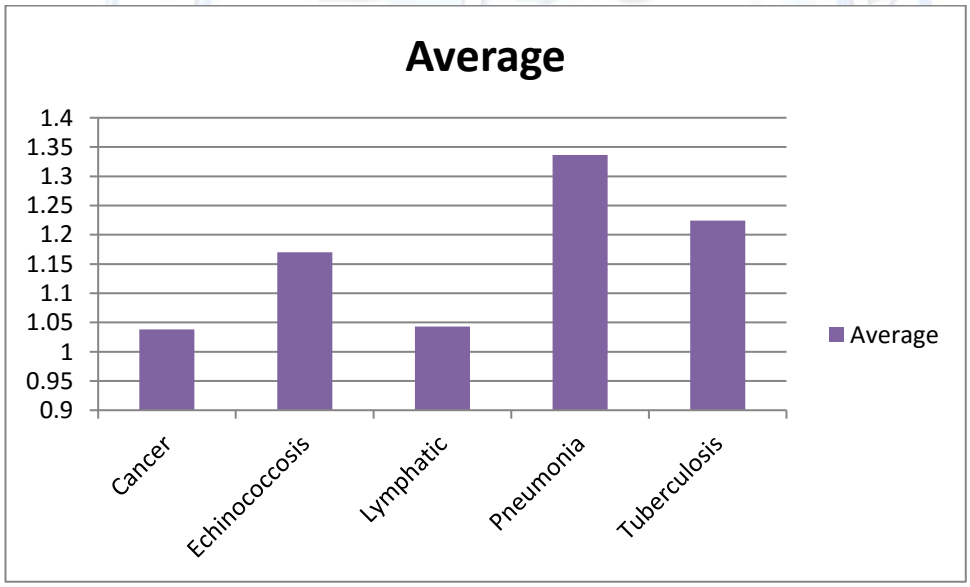


Figure (10) Shows average value of Gabor feature for the lung tumor.

Conclusions

Figure(6) shows the small number emphasis feature which obtained from the Q-matrix with the FCM, the Lymphatic case has the smallest value of (SNE), while the Pneumonia has the largest value i.e the lymphatic texture is more smooth in the contain and tend to be more flat while Pneumonia is rough in its texture , so from this conclusion one can recognize the tumor types, this helps to classify the lunge dieses .

Figure(7) shows the feature for number non uniformity, the tuberculosis has the largest value for non uniformity while the cancer has the lowest value, this is because the cancer texture is the must uniform among the other disease. This means that the cancer cell is uniform in its pattern in which the tuberculosis texture is non uniform in its pattern , this make the diagnose of lung cancer from the other disease easy.

Figure(8) shows the largest number emphasis, the lymphatic case has the large number emphasis while the pneumonia has the smallest value for this case. This means that the lymphatic indicate much emphasis in its texture compare with the other cases, while Pneumonia is the lowest emphasis in its texture.

Figure(9) gives indication about the secondary moment feature, the cancer has smallest secondary moment, while the tuberculosis has the largest value. this means that the cancer is non homogenous in its texture. since, the secondary moment gives indication about the purity of the texture and the cancer make a defect in the lung texture, while the tuberculosis is more homogenous and pure than the other diseases. Figure(10) shows Gabor texture of the average value for the three images for each dieses, since Gabor filters are used to extracted information from images. Pneumonia shows the highest value among the other tumor, this means that the Pneumonia texture shows high sensitivity than the other dieses, (i-e) the Pneumonia make no damage in lung texture. while, the cancer shows low sensitivity because the cancer make a defect in the lung texture, it damage the tissue of the lung so, it appear less sensitivity. this give indication that the cancer is non homogenous in it texture.

Reference

1. The Institute of Islamic Medicine for Education and Research Panama City, Department of Surveying Engineering University of Maine Orono, ME, " Removal of Random Noise from Conventional Digital X-Ray Images", 1992, Khali, I.J. and professor Maureen C. USA Commission VII.
2. IJISET, "Content Based Medical Image Retrieval Using Fuzzy C-means Clustering With RF",2015, Jasmine s., and Nazreen Bee.M., Issue 1, January, Vol.2.
3. IJEIT, "Lung Cancer Classification Using Image Processing".2013, Kuchanur,M,B, No. 3, Vol.2.
4. IJET, "Lung Nodule Detection Using Fuzzy Clustering and Support Vector Machines",2013, S.Sivakumar and C.Chandrasekar, No.1, Vol.5
5. IOSR, "Lung Cancer Detection and Classification by using Mechine Learning and Multinomial Bayesia", 2014, Mr.Sandeep, A.D. and Mr. Borse R.P. No.1, Vol.9.
6. IJSCE, "Detection of Lung Nodule using Multiscale Wavelets Support Vector Machaine", 2012, Aarth K.P and Ragupath, No.3, Vol.2.
7. IJCST, " A Novel Approach for Computation of Q-Matrix", 2011, V.Sneha Latha and C.Naga Raju, Issue 3, September, Vol.2.
8. Archives of Photogrammetric, Cartography and Remote Sensing, "Use of Gabor Filters for Texture Classification of Airborne Images and Lidar data",2011, Urszula Marmol, Vol. 22, pp. 325-336.