Evaluation Methodology between Globalization and Localization Features Approaches for Skin Cancer Lesions Classification

H M Ahmed¹, R J Al-azawi², A A Abdulhameed³

1Department of Computer Science, Iraqi Commission for Computers and Informatics/Informatics Institute for postgraduate studies, Baghdad, Iraq.
2Department of Laser and Optoelectronics, University of Technology, Baghdad, Iraq.
3Department of Computer Science, University of Mustansiriyah, Baghdad, Iraq.

Hussein_gqi@yahoo.com

Abstract. Huge efforts have been put in the developing of diagnostic methods to skin cancer disease. In this paper, two different approaches have been addressed for detection the skin cancer in dermoscopy images. The first approach uses a global method that uses global features for classifying skin lesions, whereas the second approach uses a local method that uses local features for classifying skin lesions. The aim of this paper is selecting the best approach for skin lesion classification. The dataset has been used in this paper consist of 200 dermoscopy images from Pedro Hispano Hospital (PH2). The achieved results are; sensitivity about 96%, specificity about 100%, precision about 100%, and accuracy about 97% for globalization approach while, sensitivity about 100%, specificity about 100%, precision about 100%, and accuracy about 100% for Localization Approach, these results showed that the localization approach achieved acceptable accuracy and better than globalization approach for skin cancer lesions classification.

Keywords: Skin cancer, Dermoscopy, Melanoma, image segmentation, Color features, Texture features.

1. Introduction
Cancer is considered one of the largest danger to the humanity, it is predictable to be a major reason for death across the following few years. Depending on the statistics of the (WHO) World Health Organization 13% is the cancer estimation for all the deaths in 2004 in the world, and in 2030 is estimated 12 million people will die of the disease Among all the familiar cancers skin cancer is the main worry in both the developed and the developing countries during the last 40 years with an occurrence that has been increased in recent years [1]. One of the most common cancers in humans is skin cancer, the numeral of the skin tumor treatments growing robustly in the last decade and the cost
of skin tumor was most elevated of all tumors in U.S, around 87,110 new instances of melanoma and roughly 9,730 new melanoma-related deaths in 2017 in U.S [2]. In a modern estimate by the Australian government, the aggregate cost of diagnoses and treatments benign skin cancer was 511 million Australian dollars in 2010 and will be 703 million in 2015 [3]. The continuous rise of skin cancer in the worldwide and the rise medical cost and death average have prioritized to early diagnosis of this tumor [4]. The survival average is very high if cancer diagnosed in its early stages. Therefore, the early diagnosis the skin lesion is important to prevention death [5]. Two main cause that required an early detection of melanoma, firstly reason, localization of melanoma is surface (skin) in the majority of the cases, therefore, the detection manner can be very simple, second reason, the probability of healing of melanoma are large if diagnosed early stage (with the thickness below 1mm) with the 10-years survival rate for the patient as large as 95% [6].

2. Skin Cancer Diagnosis

There are different ways to evaluate and diagnose skin cancers. Most dermatologists depend on biopsy of the lesion for definitive diagnosis. Pathologists then examine histological sections derived from such biopsies to make a definitive diagnosis, this depends on evaluate cell morphology and architectural distribution of the cancer cells.

In some instances, the definite histopathological diagnosis of malignancy is difficult, this is particularly so when there is overlap in morphological features between some malignant and benign lesions [5]. Computer-based automatic diagnosis system appears to be an important tool for such difficult cases. Computer is not more intelligent than the human brain, but it may be able of extracting some information, such as texture features, that may not be easily seen by human eyes [7]. Therefore, it is important to develop efficient schemes for the physicians and pathologists through supporting their decision with computer-aided diagnosis (CAD) systems this is based on digital images. It is hoped that image analysis could help in identifying early skin cancers and aids in early diagnosis to help reduce the death rate caused by these serious diseases [8].

![Figure 1. Different types of dermoscopy images for both type benign and melanoma; (A) Benign skin lesion; (B) Melanoma skin lesion; (C) Benign skin lesion with some hair; (D) Melanoma skin lesion with some hair.](image-url)
3. Related work

There are a number of researches which used images obtained from a dataset of Hospital Pedro Hispano (PH2) as shown in Figure 1. [9] that is used in this paper, and these researches are as followed below:

[10] Suggested algorithm for diagnosis of the melanoma. Based on a group of directional filters and explore colors, directional and topology properties of the network it used 55 images from the dataset (PH2) and achieved result SP = 67% and an SE = 80%. [11] Suggested an improved system for automatic diagnosis of melanomas. It uses Adaptive Boosting (AdaBoost) classifier and used 57 images from (PH2) dataset and achieving SE = 78% and SP = 77%. [12] Suggested a system to diagnose melanoma. It used a bank of directional filters and a connected component analysis, then used AdaBoost classifier and used 200 images from (PH2) dataset, achieving an SE = 91.1% and an SP = 82.1%. [13] Suggested system to diagnoses of melanoma, using texture and color features. The achieved results Sensitivity = 94.1%, Specificity = 77.4% are obtained by mixing them both. Used 163 images from the (PH2) dataset. [14], a comparison of performance Big-of-feature classifies by using a manual and automated segmentation. They using KNN classifier and achieve results Sensitivity = 98% and Specificity = 86% for a manual segmentation, and achieved Sensitivity = 88% and Specificity = 82% for an automatic segment, and combining the two segmentation strategies achieve results SE is 98% and SP is 82%. They used 176 images from (PH2) dataset. [15] Suggested two different systems for the diagnosis of melanoma. The first system uses the global method and a second system use localization feature. They using AdaBoost, SVM, and KNN classifier. They used 176 images from (PH2) dataset and achieve results an SE is 96%, SP is 80% for globalization method and SE is 100%, SP is 75% for localization method.

4. Proposed Method

The proposed method tries to find the skin cancer using image processing by applying specific operation on the image. The hair removing is applied as a pre-processing step. The output of hair removing is used as globalization dataset. Two segmentation method applied using the different method used to find the skin cancer. The skin cancer is then classified using K-Means Clustering and Region of Interest Detection and Feature Selection.

![Figure 2. General view of proposed method.](image-url)
region of interest (ROI) then used as localization dataset. Feature extraction used to find some discriminant characteristics to verify the input image for classification. Extracted features are normalized for standardization features. Features selection used to reduce dimensional features and increase accuracy. The support vector machine used as a classification method. The general view of the proposed method shown in Figure.2. The total steps of the proposed method are five steps as follow:

4.1 Image Hair Removing
The process of hair removal from skin cancer images is a pre-processing of images in the proposed method for diagnosing. The cases through medical images consist of several steps for two kinds of skin image (benign and melanoma). Pre-processing of this kind of image consists of input a color image and extract its three bands RGB. For each band applied the same operation, first applied the median filter with size (3×3), then applied image Closed morphological operation with three masks zero, 45, and 90 degrees. These three results are a union of one output to replace in the input image. These will be used in three ways first as a global feature, second as input to segmentation with thresholding method, and the last as input with segmentation using k-mean clustering. The next step is applied Gaussian smoothing filter for smoothing a removed hair image at last. The size of the filter is 3×3 and the coefficient sigma is equal to 1.1. The total steps of hair removing illustrate in figure 3 in details, and the results are shown in figure 3.

![Figure 3. Skin Image Hair Removing (Image Preprocessing)](image)

- Image Segmentation
In the proposed method, two segmentation method applied and evaluated to get a better result for The ROI because it is sensitive step and the total feature depend on it, especially structural features depend on the extracted binary image. The first is segmentation using adaptive threshold. The second is
segmentation using K-means clustering. Each one has a specific way and properties to get the ROI. Among the implemented algorithms the Adaptive automatic thresholding technique accomplished the better results and confirm to be useful and firm enough to the automatically skin lesions segmentations in CAD systems. The result is shown in figure 5, And explained more details in [16].

Feature Extraction
Accurate classification depends on a discriminant feature that extracted from ROI. In proposed method several kinds of features are extracted as follow:

Geometric Features
These features depend on the binary mask that deals with the shape of ones in the mask. The binary mask extracted from skin image using two previous segmentation method, from this binary mask geometric feature will extract as follow: Area, Perimeter, distance to Centroid, distance to Centre of Nearest, and standard deviation for each previous features [17].

\[
\text{Area} = \sum_{i=1}^{m} \sum_{j=1}^{n} \text{I}_{\text{white}}(i,j)
\]  

Where m and n is height and width of binary image (object mask) and I\text{white} is white pixel in binary image (object).

\[
P = N_{\text{e}} + \sqrt{2} N_{\text{o}}
\]  

Where Ne is the number of even values and No is the number of odd values in the boundary chain code.

\[
D(B,C) = \sqrt{(B_x - C_x)^2 - (B_y - C_y)^2}
\]  

Where Bx, By are coordinate of boundary points and Cx and Cy are coordinate of centroid point of object.

\[
D(B,C) = \sqrt{(B_x - C_{\text{row}})^2 - (B_y - C_{\text{col}})^2}
\]  

Where Bx, By are coordinate of boundary points and C_row and C_col are coordinate of centroid point of geometric center.

\[
St = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2}
\]  

Where St is standard deviation value, n is size of column data, xi is column data, and μ is mean of column data.

Intensity Color Features
The Intensity color features used in the proposed method, for global and local features, used the intensities of colors band after split them into RGB from this mask color feature will extract as follow: Mean of colors for each band, the variance of Colors, and standard deviation for each previous features [17].

\[
\mu = \frac{1}{n} \sum_{i=1}^{n} x_i
\]  

Where μ is mean value, n is size of column data, and xi is column data.

\[
\text{Var} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)^2
\]  

Where Var is variance value, n is size of column data, and xi is column data.
Where $Var$ is variance of column data, $n$ is size of column data, $x_i$ is column data, and is mean of column data.

- **Texture Features**
  The third group features are texture features, they depend on second order feature (Gray Level Co-occurrence Matrix (GLCM) and Gray-Level Run-Length Matrix (GLRLM)). These features are local feature depend on the relation of two adjacent pixels that give some information about location, the samples of grayscale image matrix. All features obtained from co-occurrence matrix for GLCM while GLRLM depend on run length matrix. These main features are as follow [17]: contrast, correlation, energy, and homogeneity for GLCM, while (Short Run Emphasis, Long Run Emphasis, Gray Level Distribution, Run Length Distribution, Run Percentage, Low Gray-level Run Emphasis, High Gray-level Run Emphasis, Short Run Low Gray-level Emphasis, Short Run High Gray-level Emphasis, Long Run Gray-level Emphasis, Long Run High Gray-level Emphasis) for GLRLM.

- **Features selection**
  Features Selection step used to find feature subsets combination that relevant to result. The feature selection restricted chance to find class labels of a given features as filter. Conditional mutual information is modeled to use. This corresponds to the smallest feature subset and starting from this objective function to increased subset using mutual information properties.

- **Classification**
  The classification is the last step in this proposed method that used 200 dermoscopy images of lesions for training and testing, where the SVM algorithm is used in this proposed method for diagnoses. Two approaches global and local approach are considered in this proposed method, the global approach depends on the global features while local approach depends on local features from ROI. Two kind of kernel function used for classifying data: linear, and nonlinear (polynomial kernel) function. Where the linear kernel function is used for solving the optimization problem and it is much faster than the others. The linear kernel is certainly not more accurate than the others but it is used when a number of features is larger than number of observations (labeled features). While nonlinear function (Polynomial kernel) is used to solving the optimization problem, and it is used usually with high dimension features as kernel trick to make the combinations of features able to split as possible. Kernel function used for solving interference in properties and making features that representing the infected region more separable and increasing accuracy in classification results.

5. **Experimental Result**
Experimental of the proposed method used Hospital Pedro Hispano (PH²) dataset. It consists of 200 dermoscopy images of lesions, Hair Removing this step is a pre-processed step. It applied to two kinds of images benign and melanoma. The same procedure applied to them. The operations performed on the images for the purpose of hair removal are done by separating the colors to the RGB band then applied the median filter on each band. Morphological closing operation applied on an image with line structure element. The horizontal line (zero angles), sloped line (45 angles), vertical line (90 angles). From these
three band getting maximum value replaced the hair area that near to dark. The last step is applying Gaussian smoothing filter on the result. The hair removing of skin image is shown in figure 4.

![Hair Removal Process](image)

**Figure 4.** Hair Removal Process: (A) Original Haired Skin Image Cancer(B) Haired Skin Cancer Mask (C) Hair Removing Skin Cancer Image (D) Smoothed Skin Cancer Image.

The second step is Segmentation, two algorithms for automatic segmentation of dermoscopy image were executed and assessed for tracking the boundary of skin cancer lesions, including the Adaptive automatic thresholding, K-Means clustering. The result of each algorithm is greatly influenced by the type of images used for analysis. Among the implemented algorithms the Adaptive automatic thresholding

![Implemented Segmentation Methods](image)

**Figure 5.** Implemented Segmentation Methods; (A) Original Skin Cancer Images; (B) Ground Truth Lesion Image; (C) Adaptive Automatic Thresholding Cluster Lesions Image; (D) K-Means Cluster Lesion Image.
technique accomplished the better results and confirm to be useful and firm enough to the automatically skin lesions segmentations in CAD systems. The skin image Segmentation step in proposed method and the evaluation of the two methods used to Region of Interest Detection are explained in details in [16]. Some of these result is shown in figure 5.

In the evaluation step to Region of Interest Detection, the Adaptive automatic thresholding technique was chosen because it accomplished the better results, after this operation is done the enhanced of segmentation is applied on the result by comparing it with ground truth (GT) lesion mask of the same image get it from a dataset. The comparison applied on each pixel if it is one on proposed segmentation or in GT image the result will be one. Some of these results are shown in Figure 6 and Figure 7.

![Flowchart Enhanced segmentation methods](image)

**Figure 6.** Flowchart Enhanced segmentation methods
The region of interest circularity is calculated after the enhanced operation to determinate cancer. Color ROI important because it used in next step of feature extraction (color features) that detected by select the whole the boundary of the lesion and cover the lesion in a rectangle that shown in figure 8.

Feature extraction applied to find three kinds of features geometric features, color features, and texture feature. Used equation (1 to 7) to calculate geometric and color features and other feature are second

Figure 7. Enhanced segmentation methods for skin cancer image.

Figure 8. Calculate the color ROI and ROI circularity.
order (GLCM and GLRLM). There are two types of features are produced the global and local features. The global features depend on the whole images while local features depend on ROI lesion. Feature normalization depends on mean and standard deviation for each feature. The normalization of features supports classification step. This necessary for keeping the range of feature close to all in numeric value. The last step before the classification step is feature selection step that used Mutual information method in these step calculate by weighted each feature vector to the class label. These weights will sort from high weight to low weight and then can choose the higher weighted feature that relative to specify the class label. These operations at all reduced complexity of the proposed method. The classification is the last step in the proposed method, it is used SVM algorithm with two functions (linear and nonlinear (Polynomial kernel)) for diagnoses. The achieved results for each function are shown in Table 1, 2.

| Table 1: Testing Linear and polynomial kernel (Globalization) |
|-------------------------------------------------------------|
| **Kernels** | **Sensitivity** | **Specificity** | **Precision** | **Accuracy** | **Time** |
|------------|----------------|----------------|--------------|-------------|----------|
| Linear     | 100%           | 77%            | 93%          | 94%         | 0.019923 sec |
| Polynomial | 96%            | 100%           | 100%         | 97%         | 0.021653 sec |

| Table 2: Testing Linear and polynomial kernel (Localization) |
|-------------------------------------------------------------|
| **Local Approach** | **Kernels** | **Sensitivity** | **Specificity** | **Precision** | **Accuracy** | **Time** |
|---------------------|------------|----------------|----------------|--------------|-------------|----------|
| Linear              | 100%       | 85%            | 95%            | 96%          | 0.019869 sec |
| Polynomial          | 100%       | 100%           | 100%           | 100%         | 0.021965 sec |

6. Conclusion
Evaluation and comparison methodology between two approaches are presented for the diagnosis skin cancer image. There are two approaches depend on localization and globalization features. Both approaches are tested on PH2 dataset, the experimental result show that localization features approach performance better than globalization features approach, as shown in Table 2 for each kernel linear and polynomial with respect to time. The selected kernel (polynomial kernel) achieved best results: sensitivity is 100%, specificity is 100%, Precision is 100%, and Accuracy is 100%. In other approach (globalization features) the achieve results are shown in Table 1 for each kernel linear and polynomial. The polynomial kernel achieved best results: sensitivity is 96%, specificity is 100%, Precision is 100%, and Accuracy 97%.

References
[1] H. W. Rogers and B. M. Coldiron, “Analysis of skin cancer treatment and costs in the United States Medicare population, 1996-2008,” Dermatol. Surg., vol. 39, no. 1 Pt 1, pp. 35–42, 2013.
[2] Siegel, R.L., Miller, K.D., Fedewa, S.A., Ahnen, D.J., Meester, R.G., Barzi, A. and Jemal, A., 2017. Colorectal cancer statistics, 2017. CA: a cancer journal for clinicians, 67(3), pp.177-193.
[3] Sinclair, Rodney. "Nonmelanoma skin cancer in Australia." British Journal of Dermatology 168.1 (2013): 1-2.
[4] Bhati, P. and Singhal, M., 2015, November. Early stage detection and classification of melanoma. In Communication, Control and Intelligent Systems (CCIS), 2015 (pp. 181-185). IEEE.
[5] Demir, Cigdem, and Bülent Yener. "Automated cancer diagnosis based on histopathological images: a systematic survey." Rensselaer Polytechnic Institute, Tech. Rep (2005).

[6] Sagar, Chiranjeev, and Lalit Mohan Saini. "Color channel based segmentation of skin lesion from clinical images for the detection of melanoma." In Power Electronics, Intelligent Control and Energy Systems (ICPEICES), IEEE International Conference on, pp. 1-5. IEEE, 2016.

[7] Ruela, Margarida, Catarina Barata, and Jorge S. Marques. "What is the role of color symmetry in the detection of melanomas?" In International symposium on visual computing, pp. 1-10. Springer, Berlin, Heidelberg, 2013.

[8] Sheha, Mariam A., Mai S. Mabrouk, and Amr Sharawy. "Automatic detection of melanoma skin cancer using texture analysis." International Journal of Computer Applications 42, no. 20 (2012): 22-26.

[9] Mendonça, Teresa, Pedro M. Ferreira, Jorge S. Marques, André RS Marcal, and Jorge Rozeira. "PH 2-A dermoscopic image database for research and benchmarking." In Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, pp. 5437-5440. IEEE, 2013.

[10] Barata, Catarina, Jorge S. Marques, and Jorge Rozeira. "Detecting the pigment network in dermoscopy images: A directional approach." In Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, pp. 5120-5123. IEEE, 2011.

[11] C. Barata, J. S. Marques, and J. Rozeira, "A system for the automatic detection of pigment network," Proc. - Int. Symp. Biomed. Imaging, pp. 1651–1654, 2012.

[12] Barata C, Marques JS, Rozeira J. A system for the detection of pigment network in dermoscopy images using directional filters. IEEE transactions on biomedical engineering. 2012 Oct;59(10):2744-54.

[13] Marques, Jorge S., Catarina Barata, and Teresa Mendonça. "On the role of texture and color in the classification of dermoscopy images." Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE. IEEE, 2012.

[14] Barata, Catarina, Jorge S. Marques, and M. Emre Celebi. "Towards an automatic bag-of-features model for the classification of dermoscopy images: The influence of segmentation." In Image and Signal Processing and Analysis (ISPA), 2013 8th International Symposium on, pp. 274-279. IEEE, 2013.

[15] Barata, Catarina, Margarida Ruela, Mariana Francisco, Teresa Mendonça, and Jorge S. Marques. "Two systems for the detection of melanomas in dermoscopy images using texture and color features." IEEE Systems Journal 8, no. 3 (2014): 965-979.

[16] Al-azawi, Razi J., Abbas Abdulazez Abdulhameed, and Hussein Majeed Ahmed. "A Robustness Segmentation Approach for Skin Cancer Image Detection Based on an Adaptive Automatic Thresholding Technique." American Journal of Intelligent Systems 7, no. 4 (2017): 107-112.

[17] Filipczuk, Pawel, Thomas Fevens, Adam Krzyżak, and Andrzej Obuchowicz. "GLCM and GLRLM based texture features for computer-aided breast cancer diagnosis." Journal of Medical Informatics & Technologies 19 (2012).