Revisited Otsu Algorithm for Skin Cancer Segmentation

ALAA TOM, JIHAD DABA
Department of Electrical Engineering,
University of Balamand,
LEBANON

Abstract: - Computer-aided technology can be used to perform a quantitative and objective evaluation of pigmented skin lesions during the clinical assessment procedure. This helps to expedite the procedure. The growing development of non-invasive techniques can be of significant benefit in the early identification of malignant melanoma, which can, in turn, help to minimize the necessity for invasive biopsies. The system is primarily focused on two principal schemes: Establishing an effective lesion border detection method and then creating an efficient classification scheme. We address two primary areas in this work. First, we study skin lesion detection to analyze any sign of malignancy for skin cancer diagnosis. This is followed by the system implementation of a color-based method for all the images from the RGB color space using a revisited OTSU thresholding segmentation scheme. The results proved to be promising with at least an 80% accuracy detection rate for a wide range of clinical skin lesion images.

1 Introduction

One of the most deadly illnesses and the main cause of death worldwide is cancer. According to WHO (World Health Organization) statistics, cancer killed 10 million people globally in 2020, [1], and by 2030, there would be 13.1 million cancer-related deaths, according to predictions, [2].

With 1.2 million cases expected in 2020, skin cancer is the fifth most prevalent malignancy overall, [2]. It is divided into three types: Skin cancer that is not melanoma, such as basal cell and squamous cell cancer, and melanoma skin cancer that affects melanocytes. Squamous cell cancer and basal cell cancer are both common but less dangerous. Compared to non-melanoma skin cancers, melanoma is less common, but it has a higher risk of spreading and becoming lethal, [3].

Although skin cancer is a life-threatening disease, with a poor prognosis, especially for melanoma, it may be treated if caught early. According to studies, the cure rate is more than 90% if melanoma is detected early on, whereas the cure rate is less than 50% if it is detected late, [4]. The largest phenotypic risk factor for this condition is the presence of a significant number of melanocytic nevi, especially in the case of atypical cutaneous syndrome. Other risk factors include fair skin, a high density of freckles, and eyes and hair colour. Melanoma is currently diagnosed mostly by a dermatologist by examining the skin with his naked eyes or with a derma-scope during the diagnosing process.

Dermatologists commonly utilize the ABCDE (Asymmetry─Border─Color─Diameter─Evolution) criteria to discover signs of skin cancer during this stage, [5]. Asymmetry, borders, colors, diameters, and evolution (or evolving) are all characteristics of a skin patch, [5]. Dermatologists can determine if a skin patch is benign or malignant by analyzing the early indications and applying the ABCDE approach, [3].

Computer-aided skin cancer detection approaches have been studied to help dermatologists in melanoma diagnosis to enhance the diagnostic rate. The main idea behind computer-aided detection (CAD) is to detect questionable patches of the skin using image processing and pattern recognition techniques. It is commonly utilized as a dermatologist's recommendation. In this work, we are mainly interested in the primary approach used in CAD systems to identify skin cancer.

2 Literature Survey

The detection of skin cancer using CAD systems has been a topic of intensive research for more than 30 years, [6]. For instance, numerous melanoma detection techniques have been created and researched, [7]. A fascinating overview of methods that have been tested on clinical and dermoscopic images from 1984 to 2012 was provided by Korotkov and Garcia in [8]. They arranged the review from the data acquisition stage to the classification and diagnosis using ABCDE criteria, [7], as well as several other methods created in clinical and CAD systems. Additionally, Maglogiannis and Doukas reported in [9] a non-exhaustive comparison of the most cardinal implementations, particularly characteristics selection like colour and border, and they offered the most widely used classifiers in the literature: ANNs (Artificial Neural Networks) and SVMs (Support Vector Machines).

Abbas et al. in [11] employed modified RACs (Region-based Active Contours) created by Lankton et al. in [10] using total variation regularization for the segmentation part. If the boundary lesion was visible and had a strong contrast with the surrounding normal skin, their approach could segment the lesion border accurately. It did not, however, pick up lesions that blend seamlessly into the healthy skin around them.

Another strategy utilized by Ma et al. in [12] is wavelet decomposition banks for artificial neural network classifiers for melanoma and non-melanoma case discrimination. 72 melanoma and 62 benign tumors from a database were subjected to this scheme, making a total of 134 skin lesion images. The drawn results reached a sensitivity of 90% and a specificity of 83%.

Celebi et al. proposed in [13] an automated fusion of the thresholding method with a Markov random field; they examined 90 dermoscopic images with 23 malignant melanoma and 67 benign lesions; they then compared their findings to the state-of-the-art methods, expressing the results using exclusive-OR errors of 9.16 ± 5.21 %.

3 Methodology

Our study is divided into four major stages as depicted in the block diagram in Fig. 1. The first stage comprises the pre-processing stage comprising image enhancement, resizing, and getting rid of all unwanted details. The second stage is the segmentation stage consisting of segmenting the skin lesions from the input image and performing image processing techniques. The third stage consists of the feature extraction stage including the ABCD parameters and some other parameters such as area, perimeter, and axis. And finally, we employ the TDS calculation stage which states that the lesion is either benign, suspicious, or malignant.
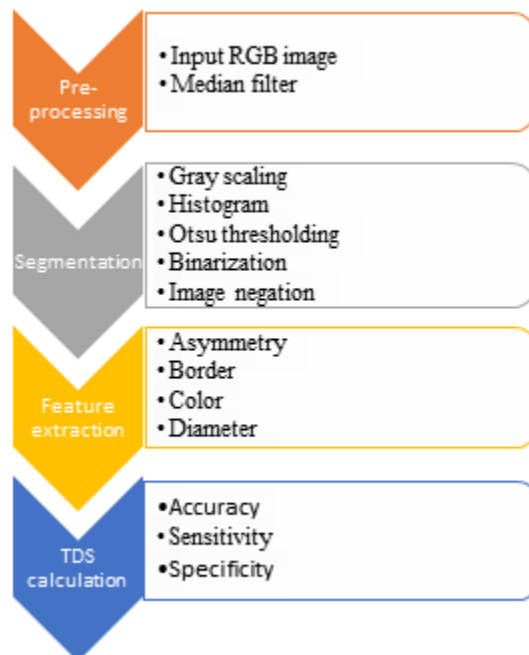

Fig. 1: Block diagram of the proposed algorithm

3.1 Preprocessing

The majority of dermoscopy images need some adjustments when it comes to brightness, contrast, thick hair, and any unwanted detail such as air bubbles, etc. For this matter, a median filter is used to get rid of every detail considered noise.

3.1.1 Data collection

Any image processing technique starts with acquiring the image, which is the process of selecting an image from the databases selected before.

3.1.2 Image input

The acquired image is usually in the RGB color space because all pictures are coded according to the jpg or jpeg standards.

Alaa Tom, Jihad Daba

3.2 Segmentation

The purpose of the segmentation process is to make work easier or to convert the image into a representation that is easier to comprehend and evaluate. Using Otsu's thresholding, a grayscale image can be converted into a binary image for clustering-based image segmentation.

Otsu's threshold-based segmentation is employed as the technique for carrying out the segmentation work for our method.

3.2.1 Gray scaling

The bulk of jpg and jpeg photos are in the RGB color space, making it one of the most frequently used color spaces, [14]. The RGB images are converted into gray images, meaning that the red, blue, and green values are averaged to get the intensity of the grayscale image using the following equation:

$$\text{Grayscale} = \frac{R+G+B}{3}, \qquad (1)$$

where R is the value of the red pixel, G is the value of the green pixel, and B is the value of the blue pixel.

3.2.2 Histogram

A critical stage before the thresholding process is the construction of the histogram of image pixel distribution. In the histogram analysis, we should distinguish between the foreground and the background. The pixel probability distribution of each level is computed by the following equation:

$$P_i = \frac{n_i}{N}, \qquad (2)$$

where P_i is the pixel probability to i, n_i is the pixel with grayscale level i, and N is the total number of pixels in the image.

3.2.3 Binarization

After gray scaling the image, we should perform binarization, which is transforming the gray image into a binary image. The binary image consists of 2-pixel values: either 0 or 1. The importance of the binary image is the increase in efficiency and decrease in load. To achieve this step, there are two ways: Global binarization and adaptive binarization. Global binarization consists of one threshold value

for all the images, while adaptive binarization consists of a changing threshold value for every pixel, [10].

3.2.4 Image negation

The act of replacing each white pixel (of value 1) with a black pixel (of value 0), and vice versa, is known as image negation. This process assures a better representation of the ROI (Image Region of Interest), where the ROC is shaded in white.

3.3 Feature Extraction

The ABCD rule is utilized in the process of distinguishing benign melanocytic tumors from malignant melanocytic tumors. The asymmetry, border, color, and diameter are the four factors that are considered while applying the ABCD formula to get the TDS (Total Dermoscopy Score) index.

3.3.1 Asymmetry

Asymmetry is a crucial element that plays a fundamental role in computer-aided diagnosis. The suspicious mole is evaluated in 0, 1, or 2 axes. In other words, the lesion is cut in half and the two parts should match one another, regardless of the axis of symmetry. Otherwise, there is an asymmetry score "A" computed as follows:

0: bi-axial symmetry
1: mono-axial symmetry
2: bi-axial asymmetry

3.3.2 Border

The border score "B" ranges from 0 to 8. To compute the border irregularity, we should find the perimeter and the area of the acquired image. And then B is determined using the following equation:

$$B_i = \frac{p^2}{4\Pi a}, \qquad (3)$$

where p denotes the lesion's perimeter, a denotes the lesion's area and Π is a controlling parameter set at 22.7.

3.3.3 Color

The RGB color space's three components are distributed among the color parameter "C". When it comes to scoring, the value ranges from 1 to 6, indicating the presence of up to six well-known

colors, such as black, white, slate blue, brown, and red.

3.3.4 Diameter
A cancerous mole usually consists of a diameter "*D*" wider than 6 mm. If this occurs, then the mole is most probably a sign of malignancy, and automatically we set *D* to a value of 5.

3.4 TDS Calculation
The total dermoscopy score or TDS index is an essential criterion for melanoma diagnosis. The ABCD parameters of the skin lesion are the four factors that go into the calculation of the TDS index. If the TDS is below 4.75, the lesion is considered to be benign. If the TDS is above 5.45, this is an index that the skin growth is malignant. If the TDS is between 4.75 and 5.45 the skin lesion is suspicious.

The TDS index can be found by the formula below:

$$TDS = 1.3A + 0.1B + 0.5C + 0.5D. \qquad (4)$$

3.4.1 Accuracy Testing
When a testing technique correctly identifies a fraction of its true positives (TP), as suggested by Eq. (5) below, it is said to have high sensitivity, which is a measurement of how well a test can detect disease. The specificity of a diagnostic test refers to the percentage of true negatives that it correctly identifies. It indicates how well the test can distinguish bad situations (normal conditions). Accuracy is defined as the percentage of true outcomes in a population, which may be true positive (TP) or true negative (TN). It evaluates the accuracy of a diagnostic test when used to treat a condition and then summarizes the findings.

Sensitivity, specificity, and accuracy are described in Eqs. (5), (6), and (7), respectively, in terms of TP (true positive assessments), TN (true negative evaluations), FP (false positive evaluations), and FN (false negative evaluations):

$$Sensitivity = \frac{TP}{TP + FN}, \qquad (5)$$

which denotes the number of true positive assessments over the number of all positive assessments;

$$Specificity = \frac{TN}{TN + FP}, \qquad (6)$$

which denotes the number of true negative assessments over the number of all negative assessments;

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}, \qquad (7)$$

which denotes the number of correct assessments over the number of all assessments.

As stated earlier, we took a sample of 30 images in our study. Eighteen of the selected images were melanoma cases and the other twelve were non-melanoma. After testing, 15 images were analyzed as TP, 6 as TN, 3 as FP, and 2 as FN. Table 1 highlights the resulting accuracy parameters.

Table 1. Accuracy parameters

Sensitivity	Specificity	Accuracy
88.2%	66.7%	80.1%

4 Results and Discussions
The Otsu threshold was developed in MATLAB for this work, which involves the segmentation of dermoscopy images in melanoma. Figure 2 is an illustration of an Otsu thresholded dermoscopy image with a resolution of 1021 by 766 pixels.

During the initial step of this research, a dermoscopy image of melanoma was entered into the system so that it could be converted from an RGB image to a grayscale image that had been calculated earlier. The result of effectively converting an RGB image into a gray level is shown in Fig. 3.

Fig. 2: Input image

Fig. 3: Gray-scaled image

The development of a histogram from a grayscale image is the next step to follow. Creating this histogram is helpful before moving on to the next segmentation stage. The histogram of the grayscale image is depicted in Fig. 4.
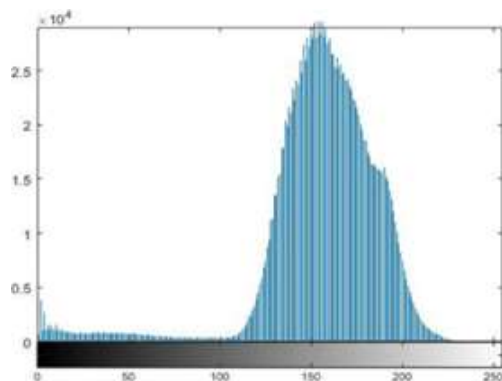

Fig. 4: Histogram analysis

Image segmentation with Otsu Thresholding is applied to the grayscale image after the histogram of the image has been created. The image shown in Fig. 5 is segmented using Otsu thresholding.

The between-class variance value of the input image is 1843, while the threshold value of the image itself is 125.
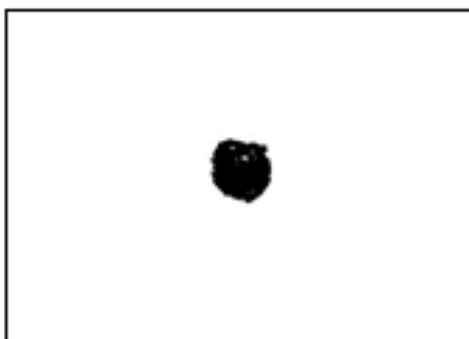
Fig. 5: Segmented legion

The color analysis of the segmented image using MATLAB identifies the presence of the known colors stated in the methodology section. Every color that has a percentage greater than 5% increases the TDS score by 1. Table 2 and Fig. 6 display the outcomes of the color analysis of the segmented image.

Table 2. Segmentation of dermoscopic images

	Color_Count	Percentage	Score
Black	858	41.131	1
White	0	0	0
Red	0	0	0
Light Brown	102	4.8897	0
Dark Brown	901	43.193	1
Blue Gray	0	0	0

Fig. 6: Estimated color score

5 Conclusion

In this study, we investigated the development of a skin lesion adaptive segmentation as well as the classification of skin lesions into malignant and benign cases using a set of extracted features.

The primary goal was the creation of a customized adaptive segmentation scheme for skin lesions. The research findings demonstrated that the revisited Otsu algorithm-based segmentation strategy performed very well for the selected database of clinical skin lesion images and successfully produced a promising minimum of 80% accuracy detection rate.

Undoubtedly, the proposed system suffers from several limiting constraints which are addressed in the next section.

6 Future Work

Since our presented system has its limitations, we propose the following recommendations for future work:
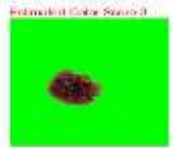
1. The proposed algorithm considered image characteristics enhancement as an important system phase, and it was mainly focused on brightness, contrast, and quality. However, for better results and image representation, it would be best to include air bubbles and hair detection as pre-processing sub-systems to produce a better reading of the image.

2. The algorithm presented a perfect malignancy detection rate, while for the benign cases, the algorithm failed to give an appropriate lesion for the area of interest. This is a main limitation of our study and deep learning-based algorithms need to be developed to address this shortfall.

3. All images were selected from a specific database obtained from the website dermis.net, and as a result, the generalizability of the results is considered limited. The database of images needs to be extended to include other open-source databases of clinical images.
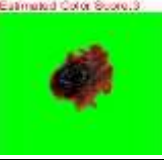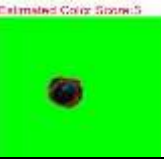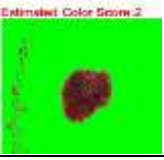
A more ambitious study would be to consider a more advanced stochastic model that captures the biological and epidemiological characteristics of the ultraviolet induction of skin cancer, [15]. Under such models, the number of cancerous cells follows a doubly stochastic marked Poisson point (DSMPP) process that is randomly distributed in time and space and whose estimated den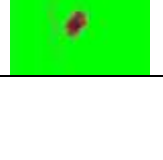sity (intensity or rate of the point process) gives an indication for the presence of skin cancer or its absence, [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26]. Unexpected spinoffs of classification and estimation algorithms employed in cellular communication systems, [27], and next-generation wireless systems, [28], [29], [30] can be made for skin cancer detection under the proposed DSMPP models.

7 Appendix A: Small Database of Segmented Dermoscopic Images

Table 3 summarizes the segmentation of 20 dermoscopic images with the results followed by an accuracy detection rate.

Table 3. Segmentation of selected dermoscopic images

Dermoscopic image	Segmentation	Result
		TN
		TP
		TN
		FN
		TN
		FP
		TP
		TP

	Estimated Color Score:3	TP
	Estimated Color Score:2	TN
	Estimated Color Score:3	TP
	Estimated Color Score:5	FN
	Estimated Color Score:2	TP
	Estimated Color Score:2	TP
	Estimated Color Score:3	TP
	Estimated Color Score:3	FP
	Estimated Color Score:2	TN

	Estimated Color Score:3	TP
	Estimated Color Score:2	TP
	Estimated Color Score:2	TP
	Estimated Color Score:2	TP

As shown in Table 3, we obtained a sample of 21 images with 12 TPs and 5 TNs, resulting in an accuracy detection rate of 80%.

References:

[1]	S. Jaiswar, M. Kadri, V. Gatty, "Skin cancer detection using digital image processing," *International Journal of Scientific Engineering and Research*, vol. 3 no 6, pp. 138-140, June 2015.

[2]	M. Chandrahasa, V. Vadigeri, D. Salecha, "Detection of skin cancer using image processing techniques*," International Journal of Modern Trends in Engineering and Research* (*IJMTER*), vol. 3, no. 5, pp. 111-114, May 2016.

[3]	M. A. Farooq, M. A. Azhar, "Automatic lesion detection system (ALDS) for skin cancer Classification Using SVM and Neural classifiers," *2016 IEEE 16th International Conference on Bioinformatics and Bioengineering*, pp. 301-308, 2016.

[4]	N. Razmjooy, S. Mousavi, et al, "A computer-aided diagnosis system for malignant melanomas," *Neural Computing & Applications*, vol. 23, no. 7/8, pp. 2059-2071, 2013. DOI: 10.1007/s00521-012-1149-1.

[5] D. Thanh, V. B. Prasath, L. Hieu, et al., "Melanoma skin cancer detection method based on adaptive principal curvature, colour normalisation and feature extraction with the ABCD rule," *Journal of Digital Imaging*, vol. 33, pp. 574–585, 2020.

[6] I. M. Widyantara, A. T. Kusuma, N. M. Wirastuti, "Preprocessing pada segmentasi citra paru-paru dan jantung menggunakan anisotropic diffusion *filter, Open Journal Systems-Teknologi Elektro*, vol. 14, no. 2, pp. 6-10, 2015.

[7] K. Korotkov, R. Garcia, "Computerized analysis of pigmented skin lesions: A review," *Artificial Intelligence in Medicine*, vol. 56, no. 2, pp. 69-90, 2012.

[8] I. Maglogiannis, C. N. Doukas, "Overview of advanced computer vision systems for skin lesions characterization," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 721-733, 2009.

[9] Z. She, Y. Liu, A. Damatoa, "Combination of features from skin pattern and abcd analysis for lesion classification," *Skin Research and Technology*, vol. 13, no. 1, pp. 25-33, 2007.

[10] N. R. Abbasi, H. M. Shaw, D. S. Rigel, R. J. Friedman, W. H. McCarthy, I. Osman, A. W. Kopf, D. Polsky, "Early diagnosis of cutaneous melanoma: Revisiting the abcd criteria," *Jama*, vol. 292, no. 22, pp. 2771-2776, 2004.

[11] Q. Abbas, I. Fondón, M. Rashid, "Unsupervised skin lesions border detection via two-dimensional image analysis," *Computer Methods and Programs in Biomedicine*, vol. 104, no. 3, pp. e1–e15, 2011.

[12] S. Lankton, A. Tannenbaum, "Localizing region-based active contours," *IEEE Transactions on Image Processing*, vol. 17. no. 11, pp. 2029-2039, 2008.

[13] L. Ma, B. Qin, W. Xu, L. Zhu, "Multi-scale descriptors for contour irregularity of skin lesion using wavelet decomposition," *2010 3rd International Conference on Biomedical Engineering and Informatics*, vol. 1, IEEE Press, pp. 414-418, 2010.

[14] S. Sabbaghi, M. Aldeen, R. Garnavi, G. Varigos, C. Doliantis, J. Nicolopoulos, "Automated colour identification in melanocytic lesions," *Engineering in Medicine and Biology Society* (*EMBC*), *2015 37th Annual International Conference of the IEEE*, IEEE Press, pp. 3021-3024, 2015.

[15] D. Pearl, "Stochastic model to explain the biology and epidemiology of the ultraviolet induction of skin cancer," *Mathematical Population Dynamics, Proceedings of the Second International Conference*, *Lecture Notes in Pure and Applied Mathematics*, vol. 131, CRC Press, 1991.

[16] J. S. Daba, M. R. Bell, "Segmentation of speckled images using a likelihood random field model," *Optical Engineering*, vol. 47, no. 1, 2008.

[17] J. Dubois, "Scattering statistics of doppler faded acoustic signals using speckle noise models," *IEEE International Conference on Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory* (*DIPED*), Lviv, Ukraine, pp. 185-189, 2003.

[18] J. S. Daba, M. R. Bell, "Estimation of the surface reflectivity of SAR images based on a marked Poisson point process model," *IEEE International Symposium on Signals, Systems, and Electronics* (*ISSSE*'95), San Francisco, CA, pp. 183-186, 1995.

[19] J. S. Daba, M. R. Bell, A. Abdi, S. Nader-Esfahani, "Comments on statistics of the scattering cross-section of a small number of random scatterers," *IEEE Transactions on Antennas and Propagation*, vol. 48, no. 5, pp. 844-845, 2000.

[20] J. P. Dubois, O. M. Abdul-Latif, "SVM-based detection of SAR images in partially developed speckle noise," IEC, Prague, pp. 321-325, 2005.

[21] J.P. Dubois, O. M. Abdul-Latif, "Detection of ultrasonic images in the presence of a random number of scatterers: A statistical learning approach," IEC, Prague, pp. 326-329, 2005.

[22] J. Dubois, "Poisson modulated stochastic model for partially developed multi-look speckle," *Proceedings of the American Conference on Applied Mathematics*, Harvard University, Cambridge, MA, USA, pp. 209-213, 2008.

[23] J. Dubois, "Segmentation of speckled ultrasound images based on a statistical model," *EURASIP Proceedings of the 16th International Biosignal Conference*, Czech Republic, vol. 16, 2002.

[24] J. S. Daba, M. R. Bell, "Statistics of the scattering cross-section of a collection of constant amplitude scatterers with random phase," *ECE Technical Reports*, Purdue University, p. 194, 1994.

[25] J. S. Daba, M. R. Bell, "Object discrimination and orientation determination in speckled images," *Optical Engineering*, vol. 33, no. 4, pp, 1287-1303, 1994.

[26] J. Dubois, "Traffic estimation in wireless networks using filtered doubly stochastic point processes," *Proceedings of IEEE International Conference on Electrical, Electronic, and Computer Engineering*, Cairo, Egypt, pp. 116-119, 2004.

[27] J. Dubois, J. S. Daba, M. Nader, C. El Ferkh, "GSM position tracking using a Kalman filter," *International Journal of Electrical, Computer and Communication Engineering*, vol. 6, no. 8, pp. 867-876, 2012.

[28] J. Dubois, O. M. Abdul-Latif, "Novel diversity combining in OFDM-based MIMO systems," *Proceedings of the American Conference on Applied Mathematics*, Harvard University, Cambridge, MA, USA, pp. 189-194, 2008.

[29] J. Dubois, O. Abdul-Latif, "Improved receiver diversity processing over SIMO fading channels," *Proceedings of the IEEE International Conference on Signal Processing and Communications*, 2007.

[30] O. M. Abdul-Latif, J. Dubois, "Performance of UWB system in a partially developed fading channel with CCI," *Proceedings of the 5th IEEE GCC Communication and Signal Processing Conference*, Kuwait, pp.1-5, 2009.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)
The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Conflict of Interest
The authors have no conflicts of interest to declare that are relevant to the content of this article.