Iranian Journal of Medical Physics
Vol. 10, No. 3, Summer 2013, 95-108
Received: October 24, 2012; Accepted: Aug 24, 2013

Original Article

Extraction and 3D Segmentation of Tumors-Based Unsupervised Clustering Techniques in Medical Images

Javad Hadadnia[1]*, Khosro Rezaee[2]

Abstract

Introduction
The diagnosis and separation of cancerous tumors in medical images require accuracy, experience, and time, and it has always posed itself as a major challenge to the radiologists and physicians.
Materials and Methods
We Received 290 medical images composed of 120 mammographic images, LJPEG format, scanned in gray-scale with 50 microns size, 110 MRI images including of T1-Wighted, T2-Wighted, and Proton Density (PD) images with 1-mm slice thickness, 3% noise and 20% intensity non-uniformity (INU) as well as 60 lung cancer images acquired using the 3D CT scanner, GE Medical System LightSpeed QX/i helical, yielding 16-bit slices taken from various medical databases. By applying the Discrete Wavelet Transform (DWT) on the input images and constructing the approximate coefficients of scaling components, the different parts of image were classified. In next step using k-means algorithm, the appropriate threshold was selected and finally the suspicious cancerous mass was separated by implementation image processing techniques.
Results
By implementing the proposed algorithm, acceptable levels of accuracy 92.06%, sensitivity 89.42%, and specificity 93.54% were resulted for separating the target area from the rest of image. The Kappa coefficient was approximately 0.82 which illustrate suitable reliability for system performance. The correlation coefficient of physician's early detection with our system was highly significant ($p<0.05$).
Conclusion
The precise positioning of the cancerous tumor enables the radiologists to determine the progress level of the disease. The low Positive Predictive Value (PPV) and high Negative Predictive Value (NPV) of the system is a warranty of the system and both clinical specialist and patients can trust the software and output.

Keywords: Discrete Wavelet Transform; K-Means Clustering; Image Processing; Lung Cancer; Mammograms; MR Images.

1- *Center for New Research of Medical Technologies Sabzevar University of Medical Sciences, Sabzevar, Iran.*
*Corresponding Author: Tel: +989151710649; Fax: +985714002174; Email: jhaddadnia@yahoo.com.

2- *Biomedical Engineering Department, Hakim Sabzevari University, Sabzevar, Iran.*

1. Introduction

Medical images are the perfect tools for viewing tissue problems. However, these images involve problems such as complex shape structures, weakness details, non-homogeneous in brightness, and poor contrast. Breast cancer, lung cancer, and brain tumors are most common causes of deaths among humans in the world. Breast cancer is currently the most common malignant cancer among women [1]. One in every 7 women loses her life due to this disease [2]. Though limited activities have been performed in terms of cancer registration, but breast cancer remains the most common cancer among Iranian women. This cancer is usually diagnosed through surgical biopsy that has the highest accuracy among the existing methods. However, the invasive procedure is both time-consuming and costly. Photographic techniques are proper instruments with high accuracy in detection of this disease. Mammography is one of these techniques that assists radiologists and doctors in the observation and diagnosis of breast cancer. In mammographic photos, due to the complexities of the breast tissues and the resemblance between cancerous tissues and natural ones, even a radiologist or a specialist might offer different explanations as to the cancerous tumors and their precise location [3]. Moreover, in mammography, 30% of breast cancer case is not correctly diagnosed because the exact location of the cancerous tumor is not properly detected [4].

Lung cancer is the uncontrolled growth of abnormal cells in the lung. This cancer is known as one of the most difficult cancers to cure and number of deaths that it causes is generally increasing. A detection of lung cancer in its early stage can be helpful for medical treatment to limit the danger. However, recent studies also indicated that among 20-60% of abnormalities by screening CT scans most of them are not lung cancer [5-6]. Therefore, development of a reliable, noninvasive, and cost-effective confirmatory test would reduce the over-diagnosis and facilitate the implementation of screening CT scan procedure in the near future [6].

This makes lung cancer the second most prevalent type of cancer to be diagnosed in both men and women, and the leading cancer-related cause of death in both sexes. Lung cancers will account for approximately 13% of new cancer diagnoses in 2006 [7]. It is the leading cause of cancer deaths in North America and results in more deaths than breast, prostate, and colon cancer combined [8].

According to a statistical report published by the Central Brain Tumor Registry of the United States (CBTRUS), the total number of new cases of primary brain tumors for all 50 states and the District of Columbia in United States in 2012 is estimated to be 66,290 with 24,300 being malignant and 41,980 being nonmalignant [9]. The brain tumor is usually diagnosed through surgical biopsy that has the highest accuracy among the existing methods, with the difference that in this way, the invasive procedure is both time-consuming and costly [10]. The cells of the brain arteries are intertwined, so the laboratory tests that are used for chemical analysis of brain tissue usually have not high accuracy [11]. One such modality that has had a great deal of attention from those researching image segmentation methods is MRI [12]. MRI and other medical images contain complicated anatomical structures that require precise and most accurate segmentation for clinical diagnosis [13]. However, MR segmentation methods have been successful on normal tissues [14], but have failed in some cases. Therefore, the necessity of an intelligent system with high precision in the detection of tumor location is felt in medical images.

There is an extensive literature on the development and evaluation of computer aided diagnosis (CAD) systems in mammography, lung cancer, and tumor detection as the one reviewed in [15]. Khuzi et al. using the region of image (ROI) method, contrast analysis, the homogeneity of pixels, and the effect of brightness, extracted texture features [16]. Based on the image processing technique and

employment of discrete wavelet transform (DWT), the cancerous tumor from mammographic images was isolated by Garge et al. [17]. In many cases, global thresholding has been used to segment the breast region from the background as proposed in [18]. For evaluating the presence or absence of tumoral masses, many researches have used combination models [19]. The effectiveness of the algorithm increasing the sensitivity of breast cancer diagnosis was further investigated by Patel et al. [20].

Automatic lung segmentation for accurate quantitation of CT images is offered by Hu et al. as a powerful method in segmentation of lung masses in CT images [21]. Sometimes, by applying unique techniques such as histogram equalization and thresholding, perfect image brightness and segmented background and foreground image in lung imaging and CAD system can be provided [22]. The histogram equalization modifies the intensity levels and make them almost equal that makes the number of pixels per each intensity level the same. Thresholding is primarily concerned with selecting an appropriate threshold according to image histogram. Computational intelligence has been used to segment an image such as a CT-scan display into its constituent parts and allow automated estimations of tissue volumes [23]. We do not have any efficient algorithm in lung imaging, because the masses are very small and thus the segmentation process has different problems.

Several techniques have been proposed for the isolation of MR images that primarily deal with their evaluation and assessment and soft computing [24]. Jayadevappa and et al. formed a hybrid model based on watershed separation technique and vector gradient which showed brain tumors in MR images and CT [25]. In their method, the images were first de-noised and the initial separation of the tumor was performed using the watershed. Narayanan and Karunakar using the geometric properties of tumors began to reconstruct three-dimensional (3D) brain lesions [26]. However, in most of these methods, such as optimization and clustering, time has not been included as a factor and identification of tumor occurs slowly. In other methods including 3D reconstruction, curvelet, conditional random field (CRF), or edge detection, there is the problem of imprecision in extraction of cancerous tumor.

2. Materials and Methods

The proposed algorithm has been implemented on a series of medical images. Among these images, 120 mammogram images taken from Digital Database for Screening Mammography (DDSM) (80 healthy and 40 with tumors), 110 MR images of brain tumors (70 healthy and 40 with tumors), and 60 lung cancer images (36 healthy and 24 with suspicious masses) were used [38].

The main parts of algorithms coded in the MATLAB Version 7.13 and were executed successfully on the low cost PCs to denoising images from noises and extract detailed coefficients and approximate coefficients respectively.

Based on various reasons including sensitivity analysis, testing and training data were divided into two categories that 70% of them were training data and 30% of them were used as testing data. Thus, in 290 selected medical images including 186 images with healthy tissue and 104 images with pathologic tissue, 203 images (including 130 images with healthy tissue and 72 images with pathologic tissue) were used for training data and the rest of them were implemented for testing data. The steps of proposed algorithm are shown as a schematic chart (Figure 1).

2.1. Pre-processing step

Medical images have so prominent information about masses especially when the mass is not centralized in canter tissue. Denoising is used as a pre-processing step in many image processing and analysis tasks such as registration or segmentation to reduce the random noise arising from the acquisition process [27].

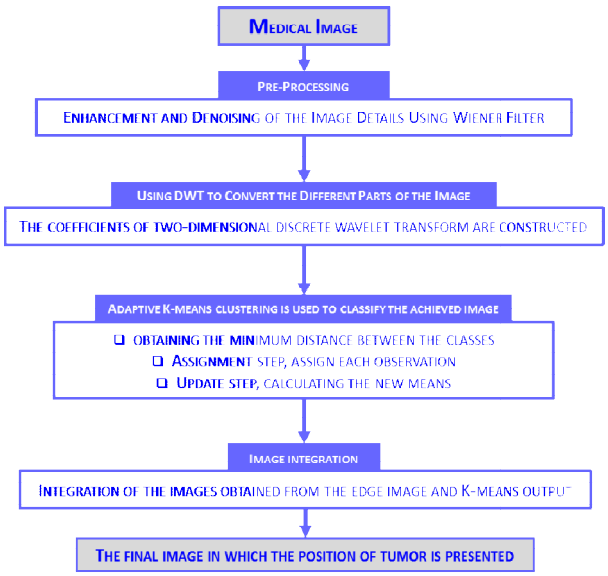

Figure 1. The steps of proposed algorithm are shown as schematic chart.

Images from various modalities need to be denoised as a pre-processing step for many planning, navigation, detection, data-fusion, and visualization tasks in medical applications and CT slices are often corrupted by noise during acquisition or transmission [28]. Radiologists need an image without any noise to recognize the place and position of the masses. Median filter is a non-linear filter that is used to eliminate impulse noises. Salt-and-pepper noise appears when there is an inappropriate function of glasses sensor cells, memory cells errors, and synchronized errors while digitizing or data transition. The main work of this filter is using neighbor pixels of the aimed pixel and while a mask pads all pixels. While a pixel is processed, all neighbor pixels are organized based on intensity of pixels ascend, then intensity of median element of organized numbers is selected as a new pixel. Wiener filter is used to remove the blurring and noise in images which are faced with high level of obscurity [27,28].

2.2. Discrete Wavelet Transform

Wavelet transform is defined both for discrete Fourier transform and continuous transform. Having numerous applications, these transforms, occur in one-dimensional and two-dimensional spaces. In the two-dimensional discrete wavelet transform, an image is treated as a matrix, each row or column is considered as a signal that its amplitude is the brightness of pixels in that certain row or column. With the application of one-dimensional discrete wavelet to each row, and fixation of columns, parallel signals are achieved. Then, assuming the constancy of rows, this procedure is applied to all columns again and as a result, the rows with a rate of 2 are sampled [29]. With the break of two-dimensional discrete wavelet transform into four components of approximate coefficients at j level, it can be inferred that the approximation is transformed into j+1 level and three other levels, i.e., horizontal, vertical, and diagonal levels. To obtain the two-dimensional discrete wavelet transform, we first need to build the wavelet series expansion function and the scaling function [30].

The two-dimensional discrete wavelet transform in four directions for MR image is shown in Figure 2. In this figure, on the top row there are images including a mammogram, MR brain tumor, and a lung cancer CT image in which locating the cancerous tumor is difficult. In the same figure on the bottom row, the image calculated using only S1 component coefficients has been displayed for 3 types of medical image and implementation of DWT (Figure 3).

K-means is one of the simplest unsupervised learning algorithms and the procedure follows a simple and easy way to classify a given dataset through a certain number of clusters (assume K clusters) fixed as a priori [33]. We have made an attempt to display the performance of existing K-means approach by varying various values of certain parameters. The K-means algorithm is an iterative algorithm that is used to partition an image into *K* clusters. In statistical computation and machine learning, *k*-means clustering is a technique of cluster analysis which aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean. Among clustering methods used in segmentation of various parts of the image, the mean-based clustering or K-means was utilized.

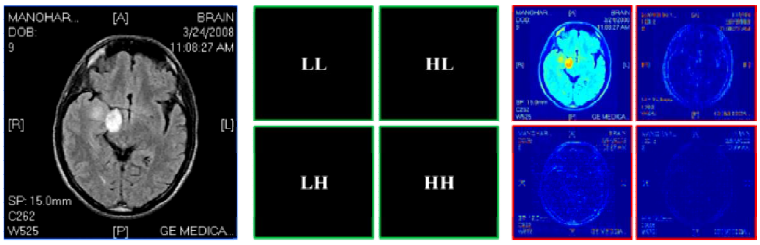

Figure 2. The two-dimensional discrete wavelet transform in four directions for MR image.

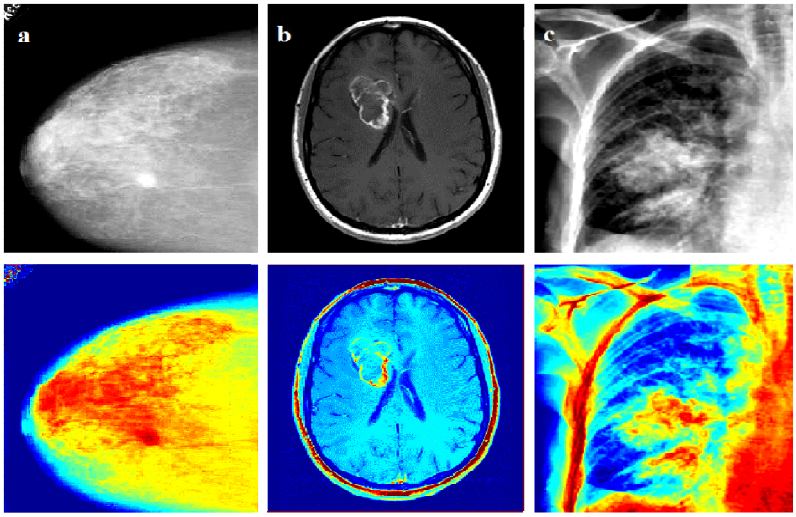

Figure 3. On the top row there are images of (a) a mammogram, (b) a MR brain tumor, and (c) a lung cancer CT image. The DWT algorithm is implemented on all of images and is shown results on the bottom row

The objective function was to obtain the minimum distance between the classes, or basically between the image pixels. In fact the aim of the K-means clustering is to segment the area of the input pattern vectors into K clusters. If input patterns is included of a set of N vectors $\{\vec{x}_1, \vec{x}_2, ..., \vec{x}_N\}$ and the Euclidian distance is used as a measure of similarity, then we can formulate K-means clustering as that of finding K cluster centers, $\vec{c}_1, \vec{c}_2, ..., \vec{c}_K$, that minimize the total square-error E [34]:

$$E(\vec{c}_1, \vec{c}_2, ..., \vec{c}_K) = \sum_{k=1}^{K} \{(1/N) \sum_{i=1}^{N} m_{ki} \|\vec{x}_i - \vec{c}_k\|^2\} \quad (1)$$

In which m_{ki}=1 if $\vec{x}_i$ belongs to cluster k, and m_{ki}=0 otherwise. The notation $\| \ \|$ denotes norm of term. When the training patterns are generated from probability density $p(\vec{x}_i)$ defined on an input space *S*, the cost function of the K-means algorithm is transformed into:

$$E(\vec{c}_1, \vec{c}_2, ..., \vec{c}_K) = \sum_{k=1}^{K} \int_S m(\vec{x}) \|\vec{x}_i - \vec{c}_k\|^2 p(\vec{x}) d\vec{x} \quad (2)$$

Where $m(\vec{x}) = 1$ if $\vec{x}$ belongs to cluster k, and $m(\vec{x}) = 0$ otherwise. For expectation maximization and standard *k*-means algorithms, the Forgy method of initialization is preferable. The algorithm proceeds by alternating between two steps [35]:

Assignment step: each observation is assigned to the cluster whose mean is closest to it.

Update step: new means are calculated to be the centroids of the observations in the new clusters.

Classification of image pixels is based on the brightness intensity. The frequency of brightness intensity of images (histogram) has been used to select the appropriate number of secretions. Based on this classification, pixels can divide into a maximum of 255 clusters. Here, based on achieved results from Ray and Salvador, the number of proposed clusters has been 4 to 7 [36,37]. In image segmentation, the choice of K is not critical because the adaptation of the intensity functions allows the same region type to have different intensities in different parts of the image. All of developed criteria have a common goal to find the clustering which results in compact

clusters which are well separated. However, we used Ray and Salvador techniques to determine the optimum number of clusters in proposed k-means clustering in color images segmentation [36,37]. We can measure the distance of intra-cluster which is simply the distance between a point and its cluster center (*DBPC*) and inter-cluster or the distance between clusters (*DBC*) according (3) and (4), respectively:

$$DBPC = \frac{1}{N}\sum_{i=1}^{K}\sum_{x\in C_i}\|x - z_i\|^2 \quad (3)$$

$$DBC = \min\left(\left\|z_i - z_j\right\|^2\right) \quad (4)$$
$$i = 1,2,...,K-1,\ j = i+1,...,K$$

Where in (3), N is the number of pixels in the image, K is the number of clusters, and z_i is the cluster center of cluster C_i and in (4) z_j is other cluster that is compared with z_i. We obviously want to minimize (3) and take the minimum of (4) value. Finally, we define the ratio as *Validity=DBPC/DBC* in which the goal is to minimize the intra-cluster distance and consequently minimize the Validity measure and also maximize the inter-cluster distance measure (because this term is denominator). Therefore, the clustering which gives a minimum value for the validity measure tells us what the ideal value of K is in the k-means procedure. In many images K = 5 is the best choice due to results reported from color space (red, green, and blue). Thus, classification of image by K-means algorithm can help to select proper threshold and not the brightness intensity acquired from clustering level of tumor detection in input image of frames.

Implementation of DWT due to selecting appropriate K number for tumor detection in various medical images and K-means clustering is an efficient technique to ridge the boundary of different parts in medical images. In Figure 4, the steps of implementation of K-means clustering is shown for K equal 3 which intensity of MR image of brain tumor was parted to 6 regions. The position of tumor in Figure 2 was detected. However, to increase the accuracy and sensitivity of algorithm, using K-means clustering method, pixels in image were reconstructed to separate from others. The K-means algorithm has been shown in Figure 5, which comparing with Figure 2, the positions of tumors is detected with high accuracy.

2.4. Thresholding, Margin Removal and Image Combination

The brightness intensity of a gray-scale image can only vary between [0-255]. This is a very useful feature. Partial separation of an image through selecting appropriate threshold is based on dividing the image into background and foreground classes. One of the effective methods in image processing, which is often used to separate the background from image, is thresholding method which is based on selecting an appropriate threshold from image histogram. The more careful is this selection, the more accurate will be the accuracy of separating image background from its main framework. In image processing, mainly symmetrical thresholding method is used.

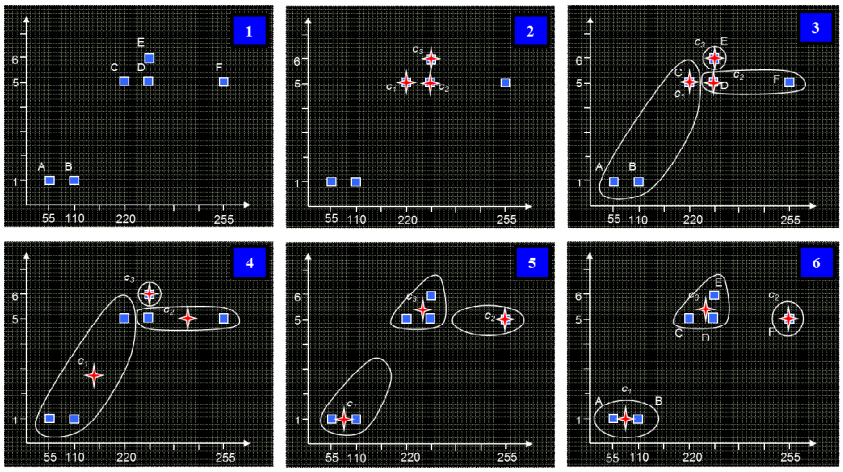

Figure 4. Applying the K-means clustering for MR image containing tumor.

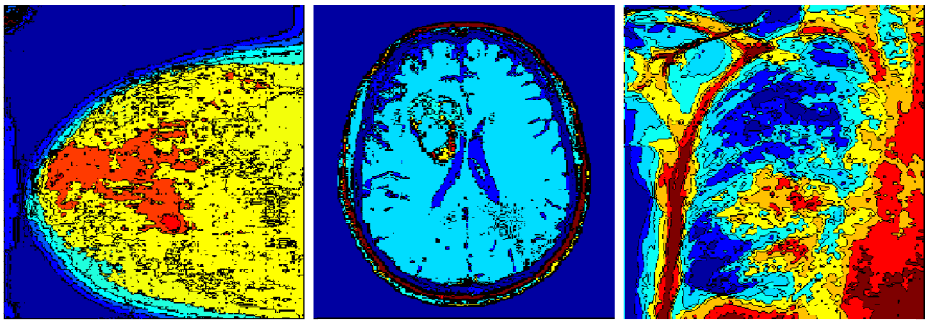

Figure 5. Implementation of K-means algorithm to determine the exact position of tumors in various medical images.

Image histogram will be split in half at t_0. Then, as to brightness intensity of $t < t_0$, $t > t_0$ average value will be calculated and we will be able to witness the greater average (either on left or right side). Thus, by changing the value of t_0 within certain brightness intensity, the means become equal, or at least, the difference between these two means becomes smaller than a minimal number. However, if partial separation of image is conducted according to brightness intensity, the threshold value or boundary is considered as the basic brightness intensity in division. That is, brightness intensities greater than threshold value will be equal to 1 and brightness intensities less than that will be equal to 0. Thus, we will have a binary image consisting of zero and one elements. This method is known as Otsu technique. Otsu shows that minimizing the intra-class variance is the same as maximizing inter-class variance:

$$\sigma_b^2(t) = W_1(t)W_2(t)[\mu_1(t) - \mu_2(t)]^2 \qquad (5)$$

Which is expressed in terms of class probabilities W_i and class means μ_i. In Table 1, the structural and procedure conditions of proposed thersolding are shown.

The histogram of three types of medical images has been displayed (Figure 6). To delete the extra elements that are associated with the threshold process, we use an area with a [0-1] range for edge detection. That is, if after the processing of an edge pixel, the number of its neighboring pixels is less than a base value, the brightness intensity of the targeted pixel will be transformed into zero; otherwise, it will maintain its previous value. Using the separation of the pixels surrounding the targeted tumor which is obtained during the edge detection stage of the binary image, the matrix of pixels around the edge is formed. In fact, the pixel selected as the periphery pixel or environmental pixel needs to be both non-zero and connected only to one pixel with zero value.

Table 1. The proposed thersolding

Begin

Compute histogram and probabilities of each intensity level

Initialize $W_i \leftarrow 0$, $\mu_i \leftarrow 0$.

Step through all possible thresholds t=1… maximum intensity.

1. Update W_i and μ_i
2. Compute $\sigma_b^2(t)$

Desired threshold corresponds to the maximum $\sigma_b^2(t)$

Compute two maximum $\sigma_b^2(t)$ (greater max) and $\sigma_b^2(t)$ (the greater or equal maximum).

Desired *Threshold*$=(Thr_1 + Thr_2)/2$

End

Considering an appropriate threshold in [0-1] range and with combination of the image resulted from the K-means step and matrix of pixels around the edge, the precise location of cancerous tumor will be detected as in Figure 7 for three types of medical images.

3. Results

The proposed algorithm has been implemented on a series of medical images. The images taken from DDSM database [38] were in LJPEG format and were scanned in gray-scale with 50 microns size. The resolution of these

images was 200 micrometers. MR images obtained from MedPix^{Tm} and Harvard Medical School databases [39,40] including of MRI T1-wighted, T2-wighted and PD (Proton Density) images which 40 images contain tumor or edema and other images only represent normal healthy tissue (white and gray tissue). Images have characteristics as the Modality including T1 and T2, Slice thickness equal 1mm, Noise equal 3% and INU equal 20%.

Experts confirmed the existence of tumor or edema in 40 images and in the rest of images, the existence of cancer was substantiated. Images were divided into two types: a) Three sets corresponding to healthy tissues and fluid (WM, GM, and CSF) and b) Two sets corresponding to pathological tissues (tumor and edema).

The lung cancer images were scanned by a Siemens machine, with processed image size as 512×512 and the pixel resolution of 0.59×0.59 mm2. Histologic diagnoses were made for all of the patients by radiologists, and the histologic types were bronchioloalveolar carcinoma, adenocarcinoma, and idiopathic pulmonary fibrosis. Images were acquired using the 3-D CT scanner, GE Medical System LightSpeed QX/i helical, and yielded 16-bit slices of 512×512 pixel arrays. The image values were recorded as Hounsfield Unit (HU) values, representing the densities of different human tissues. All images were converted to the 256×256 dimension. All three categories of images were resized to 256×256 pixels so that the algorithm would generate its output in the specified time period. Except for 7 images, the algorithm was successful in recognition of the mass and desired section in 267 medical images. Of 104 disease images, it did not diagnose properly the disease in 11 images. Three factors, i.e., accuracy (AC), specificity (SP), and sensitivity (SE), which were introduced for assessing the accuracy of the system in performance detection, were calculated according to the (6) to (8).

$$Senestivity = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (6)$$

$$Specificity = \frac{N_{TN}}{N_{TN} + N_{FP}} \quad (7)$$

$$Accuracy = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{FN} + N_{TN} + N_{FP}} \quad (8)$$

In these equations, N_{TP} is the number of images containing the tumor tissues that has been detected by algorithm and N_{FN} is the number of images containing the tumor tissues which algorithm has not been successful in their detection. Also N_{TN} is the number of images that does not contain the tumor tissues and algorithm has not identified them and finally N_{FP} is the number of frames that does not contain the tumor tissues but the algorithm has misidentified them. After calculating of these parameters, 89.42% sensitivity, 93.54% specificity, and 92.06% accuracy were achieved. Kappa coefficient shows the reliability of the system performance which is introduced in (9).

$$Kappa = \frac{2(N_{TP}N_{TN} + N_{FN}N_{FP})}{(N_{TP} + N_{FN})(N_{TN} + N_{FN}) + (N_{TN} + N_{FP})(N_{TP} + N_{FP})} \quad (9)$$

The results indicate Kappa equal to 0.8241 which is suitable for system performance. The coefficients of three factors have been calculated for each image and shown in Table 2.

The F-measure factor is a much more appropriate measure than accuracy for analyzing segmentation outputs. For the unsupervised methods, parameters such as initial values are optimized by exhaustively determining the values which obtained in the best possible F-measure for the training set. However, we kept these parameters fixed for the testing stage. The unbiased F-measure, on the other hand, is given by (10):

$$\text{FMeasure} = \frac{2N_{TP}}{2N_{TP} + N_{FP} + N_{FN}} \quad (10)$$

We used overlap procedure to evaluate the performance of the system and the output of it that based on the output of algorithm compared with the ground truth images which are identified by physicians.

Table 2. Implementation of the system and the results of the evaluation.

Databases		No. Images	Tumor		Tissue		F-Measure	Accuracy	A_z	Kappa
			N_{TP}	N_{FN}	N_{TN}	N_{FP}				
Mammograms	Normal	80	-	-	76	4	0.9268 (±0.02)	0.9500 (±0.01)	0.8447 (±0.06)	0.8938
	Abnormal	40	38	2	-	-				
Brain	Normal	70	-	-	65	5	0.9024 (±0.02)	0.9357 (±0.02)	0.7921 (±0.17)	0.8551
	Abnormal	40	37	3	-	-				
Lung Cancer	Normal	36	-	-	33	3	0.8000 (±0.07)	0.8500 (±0.06)	0.6904 (±0.06)	0.7234
	Abnormal	24	18	6	-	-				
Total	A and N	290	93	11	174	12	0.9029 (±0.037)	0.9206 (±0.03)	0.7758 (±0.097)	0.8241

The pixels of output in each image were obtained by counting and the similarity of cancerous was calculated based on (11):

$$OvL = \frac{2|A \cap B|}{(|A| + |B|)} \qquad (11)$$

In this equation the OvL is similarity factor, A is the number of cancerous pixels of ground truth, B is the number of our output and notation $|\ |$ is size of the target sets. We can decide that the obtained cancerous mass is classified as true positive class without any sign of mass and true negative case with sign of mass by choosing an appropriate threshold in similarity factor (SF) equal 0.8 in output of this equation. If the image did not have any sign of mass or illness and SF<0.8, then output was classified as false positive and if it had sign of illness and SF>0.8, then the output was classified in false negative case. Software is also capable of displaying 3D images. 3D display of medical images makes it possible that the physician find suspected cancer comfort zone and also, he/she can identify the center of mass and its related sectors. As a practical example of displaying 3D images Figure 8 displays the complete information about the masses. The Receiver Operating Characteristics (ROC) diagram is a two-dimensional curve in which the x-axis displays the variation rate of the positive change and the y-axis displays the sensitivity or true positive rate of change. Moreover, we calculated A_z factor which is the area under the ROC curve. ROC curves in the left side of Figure 9 show performance of three-detect cancerous tumors.

In the right side of Figure 9, the average performance of final system is shown. The area under the ROC curve shows a good performance in the parameters that is determined between the two groups (healthy and with tumor).

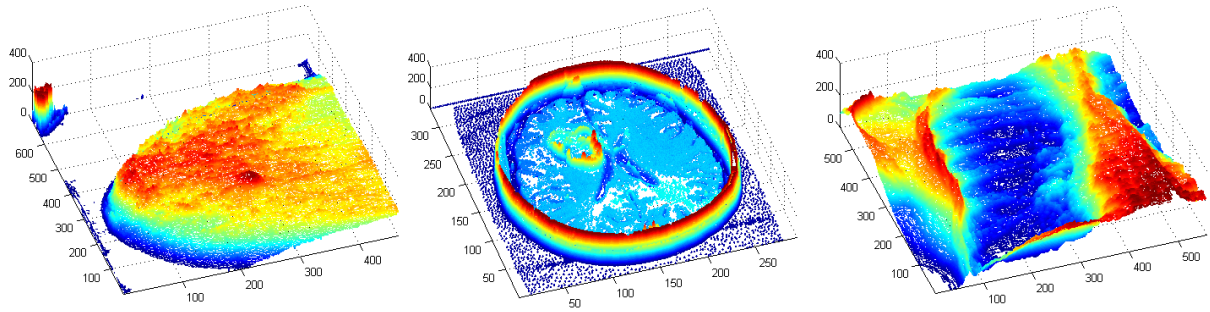

Figure 8. Displaying three-dimensional medical images to identify the relevant parts of the tumor.

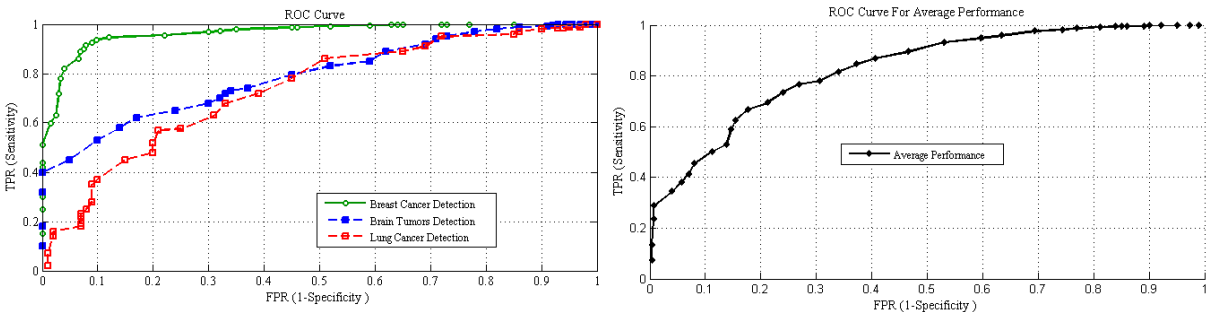

Figure 9. ROC curves in the left side of figure show performance of algorithms and in the right side of figure, the average performance of the final system is shown. The number of random images is 50.

4. Discussion

The proposed method has an adequate accuracy, but in the cost of poor Kappa coefficient. In Table 3, a comparison with some valid methods is represented. It should be noted that the proposed method is a new technique and a comparison has been made between the performance of this system and other methods used for detection of the cancerous masses. Different methods use various databases. The proposed method in this paper, however, has been applied to a larger database. At first, it should be noted that the use of a single technique such as only discrete wavelet transform or only K-means clustering would decrease the sensitivity and specificity in contrast with other techniques. Thus, the occurrence of error is normal. Using multiple techniques, which simultaneously identify the cancerous masses of the medical images, increases the accuracy, sensitivity, and specificity of this system more than the other techniques. According to the changes in light, intensity level does not affect boundaries and K changes as well as the maximum value of 9 in images (Kmax=9), so by adding to this amount, sensitivity and specificity would not be affected. However, in a research, authors showed that the K-means algorithm is very sensitive to the choice of cluster centers [41]. For example, Su et al. emphasized that the initial center affects the clustering results, but the majority of the ten trials have the same clustering results. In some cases, authors considered estimating of the number of classes as a part of their segmentation algorithms [42,43] and as result choosing the wrong number of classes could be disastrous in image segmentation. However, our adaptive algorithm was quite robust to the choice of K. This is because the characteristic levels of each class adapt to the local characteristics of the image, and thus regions of entirely different intensities can belong to the same class, as long as they are separated in space. We can understand this result from Figure 10 that classification procedure is shown for different Ks.

Our system is compared with two cancer breast detection techniques [19,20], two lung cancer detection systems [22,23], and two brain tumor methods [24,25]. The performance of Maitra method is suitable, but output is relatively high access time. Patel method (Adaptive K-means) has low accuracy, but in lung cancer detection, the performance is better. In brain tumor algorithm (Jayadevappa), accuracy is high, but the time elapsed is very long.

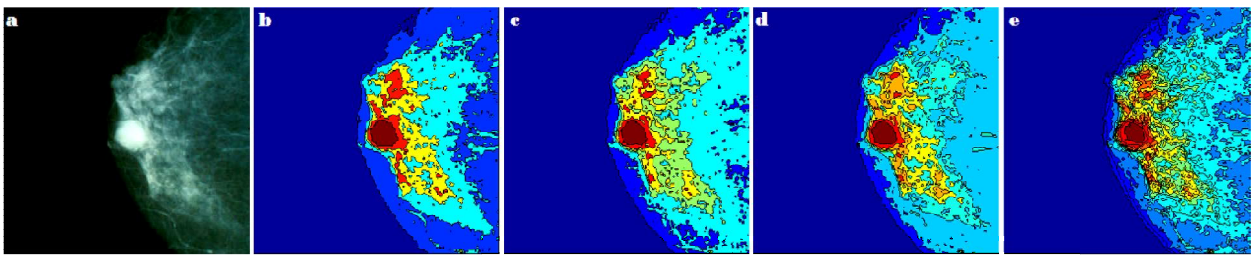

Figure 10: the poor effect of K in clustering (a) original image, (b) K=5 and OVL= 0.9332, (c) K=8 and OVL= 0.9763, (d) K=10 and OVL= 0.9434, and (e) K=18 and OVL= 0.8715.

Table 2. Comparison algorithm with some valid methods.

The Technique	Accuracy Ratio	Kappa	Time(sec)
Maitra [19]	92.32%	~0.8	~3.5
Patel [20]	84.00%	~0.75	~1.5
Sharma [22]	90.00%	~0.8	Unknown
Altarwneh [23]	92.86%	~0.78	Unknown
Logeswari [24]	>90.00%	>0.8	~3.5
Jayadevappa [25]	>96.00%	>0.8	28
Our System	92.06%	~0.83	~2.5

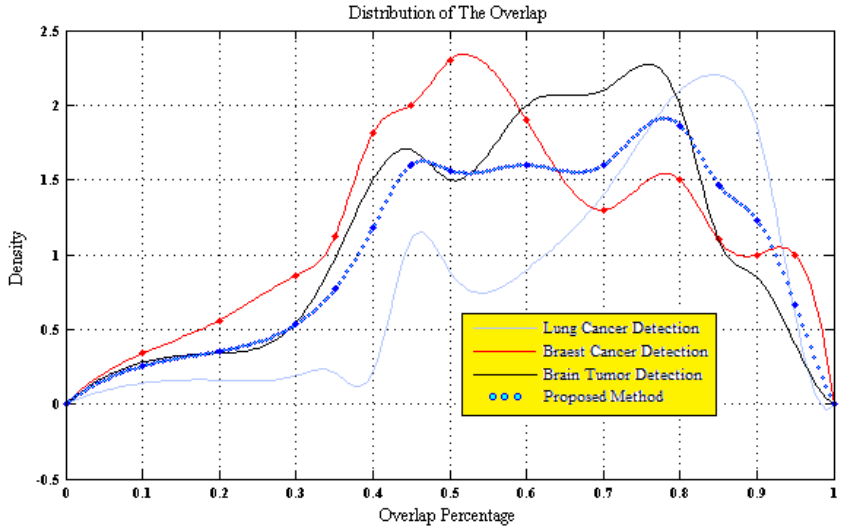

Figure 11. The distribution of the overlapping percentage for 3 detection algorithms and the proposed algorithm.

Figure 11 presents the overlapping percentage of the proposed method with what stated by physicians and radiologists, which is approximately 75%, while three detection algorithms have respectively 50% and 57% and 62% overlapping percentage. Such statistical difference is highly significant ($p<0.01$).

5. Conclusion

Due to the proper performance, high-speed process of the input data, and precise separation of cancerous tumors in medical images, the proposed system in this paper would assist radiologists and specialists. Two hundred and ninety medical images were taken from various databases with the optimum 92.06% accuracy and 89.42% sensitivity separation of breast cancer tumors. Finally, the proposed algorithm was implemented on a couple of sample images. The results can be seen in Figure 7. Compared with the similar methods such as edge detection separation, morphological methods, and separating algorithm, this method is superior in terms of speed and accuracy and does not suffer their defects including the partial deletion of the original image data and reduced quality. The employment of this system in device used for observing breast cancer, such as mammography, MRI, and CT-scan increases the accuracy and minimizes human errors. The low Positive Predictive Value (PPV) and high Negative Predictive Value (NPV) of the system (the reliability coefficients for clinical specialist and the patient, respectively) is a warranty of the system and both clinical specialist and patients can trust its software and output.

Acknowledgment

We are grateful to radiologists A. Rezaeemanesh, M. Bahadorpur, and J. Saeedi for suggestions regarding this work. This study was supported by Hakim Sabzevari University and Sabzevar University of Medical Sciences.

References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010 Dec 15; 127(12):2893-917.
2. Amer. Cancer Facts & Figures 2008, Cancer Society, Available from: http://www.cancer.org.
3. Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists. Findings from a national sample. Arch Intern Med. 1996 Jan 22; 156(2):209-13.
4. Majid AS, de Paredes ES, Doherty RD, Sharma NR, Salvador X. Missed breast carcinoma: pitfalls and pearls. Radiographics. 2003 Jul-Aug; 23(4):881-95.
5. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011 Aug 4; 365(5):395-409.
6. Barba M, Felsani A, Rinaldi M, Giunta S, Malorni W, Paggi MG. Reducing the risk of overdiagnosis in lung cancer: a support from molecular biology. J Cell Physiol. 2011 Sep; 226(9):2213-4.
7. McKee DW, Land WHJ, Zhukov T, Song D, Qian W. An adaptive image segmentation process for the classification of lung biopsy images. In: Reinhardt JM, Pluim JP, editors. Medical imaging 2006: image processing Technical conference; Proceedings of the SPIE; 2006. p. 1628-37.
8. Jemal A, Tiwari RC, Murray T, Ghafoor A, Samuels A, Ward E, et al. Cancer statistics, 2004. CA Cancer J Clin. 2004 Jan-Feb; 54(1):8-29.
9. Siegel R, Naishadham D, Jemal A. Cancer Statistics, 2012. CA Cancer J Clin 62:10-29, 2012.
10. Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening radiology by US radiologists, findings from a national sample. Arch Intern Med, 1996; 156(2):209–13.
11. Qurat-Ul-Ain GL, Kazmi SB, Jaffar MA, Mirza AM, editors. Classification and segmentation of brain tumor using texture analysis. Proceedings of the 9th WSEAS International Conference on Recent Advances in Artificial Intelligence, Knowledge Engineering and Data Bases; 2010.
12. Vannier MW, Speidel CM, and Rickmans DL. Magnetic resonance imaging multispectral tissue classification. News Physiol Sci, August 1988; 3(4):148-54.
13. Hall LO, Bensaid AM, Clarke LP, Velthuizen RP, Silbiger MS, Bezdek J. A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. IEEE Trans Neural Network, 1992; 3(5):672–682.
14. Tsai C, Manjunath BS, Jagadeesan R. Automated segmentation of brain MR images. Pattern Recognition. 1995; 28(12):1825-37.
15. Sampat MP, Markey MK, Bovik AC. Computer-aided detection and diagnosis in mammography. Handbook of Image and Video Processing, 2rd Ed. New York: Academic; 2005. p. 1195–1217.
16. Mohd Khuzi A, Besar R, Wan Zaki W, Ahmad N. Identification of masses in digital mammogram using gray level co-occurrence matrices. Biomed Imaging Interv J. 2009 Jul; 5(3):e17.
17. Garge DM, Bapat VN. A low cost wavelet based mammogram image processing for early detection of breast cancer. Indian Journal of Science and Technology, Sep 2009; 2(9): 63-5.
18. Masek M, Attikiouzel Y. Skin-Air Interface Extraction from Mammograms Using an Automatic Local Thresholding Algorithm. In proceedings of the 15th Biennial International Conference Biosignal; ICB Brono CR; 2000; p. 204-206
19. Maitra I, Nag S, Bandyopadhyay SK. Detection of Abnormal Masses using Divide and Conquer Algorithm in Digital Mammogram. International Journal of Emerging Sciences. 2011; 1(4):767-86.
20. Patel BC, Sinha GR. An Adaptive K-means Clustering Algorithm for Breast Image Segmentation. International Journal of Computer Applications. 2010; 10(4):35-8.
21. Hu S, Hoffman E, Reinhardt J. Automatic lung segmentation for accurate quantitation of volumetric x-ray CT images. IEEE Transactions on Medical Imaging. 2001; 20(6): 490–498.
22. Sharma D, Jindal G. Computer Aided Diagnosis System for Detection of Lung Cancer in CT scans Images. International Journal of Computer and Electrical Engineering. 2011; 3(5):714-8.
23. Altarwneh MS. Lung Cancer Detection Using Image Processing Techniques. Leonardo Electronic Journal of Practices and Technologies. 2012; 11(21): 147-58.
24. Logeswari T, Karnan M. An improved implementation of brain tumor detection using segmentation based on soft computing. Journal of Cancer Research and Experimental Oncology. 2010; 2(1):006-14.

25. Jayadevappa D, Kumar SS, Murty DS. A Hybrid Segmentation Model based on Watershed and Gradient Vector Flow for the Detection of Brain Tumor. International Journal of Signal Processing, Image Processing and Pattern Recognition. September 2009; 2(3):29-42.
26. Narayanan K, Karunakar Y. 3-D Reconstruction of Tumors in MRI Images. International Journal of Research and Reviews in Signal Acquisition and Processing. June 2011; 1(2): 24-29.
27. Manjón JV, Coupé P, Buades A, Louis Collins D, Robles M. New methods for MRI denoising based on sparseness and self-similarity. Med Image Anal. 2012 Jan; 16(1):18-27.
28. Barzigar N, Roozgard A, Verma P, Cheng S. Removing Mixture Noise from Medical Images Using Block Matching Filtering and Low-Rank Matrix Completion. Healthcare Informatics, Imaging and Systems Biology (HISB), 2012 IEEE Second International Conference on San Diego, CA, 27-28 Sept. 2012; p.134.
29. Kim JW, Song J, Lee S, Park IC. Tiled interleaving for multi-level 2-D discrete wavelet transform; International Symposium on Circuits and Systems; May 2007; New Orleans, LA; IEEE; 2007. p. 3984–87.
30. Gonzalez R, Wintz P. Digital image processing. 5th edition, Addison-Wesley Publishing Co., New York., Oct 2008; p. 139-203.
31. Michael J, Christopher B. Neural Networks. Second Edition (Section VII: Intelligent Systems). Boca Raton, FL: Chapman & Hall. The Computer Science and Engineering Handbook; 2004.
32. Kaufman L, Rousseeuw PJ. Finding Groups in Data: an Introduction to Cluster Analysis. John Wiley & Son; 1990.
33. Niknam T, Taherian Fard E, Pourjafarian N, Rousta A. An efficient hybrid algorithm based on modified imperialist competitive algorithm and K-means for data clustering. Engineering Applications of Artificial Intelligence. March 2011; 24(2):306–17.
34. Chinrungrueng C, Sequin CH. Optimal adaptive k-means algorithm with dynamic adjustment of learning rate. IEEE Trans Neural Netw. 1995; 6(1):157-69.
35. MacKay D. An Example Inference Task: Clustering. Information Theory, Inference and Learning Algorithms. Cambridge University Press. 2003; p. 284–292 Chapter 20.
36. Ray S, Turi RH. Determination of Number of Clusters in K-Means Clustering and Application in Colour Image Segmentation. Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques, Narosa Publishing House. 27-29 December, 1999; p.137-143.
37. Salvador S, Chan P. Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms. In Proc. ICTAI, Boca Raton, FL, November 2004.
38. Heath M, Bowyer K, Kopans D. The Digital Database for Screening Mammography. In Proc. International Digital Mammography. 2000. p. 212-218.
39. Online Medical Image Database and Teaching File [Online]. Available: http://rad.usuhs.edu/medpix/ parent.php. Accessed Oct 20, 2013.
40. Yun Z, Huang SC, Shanglian B, Wong DF, editors. Parametric imaging and statistical mapping of brain tumor in Ga-68 EDTA dynamic PET studies. Nuclear Science Symposium Conference Record, 2001 IEEE; 2001 4-10 Nov. 2001.
41. Mu-Chun S, Chien-Hsing C. A modified version of the K-means algorithm with a distance based on cluster symmetry. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2001; 23(6):674-80.
42. Ahmed MM, Mohamad DB. Segmentation of brain MR images for tumor extraction by combining Kmeans clustering and Perona-Malik anisotropic diffusion model. International Journal of Image Processing. 2008; 2(1):27-34.
43. Rezaee K, Haddadnia J. Designing an Algorithm for Cancerous Tissue Segmentation Using Adaptive K-means Cluttering and Discrete Wavelet Transform. J of Biomed Phys and Eng, 2012 Sep; 3(3): 93-104.