JJMIE

Volume 4, Number 6, December 2010
ISSN 1995-6665
Pages 677 - 684

Jordan Journal of Mechanical and Industrial Engineering

www.jjmie.hu.edu.jo

A Fusion Technique Based on Image - Statistical Analysis for Detection of Throat Cancer Types

Adnan Al-Bashir [a,*], Bassam Al-Naami [b]

Department of Industrial Engineering, Hashemite University, P.O. Box 150459, Zarqa 13115, Jordan

Department of Biomedical Engineering, Hashemite University, P.O. Box 150459, Zarqa 13115, Jordan

Abstract

The aim of this study is to establish a simple approach to classify the throat tumor type by using statistical analysis techniques without a need for a biopsy or further testing. In this study, around 35 patients were investigated to be classified and to provide appropriate diagnostic for the throat cancer type. MRI images and their properties were processed and converted into number of pixels and intensities that located in the region of interest which covered the area of tumor. This extracted information was employed by traditional statistical methods such as Descriptive Analysis, Box Plots and Testing of Hypothesis to enable reasonable accuracy in differentiation between the tumor types. As a result of this study, it is observed that the using of statistical analysis for the data taken from the MRI images is reliable to diagnose and determine the type of the throat cancer with 95% confidence.

Keywords: Throat cancer (TCa), filtration, segmentation, box plot, T-test, Inter Quartile Range (IQR) .

1. Introduction

Throat cancer is one of the most dangerous cancer's type and forms in tissues of the pharynx (the hollow tube inside the neck that starts behind the nose and ends at the top of the windpipe and esophagus). Throat cancer includes cancer of the nasopharynx (the upper part of the throat behind the nose), the oropharynx (the middle part of the pharynx), and the hyp-opharynx (the bottom part of the pharynx). Cancer of the larynx (voice box) may also be included as a type of throat cancer. Most throat cancers are squamous cell carcinomas (cancer that begins in thin, flat cells that look like fish scales). Also called pharyngeal cancer (Fig.1) Estimated new cases and deaths from throat cancer (including cancers of the larynx) in the United States in 2009 are 12,290 (laryngeal) and 12,610 (pharyngeal), while the death cases are 3,660 (laryngeal) and 2,230 (pharyngeal) [1].

In 2004, according to the Jordan National Cancer Registry, (JNCR), about 3,591 new cancer cases have been registered among Jordanians with an incidence rate of 67.1 per 100,000 populations (63.9 for males and 70.5 for females). Among the most common cancers affecting Jordanian population, the throat tumor (TCa) was ranked as the eighth common type in children and the 10th in adults and the prevalence in males 3.5% more than females (2.1%) [2].

Different diagnostic procedures have been followed in attempt to differentiate between the benign and malignant tumor such as: 1) Physical exam. 2) Indirect laryngoscopy; the doctor looks down your throat using a small, long-handled mirror to check for abnormal areas and to see if your vocal cords move as they should. 3) Direct laryngoscopy; the doctor inserts a thin, lighted tube called a laryngoscope through your nose or mouth. As the tube goes down your throat, the doctor can look at areas that cannot be seen with a mirror. 4) Biopsy is removing tissue to look for cancer cells and a pathologist then looks at the tissue under a microscope to check for cancer cells [1]. A biopsy is the only sure way to know if a tumor is cancerous.

However, it is important in many cases to validate a diagnosis and be certain of its accuracy. On the other hand, hoping for a misdiagnosis should not be used as a way to avoid treatment for a serious medical problem as the throat cancer. Nevertheless, it is sensible to attempt to confirm a diagnosis via methods such as seeking second opinions, consulting specialists, getting further medical tests, and researching information about the medical condition. Also, misdiagnosis can and does occur and is reasonably common with error rates ranging from 1.4% in cancer biopsies to a high 20-40% misdiagnosis rate in emergency or ICU care.

[http://www.cureresearch.com/intro/overview.htm].

* Corresponding author. abashir@hu.edu.jo

Therefore, in this paper, we focused on building an algorithm – software based on the use of LabView software to analyze the image of Throat Cancer, where part of this technique was tested in our previous work for brain tumor classification [3]. This target was achieved by the developed algorithm consisted of image enhancement, filtering and applying the region of interest threshold techniques to extract the number of pixels and their intensities for both types of throat cancers (benign and Malignant). This procedure is followed by the statistical analysis based on Box Plot and Test of Hypothesis.

Various segmentation techniques and methods have been cited in the literature for improving the segmentation processes to maximize the possibility and reliability of TCa classification. These techniques can be categorized into:

1) threshold-based segmentation, 2) statistical methods for TCa segmentation and 3) region growing methods [4, 5, 6, 7].

In [4] a semiautomatic system for segmentation of a diverse set of lesions in head and neck CT scans has developed. The system takes as input an approximate bounding box, and uses a multistage level set to perform the final segmentation. Then contours from automatic segmentation were compared to both 2D and 3D gold standard contours manually drawn by three experienced radiologists. The average absolute area error was 21.1% compared to 10.8%, and the average 2D distance was 1.38 mm compared to 0.84 mm between the radiologists. The automatic contours approximated many of the lesions very well. Haibo Zhang et al. developed a new three-dimensional adaptive region growing algorithm for the automatic segmentation of three-dimensional images [5]. The principle of this algorithm is to obtain a satisfactory segment result by self-tuning the homogeneity constraint step by step. Results of segmentation based on the use of this algorithm are close to that of manual segmentation. Shiping Zhu et al. proposed a new segmentation algorithm that each pixel in the image has its own threshold [6]. In this algorithm, the threshold of a pixel in an image is estimated by calculating the mean of the grayscale values of its neighbor pixels, and the square variance of the grayscale values of the neighbor pixels are also calculated as an additional judge condition. The results demonstrate that the proposed algorithm could produce precise image edge, while it is reasonable to estimate the threshold of a pixel through the statistical information of its neighbor pixels. In [7] a novel object identification algorithm was developed in Java to locate immune and cancer cells in images of immune histochemically- stained lymph node tissue. This algorithm focuses on the interactive feature extraction from color images, so that the classification is improved with an interactive visualization system. Then, in order to increase the accuracy it coupled with the statistical learning algorithms and intensive feedback from the user.

Statistical methods represent another important category in the segmentation process and most of the approaches proposed in this category were using some statistical classifications combined with different image processing techniques in order to segment the MRI images [3, 8, 9, and 10].

One of the reliable and suited MRI and CT segmentation techniques is the region growing that can be applied to TCa tumor is generally, presented in many studies [3 and 11-16]. Cheng-Long Chuang et al. proposed an image-based object segmentation algorithm for extracting tumor-like objects in CT images based on intensity regions automatically in a CT volume image [15]. Mancas et al. provided segmentation using region growing threshold; this technique used threshold and spatial information to segment the region of interest [11]. Jiang et al. also provided producer for segmentation and quantification of brain tumor, by semi automatically detecting the area of the tumor in the brain that help the radiologist to treatment [16].

This paper is organized as following: a review about the TCa types and the related work was presented. A detailed description of image preprocessing and extraction parameters from the ROI, data collection, and statistical analysis for the collected data is described in second section. The last section describes the experimental results and discussion and concludes the paper along with outline future direction.

2. Methodology

In this study, the processed images of TCa are divided into two different groups:

The first group is representing the sample of 21 (n_1) images which was already diagnosed as a malignant tumor , and the second group is representing the sample of 12 (n_2) diagnosed as benign tumor were collected randomly. All data were provided by the Hussein Medical City hospitals and the Islamic hospital in 2006 and 2007 respectively, Amman, Jordan. The common procedure of determining throat tumor type was by examining the patient's MRI images by an experienced radiologist. According to these comments, that would be our reference later on, the type of tumor will be decided which may not be accurate as mentioned above as misdiagnoses. Our proposed approach is about using the statistical analysis of the data obtained form MRI images and makes inferences to distinguish between the two different types of tumor which may be more accurate than the traditional method and be supporting technique for the physicians.

2.1. Preprocessing of image data

MRI image should be prepared and treated by applying the steps:

- Convert all images to the gray scale
- Enhancement image to improve quality
- Image segmentation to find the region of interest (ROI).
- Histogram of region of interest (ROI)
- Apply statically process on (ROI) to find the final result that we need to make the comparison between malignant and benign tumors based on the use of test of hypothesis.

The tumor of TCa is surrounding by others part of the neck (nasopharynx, oropharynx, pharynx and the hyp-opharynx) that looks like tumor in the normal form. Hence, the region that covers the tumor should be carefully enhanced and improved so that after applying the

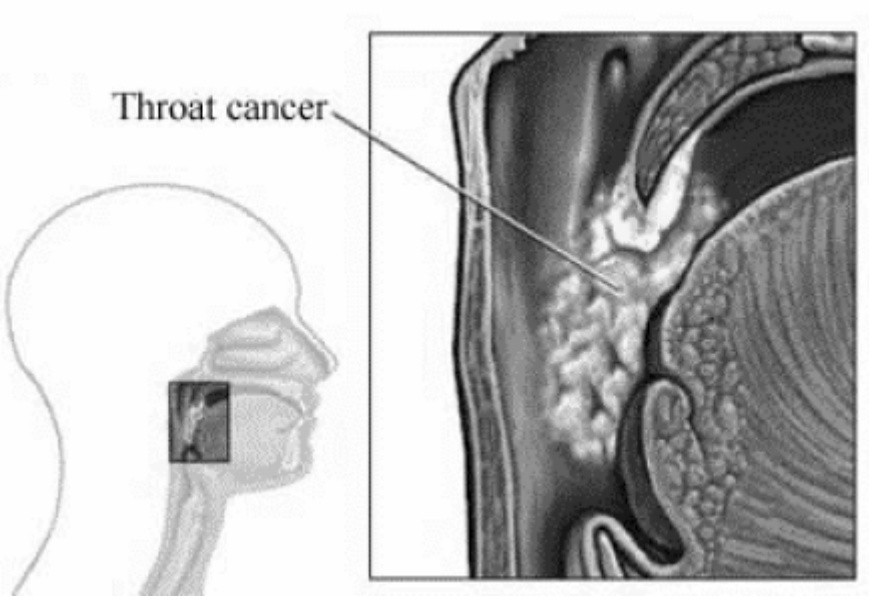

Figure 1: Throat Cancer Tumor

segmentation technique we could extract the needed features accurately.

Some of the images characteristics such a small size, or low contrast can be seen, but this problem was solved by using the special low pass filter. The clarity of the filtered image depends on many factors such as filter mask, original image, boundary option and the standard parameter called alpha. These entire factors were considered in performing the filtration process. The low pass filter is used and implemented using the LabView Software. It allows low frequency data, or data that does not change much from pixel to neighboring pixel, to pass through, For an image that contains a lot of noise, such a filter would smooth out the image and reduce the noise with minimal affect on large features in the image. A low pass filter will affect large features in the image, and will reduce or eliminate the smaller features [17, 3].

The next step is the image segmentation.

In the normal cases, when the throat image of the patient arrives to the radiologist, he studies the up normal regions in the image to discover the type of the tumor. So, the diagnostic depends on segment these regions. Segmentation process is to divide an image into its constituents regions or objects, then take a specific region; it is called the region of interest (ROI**).** Histogram is a process followed by segmentation. It counts the total number of pixels in each grayscale value and graphs it.

2.2. Statistical Analysis

For each MRI image the data is collected and tabulated. Analysis of such data will be demonstrated in two ways; the descriptive analysis utilizing the Box Plot presentation, and the testing of hypotheses approach. A hypothesis stated that there is no difference between the malignant and benign will be tested using the difference between two sample t- test.

The test statistics is a single number that calculated from the sample mean, which can be located in the rejection region or in the acceptance region. Depending on the location of test statistics the null hypothesis could be rejected or accepted as illustrated in figure 2.

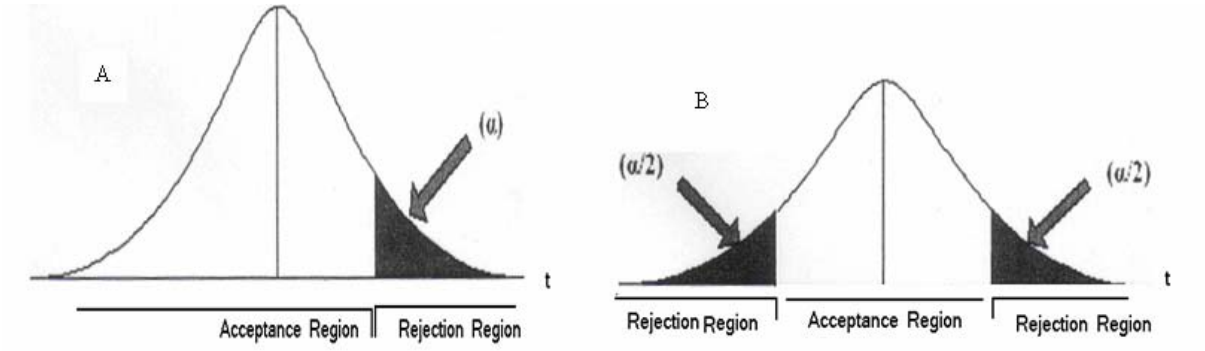

Figure 2: Rejection and acceptance regions for one tail and 2-tail test

  - Volume 4, Number 6 (ISSN 1995-6665)

Applying the test of hypotheses is done by following these steps [18-20]:

- Null hypotheses H_o :(μ_m- μ_b) =D_0, where D_0 is some difference between the mean values .and we will consider D_0 =0.

Test statistics (t_0)

$$t_0 = \frac{(\bar{X}_1 - \bar{X}_2) - D\circ}{S_p\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \quad (1)$$

$$s_p = \sqrt{\frac{(n_1-1)s_1^{\,2} + (n_2-1)s_2^{\,2}}{n_1+n_2-2}} \quad (2)$$

Where *Sp*: pooled standard deviation.

d_f : degree of freedom;

$$d_f = n_1 + n_2 - 2 \quad (3)$$

- Find The critical value of t_α *or* $t_{\alpha/2}$ depending on α (type I error)
- Compare the two values of the test and decide on H_0

3. Result and Analysis

3.1. Image Preparation

Figures 3 and 4 show the image processing result for malignant and benign tumors respectively contain filtered image and histogram for the selected ROI that contain many statistical variables. It is noted that the output image have higher quality than input image, also there is no difference between applying the enhancement on malignant or benign.

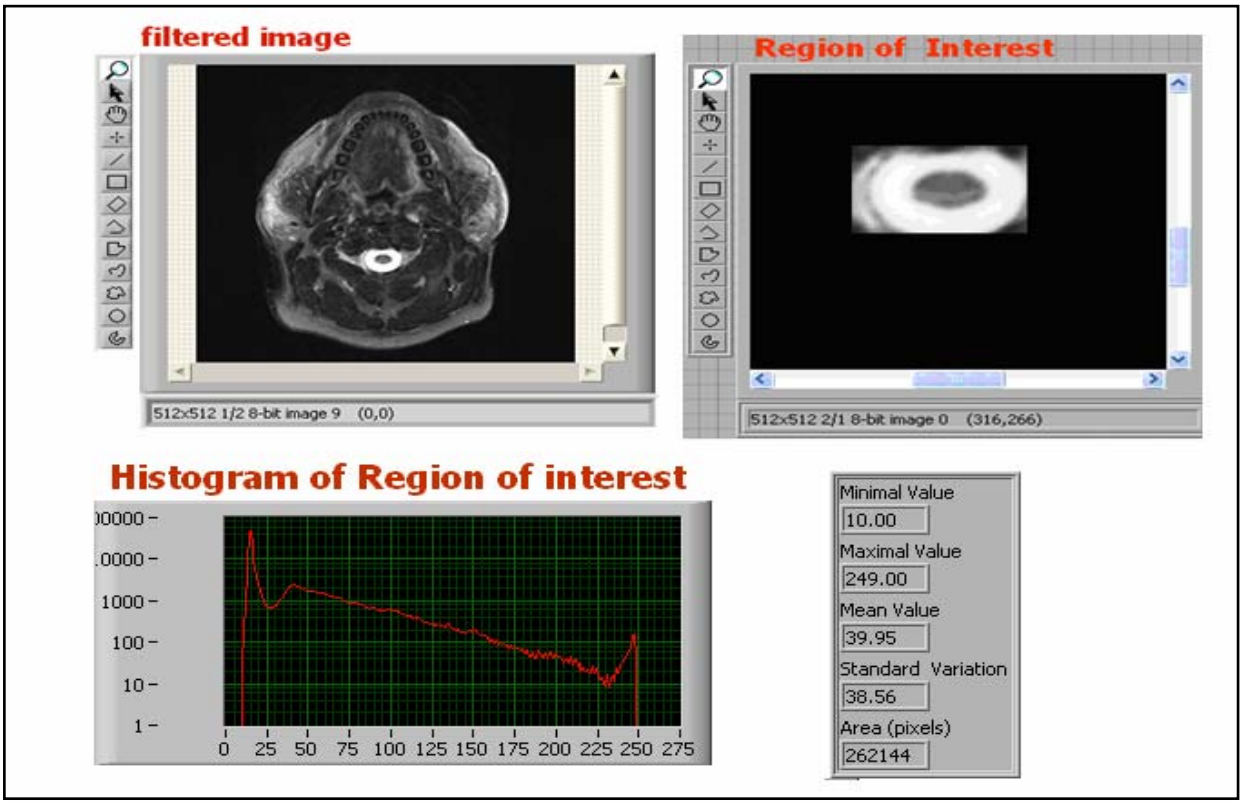

Fig. 3 Image processing result of malignant tumor

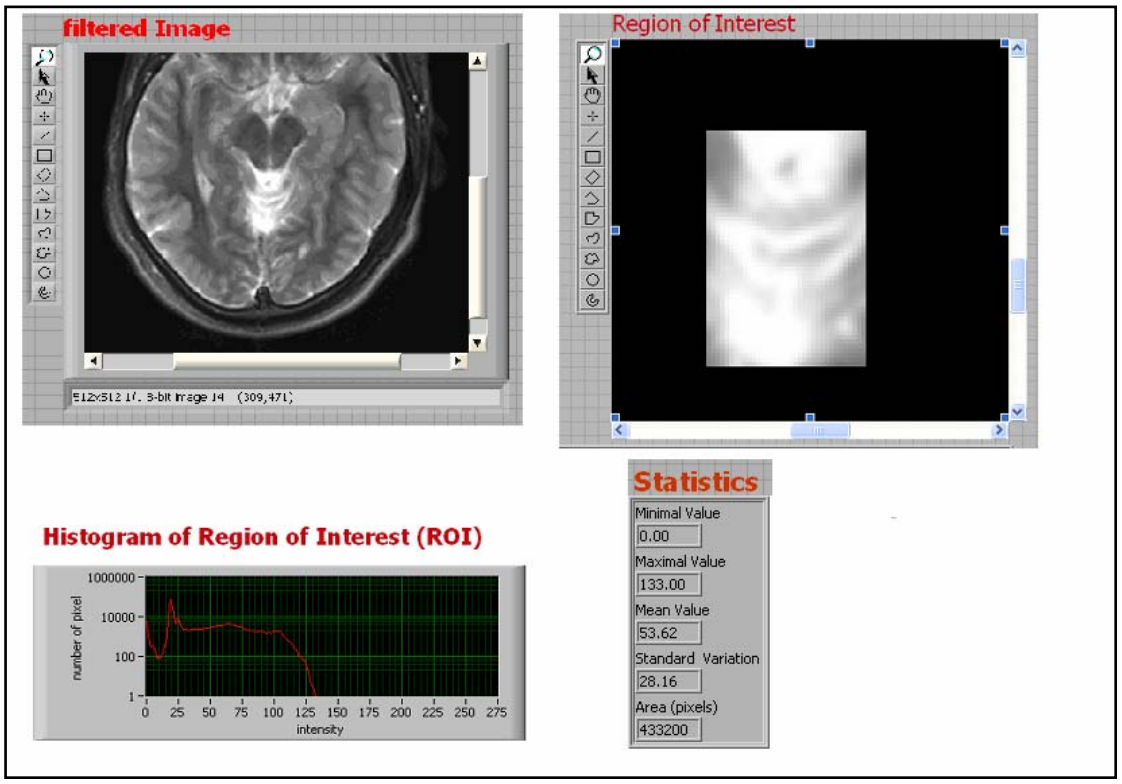

Fig. 4 Image processing result of benign tumor

By comparing figure 3 and figure 4, the two figures, the histogram for ROI in both malignant and benign tumors are different, since the histogram of region of interest in malignant image reaches high level of pixel value than the histogram of region of interest in benign image

4. Statistical Result

Statistical analysis for all images data was performed in two stages; the first stage is the descriptive analysis and construction of the box plot for each type of the tumor, while the second stage id to perform the testing of hypothesis on the mean of the two types.

4.1. Descriptive Result

The average pixel value, the standard deviation and the range for the malignant and for the benign were calculated for the two sample data available. it is noted that there is a difference between the malignant and benign image in many parameters such as, the mean pixel for the malignant image is higher than in the benign images.

Table 1 shows the summary of the descriptive analysis for the maximum and average pixel values in all 21 images and 12 benign images.

The above table shows that there is a difference in the mean value of the averages between the malignant and the benign tumor images. The mean for the malignant images was 95.58 while it was 40.1 for the benign images. It is also noted that the value of the maximum number in benign tumor is small than max value of the malignant tumor and the average mean of the pixel is below from 100.

The Box Plot in figure 5, clearly indicates the same results that the average value of the malignant tumor which are in the range of (80-115) is higher than the average values of the benign tumor which are in the range of (22-58).

The mean value for the maximum values in malignant images was 227, while the mean value for the maximum values in benign was 163. The standard deviation and median for the malignant is higher than the benign images. Moreover, the maximum and minimum in the malignant is higher than the benign. All the above descriptive give an indication that the malignant images are differ from the benign.

The Box Plot in figure 6, clearly indicates the same results explained in table 1, that is the maximum values of the malignant tumor which are in the range of (208-256) is higher than the maximum values of the benign tumor which are in the range of (135-190).

The difference between the two samples is clearly obvious as is shown in the figures above since the malignant images have a box plot with higher range and higher spread than the benign box plot. Considering the average pixel value to be the criterion to distinguish between the two tumors types, it is clear that the malignant type have greater pixel values than the benign one. Also, the dispersion for the benign tumor is smaller than the malignant one.

Table:1 Statistical parameters for maximum and average pixel values for malignant and benign

Subject	Maximum Pixel Values		Average Pixel Values	
Parameters	Malignant Tumor	Benign tumor	Malignant Tumor	Benign tumor
Mean	226.76	163.33	95.58	40.1
Maximum	255	206	120.93	73.29
Minimum	177	123	50.11	14.5
Standard deviation	26.6	29.85	20.02	19.28
Variance	707.79	890.97	400.7	371.85
The lower quartile (Q1)	203.5	134.5	77.93	23.33
the upper quartile (Q3)	251	191.5	114.02	53.09
Range	78	83	70.82	59.04

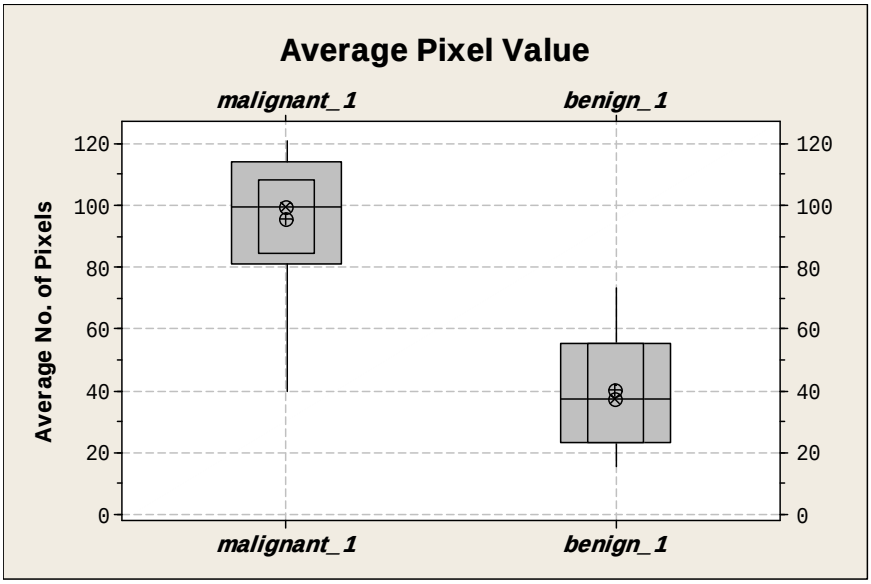

Figure 5: Box plot for the average pixel values

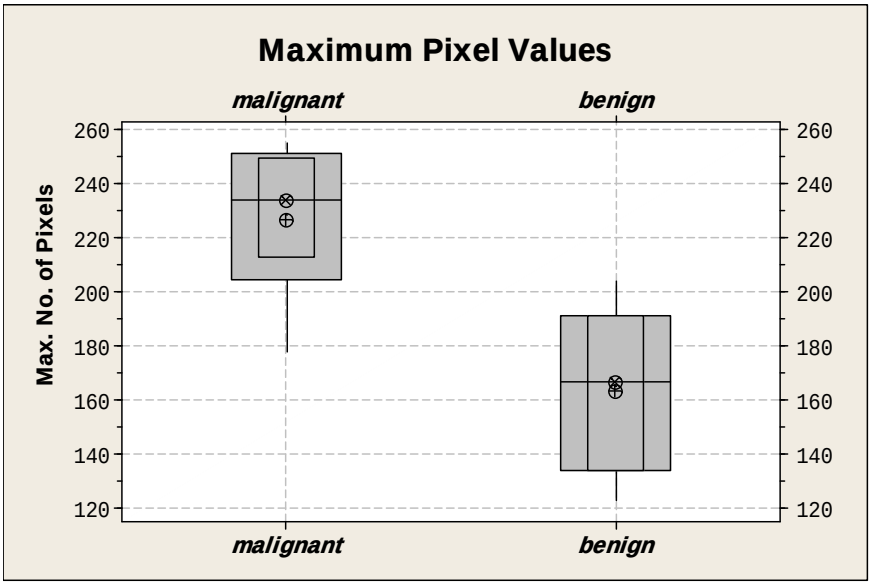

Figure 6: Box plot for the maximum pixel values

4.2. Testing of the Hypothesis

In the pervious section, the information in table 1, is needed to perform the test of hypothesis that say the mean of maximum number of pixel and the mean of average number of pixels in the malignant tumor is higher than from the benign ones .A two sample t – test will be used because the number of the two sample is different and the samples were drawn from a population with unknown variances. .

4.2.1. Test of the hypothesis for mean average number of pixel

In order to make a conclusive decision about the difference between the malignant and the benign images, the null hypothesis is tested against the alternative one as below:

$H_0 : (\mu_m = \mu_b)$

$H_a : (\mu_m \neq \mu_b)$,

Where: H_0: is the null hypothesis

H_a: is the alternative hypothesis

μ_m is the mean of the average number of pixels in malignant images

μ_b is the mean of the average number of pixels in benign images.

Since the two samples have a probability plot approximately linear, the (t) test for difference between two sample means will be used in testing the null hypothesis. If t- calculated value> t- tabulated value, we reject H_0 and accept H_1. The results for using the two sample (t) test are shown below; these parameters are used to calculate the (t) statistics.

Table 2: Parameters used in the t test

Parameters	Value
Mean of M	226.8
Mean of B	163.3
Standard deviation S_m	26.6
Standard deviation S_B	29.8
t- calculated value	6.31
Pooled standard deviation S_p	27.7991
Significant level α	0.05
t- tabulated value	1.696
Degree of freedom d*f*	31
Number of sample n1	21
Number of sample n2	12

From table 2, the t- calculated value (6.31) > t- tabulated value (1.696), and since the t (calculated) is located in the rejection region, we reject the null hypotheses H_0 and strongly conclude that there is a difference between the mean value of malignant and the mean value of benign images with 95% confidence, this means that type I error is 0.05. Then the average value of the number of pixel of malignant tumor is higher than the average value of the number of pixel of benign.

4.3. Test of the hypothesis for mean maximum number of pixel

Another important testing of hypothesis is that whether the maximum pixel value of the malignant is equal to the maximum pixel value of the benign image.

$H_0 : (\mu_{mm} = \mu_{mb})$

$H_a : (\mu_{mm} > \mu_{mb})$,

Where: H_0: is the null hypothesis

H_a: is the alternative hypothesis

μ_{mm}: is the mean of max. Pixel value in malignant images

μ_{mb} : is the mean of max. Pixel value of benign images

. The data and the results for using the two sample (t) test are shown below; these parameters are used to calculate the (t) statistics.

Table 3: Parameters used in the t test

Parameters	Value
Mean of M	95.6
Mean of B	38.5
Standard deviation S_m	20
Standard deviation S_B	19.3
t- calculated value	7.98
Pooled standard deviation S_p	19.7601
Significant level α	0.05
t- tabulated value	1.309
Degree of freedom d*f*	31
Number of sample n1	21
Number of sample n2	12

From the table the t- calculated value (7.98)> t- tabulated value (1.309) and since the t(measured) is located in the rejection region, we reject the null hypotheses H_0 and strongly conclude that the maximum pixel value for malignant is greater than the maximum pixel value for benign images with 95% confidence. Then the average of average value of the number of pixel of malignant tumor is higher than the average of average value of the number of pixel of benign.

5. Conclusion

In this study, our goal is to automatically and statistically diagnose the type of tumor in the throat by using MR images. In order to do this, we addressed simple box plot technique to differentiate between the two types. Also, testing of hypothesis was applied for the same purpose. Region of interest, filtration and segmentation techniques were utilized to be the base of the information to get the statistical data for each case. It has been proven that a simple, harmless and accurate statistical technique can efficiently distinguish between the malignant and benign tumor (TCa). The huge advantage of this approach is that there will be no need to make any further tests or examinations on the patient after making the MRI. Some

of these tests or examinations are difficult to be performed or may be dangerous such as throat biopsy.

The diagnosis system achieves accuracy over 95.0% in differentiation between the different tumor types on various qualities of MR images. Although some satisfactory results are obtained, the diagnosis system still needs to be improved. The results of the experiments and application of the proposed method may also be applied to the other areas of medical image analysis. So, another future task is to apply this proposed method to other areas in medical image analysis to improve the achieved accuracy in this study and the previous one in [3].

Some weaknesses of the algorithm are the small area of the overlap between the extreme values of the two different tumor types. To overcome this drawback, it is suggested to increase the number of samples used for the study.

This automatic statistically based diagnosis method is developed to classify the images of large medical databases. Here, this simple and fast diagnosis method may be used to extract information from the large medical databases. An important future work would be measuring the performance of this method in mining medical databases.

References

[1] National Cancer Institute, USA. http://www.cancer.gov.

[2] King Hussein Cancer Centre, Jordan. http://www.khcc.jo.

[3] B. Al-Naami, A. Bashir, H. Amasha, J. Al-Nabulsi, and A. Almalty, "Statistical Approach for Brain Cancer Classification Using a Region Growing Threshold". Journal od Medical Systems, DOI: 10.1007/s10916-009-9382-6.

[4] S. Ethan, H. Lubomir, S. Berkman, G. Sachin, I. Mohannad, K. Suresh and C. Heang-Ping, "Automated volume analysis of head and neck lesions on CT scans using 3D level set segmentation". Med Phys. Vol. 34, No. 11, 2007, 4399–4408.

[5] Z. Haibo, S. Hong, D. Huichuan, "An Automatic and Robust Algorithm for Segmentation of Three-dimensional Medical Images". Sixth International Conference on Parallel and Distributed Computing Applications and Technologies (PDCAT'05), Dalian, China, 2005.

[6] Z. Shiping, X. Xi, Z. Qingrong, B. Kamel, "An Image Segmentation Algorithm in Image Processing Based on Threshold Segmentation". Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, Shanghai, Chaina, 2007.

[7] H. Susan, K. Adam and P. Peter, "An Interactive Java Statistical Image Segmentation System". GemIdent, Journal of Statistical Software, Vol. 30, No 10, 2009, 1-20.

[8] M. Justin, K. Iftekharuddin, "Statistical analysis of fractal-based brain tumor detection algorithms". Magnetic resonance imaging, Vol. 23, No. 5, 2005, 671-678.

[9] M. Ashraf, I. Evangelia, S. Dinggang, D. Christos, "Deformable registration of brain tumor images via a statistical model of tumor-induced deformation". Medical Image Analysis, Vol. 10, No. 5, 2006, 752-763.

[10] X. Xiao, L. Qingmin, "Statistical Structure Analysis in MRI Brain Tumor Segmentation. Image and Graphics", ICIG. Fourth International Conference, Sichuan, Chaina, 2007.

[11] M. Mancas, B. Gosselin, B. Macq, "Segmentation Using a Region Growing Thresholding", Proc. of the Electronic Imaging Conference of the International Society for Optical Imaging SPIE., San Jose (California, USA), 2005.

[12] M. Sezgin, B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation". J. Electron. Imaging Vol. 13, No. 1, 2004, 146-165.

[13] M. Chowdhury, W. Little, "Image thresholding techniques" IEEE Pacific Rim Conference on Communications, Computers, and Signal Processing, Victoria, BC, Canada, 1995.

[14] Z. Pan, L. Jianfeng, "A Bayes-Based Region-Growing Algorithm for Medical Image Segmentation. Computing in Science & Engineering, Vol. 9, No. 4, 2007, 32 – 38.

[15] C. Cheng-Long, and C. Chun-Ming "A Novel Region-based Approach for Extracting Brain Tumor in CT Images with Precision. World Congress on Medical Physics and Biomedical Engineering. IFMBE Proceedings, Vol. 14, No. 4, 2006, 2488-2492.

[16] J. Chunyan, Z. Xinhua, H. Wanjun, C, Meinel, "Segmentation and quantification of brain tumor, Virtual Environments". IEEE Symposium on Human-Computer Interfaces and Measurement Systems, Boston, USA, 2004.

[17] Rafael C. Gonzalez, Richard E. Woods and Steven L. Eddins. Digital image processing, Second Edition , Prentice Hall, Upper saddle River, new Jersey, USA, 2004.

[18] Doouglas C. Montgomry and George C. Runger; Applied Statistics and Propability for Engineers, Fourth Edition, John Wiley and Sons, New York, USA, 2008.

[19] Ronald E. Walpole, Roymond H. Myers, Sharon L. Myers and Keying YE; Probability and statistics for Engineering and Scientists, Eighth Edition, Prentice Hall, Upper saddle River, new Jersey, USA, 2008.

[20] Williams W. Hines and D. Montgomery, Probability and Statistics in Engineering and Management Science, Forth edition, John Wiley & Sons, Inc., New York, USA 2006.