Improved Skin Lesion Detection and Segmentation by Fusing Texture and Geometric Features

Nidhi Bansal[1,*], S. Sridhar[2] , P.L. Daisy Priya[3]

[1] Research Scholar, Department of IST, CEG, Anna University, Chennai, India.

[2] Professor, Department of IST, CEG, Anna University, Chennai, India.

[3]M.E. Student, Department of IST, CEG, Anna University, Chennai, India.

Abstract

Melanoma is the greatest carcinogenic skin cancer. In the last years, the prevalence degree of melanoma has risen by 50 percent. There is a necessity to provide an onscreen system for the diagnosis of skin lesions. The system will reduce the unnecessary biopsy and the cancer can be diagnosed at an early stage. In this paper a framework is proposed for the automated skin lesion detection in an input image. A segmentation algorithm based on texture is used to classify normal skin class or lesion class. Also, fusion of texture and geometric features is presented in this work. SVM classifier is trained to identify lesions as malignant melanoma or benign lesion. The system yielded an efficiency of 84.7%, 89.4% and 83.5% for Haralick features, features given by Soh and Clausi and Histogram based features respectively. The fusion based on texture and geometric features enhanced the performance of the system. The evaluated performance metrics are better than the previous methods. The improved system helps diagnose at an early stage reducing the mortality rate.

Keywords: Melanoma, texture features, geometric features, segmentation, classification

1   INTRODUCTION

Melanoma is a cancer that is caused by the cell that produces melanin found under the skin [1]. The body parts affected by melanoma are the trunk, head, and neck. Melanomas can be caused by revelation to sunlight. The risk of melanoma seems is increasing in people under 50, especially women. It is "the 19th most" communal cancer across the globe [2]. Around 300,000, new cases of the disease were found in 2018 [2]. 75% of the deaths associated with skin cancer are caused by melanoma [3]. According to a report in [4], the number of serious melanoma cases in the past 11 years has increased by 56 percent. There is an estimation of around 2 million melanoma cases in 2020 [4]. If melanoma is detected early in Stage I, it can be cured but becomes dangerous in Stage IV.

Also, the treatment is costlier for Stage IV melanoma. With the rising prevalence rate of cancer, early melanoma screening is beneficial. Several new diagnostic tools such as Photograph of Body parts, Handheld Dermoscopy, Sequential Dermoscopic follow-up, etc [5] have developed in the recent year. Expensive surgical treatments can be evaded if cancer is detected at an initial stage. Dermatologists use a tool known as the Dermatoscope that detects the extremity of melanoma using images acquired [6]-[8] by magnifying, illuminating and enhancing skin lesions. However, it is tough to segment the accurate boundary of lesion especially with blurred or unclear edges.

Most of the segmentation algorithms use pixel color, intensity or texture to assign a label. To locate the lesions accurately in the skin image, the method used is to incorporate the textural and geometric information, as there is a huge difference between these properties of normal skin and lesion. Geometric features are features of objects. It can be corners, edges, contours, and salient regions. Geometric features are a great success in many imaging applications. The features of lesion are extracted and segmented to find precise border. Therefore, in this work the objective is to incorporate both textural and geometric information for more subtle boundary segmentation.

In this paper, a novel "textural segmentation algorithm (TDLS)" [9] is used to pinpoint the skin lesions in digital images. From the segmentation result significant features which include, the texture features describing the different patterns of pigmented networks, and the geometrical features based on the "ABCD scale (asymmetry, border irregularity, color irregularity, and diameter)" [10] used in clinical diagnosis are extracted. After analyzing each feature, the segmented region is analyzed and classified as malignant or benign. This way automated diagnosis of melanoma reduces the time complexity and increases the classification accuracy. The overall steps involved in the automated detection of skin lesion are shown in Fig.1. In Section 2, texture distinctiveness, region classification and feature extraction is described. Experimental

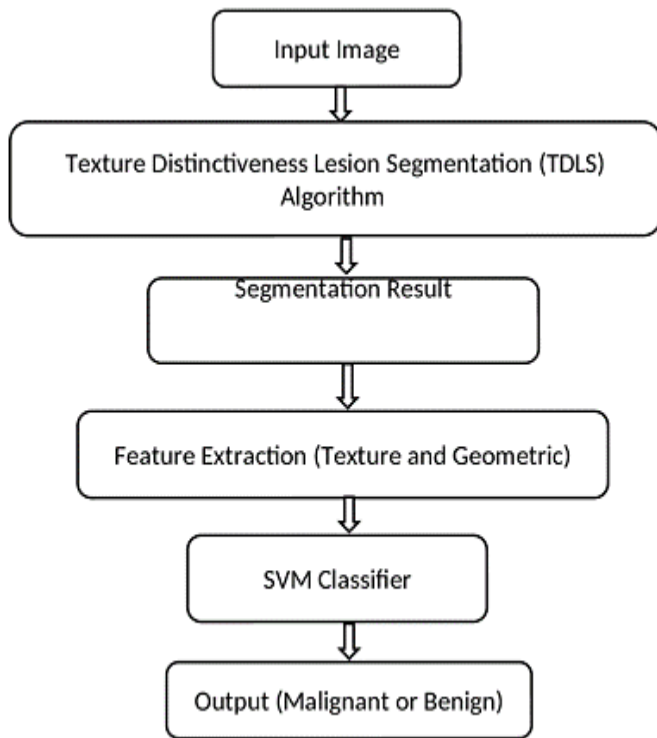results are shown in Section 3 and the work is concluded in Section 4.

Fig. 1. Overall Steps Involved in Automated Detection of Skin Lesion

2 TEXTURE DISTINCTIVENESS LESION SEGMENTATION ALGORITHM (TDLS)

2.1 Sparse Texture Distributions

Sparse texture representations divide the image in sparse regions, it can be texture patches. The model learns the lesion texture after the region detection. The problem of learning can be a clustering or an optimization problem. The sparse texture distributions can be determined by acquiring a texture vector for each pixel. Each pixel has a given no of channels. The pixel location is denoted by s and the patch centered at pixels is represented by a vector t_s given in Eqn(1). All texture vectors are consolidated to obtain T, which is a set of N x M texture vectors [10].

$$T=\{ts_j \,|1 \leq j \leq N \times M\} \tag{1}$$

2.1.1 Clustering

A clustering algorithm is used to calculate a TD metric that differentiates two texture distributions. There are two important steps:

Step1: K-means clustering technique is run to find K clusters of texture data points and categorize them according to similar properties. The optimization function is given in Eqn. (2),

where "C_k is the kth set of texture vectors, and μ_k is the mean vector for the corresponding set"[10].

The clustering is trailed by learning a finite mixture model. The finite mixture model is required as K-means clustering does not consider any probabilistic information. "The model parameters in the set Θ maximize the log-likelihood function shown in Eqn. (3)" [10].

$$C^k = \arg \min\sum_{k=1}^{K} \sum_{tsj\in ck} |t_{sj} - uk|2 \tag{2}$$

Step 2: In Step 2, a Gaussian distribution is applied to all clusters. The model parameters used are the mean (μ), covariance (Σ) and α is the mixing proportion.

$$\Theta = \arg \max \sum_{j=1}^{n} \sum_{k=1}^{K} \log(\alpha_k \, \rho(\,|\mu_k \sum k\,|)) \tag{3}$$

where $\sum_{k=1}^{K} \alpha_k = 1$ and $\Theta= \{\mu_1, \mu_2,\ldots\ldots\mu_k,\ \sum_1, \sum_2,\ldots\ldots \sum_k, \alpha_1, \alpha_2,\ldots\ldots\alpha_k\}$

The expectation-maximization algorithm is used which is an iterative algorithm. The outputs obtained by k-means clustering are used to obtain the initial parameters for the Gaussian model.

$$\mu_k = \mu c_k \tag{4}$$

$$\sum_k = \sum c_k \tag{5}$$

$$\alpha_k = \rho(t_{sj} \in c_k) \tag{6}$$

To maximize the weighted probability $t_{sj}|\mu_k, \sum k\,|$ each texture vector is allotted to belong to the distribution. Fig. 2 shows the result of representative Texture Distributions.

2.2 Region Classification

The next significant process is to discover the regions which depends on the texture distributions. The regions in a lesion are categorized based on similar pixels and the allied TD metric is also considered for classification. The output image gets divided into a huge number of regions. Then, each region is separately classified as belonging to normal skin or lesion, based on texture features.

2.2.1 Initial Regions

The image is partitioned into many regions. The initial phase is to reduce noise in the image by doing over segmentation. A fast classification algorithm can be used to find which regions belong to the lesion class. From the "statistical region merging (SRM) algorithm" in [11], the initial over segmentation algorithm is adapted. The advantage of using the SRM algorithm is that it considers the pixel location and is computationally efficient.
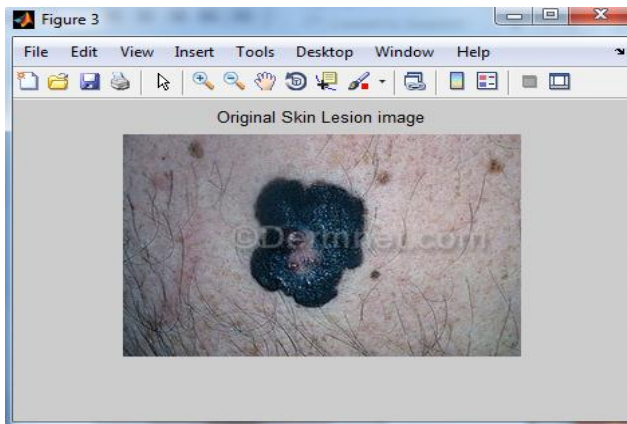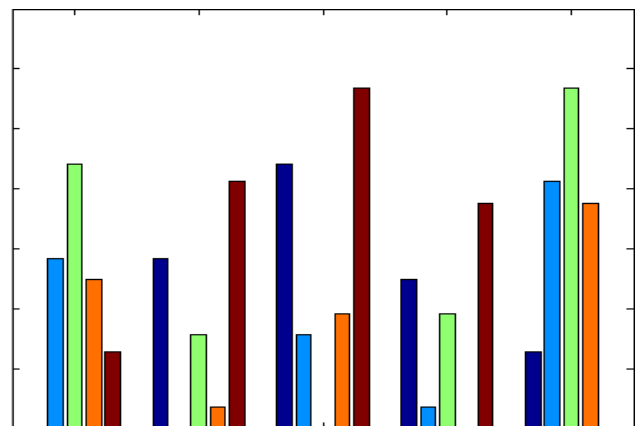
(a)
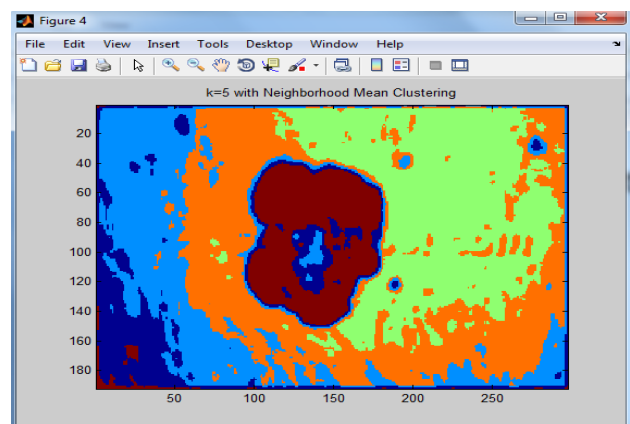
(b)
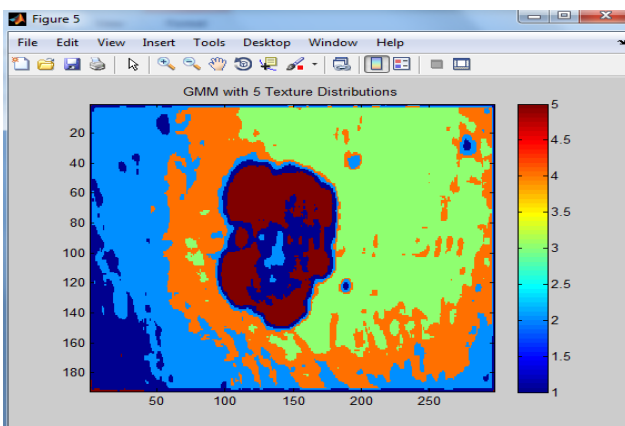
(c)

(d)

Fig. 2. Result of representative Texture Distributions. In (a) the original image is shown. In (b) the lesion is represented by texture distributions. In (c) depending on texture distribution each pixel is replaced by a colour. In (d) the texture distinctiveness among different texture distributions is shown.

2.2.2 Distinctiveness Based Segment Classification

Once the initial step of segmentation is completed, each region is partitioned into a few classes. The two classes considered are the normal skin class or lesion class. The classification step is illustrated in eq. (7), it is a skin lesion if y is 1 otherwise it signifies normal if y is 0.

$$y(R) = \begin{Bmatrix} 1 \; if \; D_R \; \geq \tau \; (Lesion) \\ 0 \qquad\qquad otherwise \end{Bmatrix} \qquad (7)$$

For each texture distribution a TD metric is calculated and used to determine a regional TD metric, D_R. The average TD is denoted by D_R, across region R. The probability of a pixel being associated with the jth texture distribution is given by $P\,(T_j^{\,r}|R)$.

$$D_R \;=\; \sum_{J=1}^{K} D_j \; P\,(T_j^{\,r}|R) \qquad (8)$$

Otsu's threshold [10] τ is used "to divide the set of texture distributions" into normal skin and the lesion class. The total

intra class variance of the TD metric is minimized as:

$$\tau = \text{argmin}\,(\sigma_{c1(\tau)}^2\,\rho(T_{c1(\tau)}^{\,r}) + \sigma_{c2(\tau)}^2\,\rho(T_{c2(\tau)}^{\,r})) \qquad (9)$$

The classes depend directly on τ. The texture distribution of skin is in class C1 (τ) if the associated distinctiveness metric is above τ otherwise in class C2 (τ).

2.3 Feature Extraction and Classification

Feature extraction is applied to excerpt the features of a melanoma lesion. It includes features that can be visually recognized by dermatologists. The automated system covers the core features of a melanoma viz texture asymmetry, color of the lesion, lesion boundary, and structural patterns.

2.3.1 Texture Features

Various features are extracted from R, G, and B channels of the image. a) Eight Haralick statistical features selected are

1118

Contrast, Correlation, Sum of Average, Sum of Variance, Sum of Entropy, Difference of Variance, Difference of Entropy and Energy, [12] b) the features like: *Cluster Prominence, Dissimilarity, Entropy, Homogeneity, Cluster Shade and Maximum Probability*, are considered [12] and c) the features suggested by Clausi are *Inverse difference and Inverse difference moment* [12] and d) four histogram based texture features are also taken for classification. It includes *mean, variance from mean, moment calculation based on skewness and kurtosis*[12]. These features include a total of 60 (20 x 3) extracted from R G B channels describes various patterns of pigmented network found in skin lesions.

2.3.2   Geometric Features

Total of 9 features were extracted from the lesion's binary mask. The various significant features computed are used to classify the malignancy of lesions. These geometric features are given below:

Area (A): Number of pixels of the lesion.

Perimeter (P): The entire number of pixels along the border of the lesion.

Greatest diameter (GD): A line between the two furthest border points and pass through the centroid given by,

$$(xc, yc) = \left(\frac{\sum_{i=1}^{n} xi}{n} \frac{\sum_{i=1}^{n} yi}{n}\right) \tag{10}$$

Shortest diameter (SD): A line between the two nearest border points.

Circularity index (CI): Indicates uniformity in shape.

$$CI = \frac{4Api}{p2} \tag{11}$$

Irregularity index A (IrA):

$$IrA = \frac{P}{A} \tag{12}$$

Irregularity index B (IrB):

$$IrB = \frac{P}{GD} \tag{13}$$

Irregularity index C (IrC):

$$IrC = Px\left(\frac{1}{SD} - \frac{1}{GD}\right) \tag{14}$$

Irregularity index D (IrD):

$$IrD = GD\text{-}SD \tag{15}$$

2.3.3   Classification

Classification is the next major phase to classify the segmented regions. It uses the set of features extracted. The training set consists of the model trained and the test set is to evaluate the classification accuracy. Machine learning approaches show good results for lesion detection from dermoscopic images [14]. Several algorithms can be used for classification but Support Vector Machine (SVM) so far gives a good performance. The SVM [15] is the most used classifier in literature for lesion skin classification. SVM is an example of Supervised Learning, it looks for finding the optimal hyperplane that separates two classes. However, hyperplane with maximal margin is a good choice.

3   EXPERIMENTAL RESULTS

Around 300 images (115 benign and 185 malignant) were taken from www.dermnet.com [16]. The training images are 80% of the set and used to train the SVM classifier. Classification results for various texture features and, texture features combined with geometric features are compared and the results are shown in Fig. 3 (a) and Fig. 3 (b).
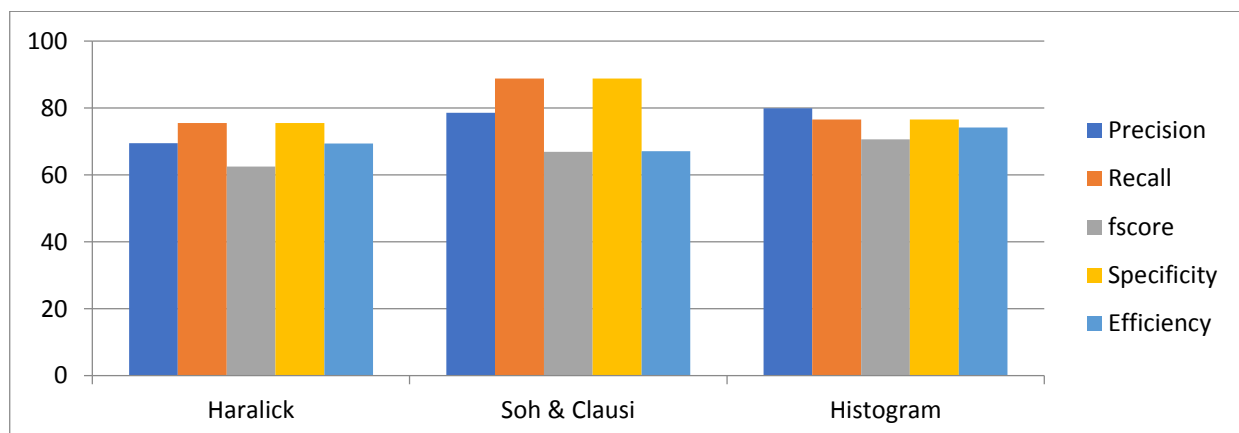

Fig. 3 (a). Comparison of Classifier Performance with various Texture Features
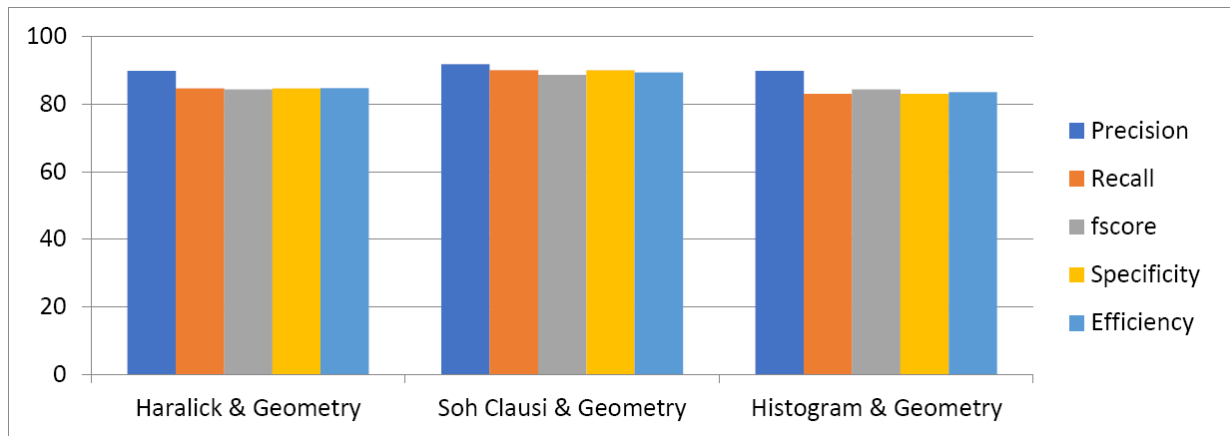
Fig. 3 (b). Comparison of Classifier Performance with Fused Texture and Geometric Features

A set of 69 features describing the geometrical (9) and textual properties (60) of the lesions are captured which is used to train the SVM classifier. The performance of the classifier (SVM) is evaluated using a confusion matrix in which the actual and predicted values are calculated. From the confusion matrix different metrics viz *Precision(P), Recall(R), Sensitivity(SN), Specificity(SP) and Efficiency(EFF)*, evaluating the performance of the classifier are computed.

True Positive (TP) and True Negative (TN) are the properties of positive and negative cases that are correctly classified by the classifier. False positive (FP) is the proportion of the negative cases; not having symptoms of skin cancer, that are inaccurately classified by the classifier. False negative (FN) is the proportion of the positive cases; those having cancer, that are wrongly classified as negative. Various texture features that include, the Haralick features, features suggested by Soh and Clausi and Histogram based texture features were trained separately.

The results of classification accuracy are shown in Table 1 and Table 2. Table 2 depicts that the classification results are better when the classifier is trained by fusing texture and geometric features. Table 2 shows the classification efficiency computed by Sou and Clausi is 5-6 % better than the Haralick and Histogram-based features. This is because the feature suggested by Soh and Clausi includes features that describe more about the symmetric/asymmetric distributions of the pigmented network. Therefore, the skin lesion images are more accurately classified with the fusion of texture and geometric features.

Table 1. Classification Accuracy for various Texture Features

Features	P	R	F-Score	SP	EFF
Haralick	69.5%	75.5%	62.5%	75.5%	69.41%
Soh and Clausi	78.6%	88.8%	66.9%	88.8%	67.1%
Histogram	**79.9%**	**76.5%**	**70.6%**	**76.5%**	**74.11%**

Table 2. Classification Accuracy of Texture Features Fused with Geometrical Features

Features	P	R	SP	SN	EFF
Haralick	89.8%	84.6%	84.4%	84.6%	84.7%
Soh and Clausi	**91.8%**	**90%**	**88.6%**	**90%**	**89.4%**
Histogram	89.8%	83%	84.4%	83%	83.5%

4 CONCLUSION AND FUTURE WORK

The objective of this paper is to develop a framework to segment skin lesions images taken using a digital camera and to classify and detect melanoma lesions. To do this, a novel texture distinctiveness lesion segmentation algorithm is implemented and significant features describing the malignancy and benign nature of the lesions are extracted. The fusion of two features i.e. texture and geometric, improved the performance of the system. In the recent emerging field called tele dermatology this framework can be used as a diagnostic tool by the dermatologist, which could serve as a procedure for the initial assessment of patients at the risk of being affected by melanoma. Images with a fair amount of hair on the skin are not considered, in future this work can be extended for large datasets. Also, deep learning methods can be incorporated to improve the overall efficiency and time complexity of the classification result.

REFERENCES

[1] American Cancer Society, "Melanoma skin cancer," https://www.cancer.org/cancer/melanoma-skin-cancer/about/what-is-melanoma.html

[2] World Cancer Research Fund, https://www.wcrf.org/dietandcancer/cancer-trends/skin-cancer-statistics

[3]   A. F. Jerant, J.T. Johnson, C.D. Sheridan, & T.J. Caffrey, "Early detection and treatment of skin cancer," American family physician,vol. 62, no. 2, pp. 1-6, July 2000.

[4]   *Cancer Facts and Figures 2019*. American Cancer Society. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2019.html. Accessed January 14, 2019.

[5]   H.Tsao, et al.. "An estimate of the annual direct cost of treating cutaneous melanoma," Journal of the American Academy of Dermatology, vol. 38, no. 5, pp. 669-680, May 1998.

[6]   Elisabeth M Wurm and H Peter Soyer, Dermatology Research centre, the University of queensland, School of Medicine, Brisbane, "http://www.australianprescriber. com/magazine/33/5/150/5.

[7]   H. Ganster, A. Pinz, R. Rohrer, E. Widling. M. Binder, and H. Kittler "Automated Melanoma Recognition," IEEE Transactions on Medical Imaging, vol. 20, no. 3, pp. 233–259, Mar. 2001.

[8]   B. Erkol, R. H. Moss, R. Joe Stanley, W. V. Stoecker, and E. Hvatum, "Automatic Lesion Boundary Detection in Dermoscopy Images using Gradient Vector Flow Snakes," Skin Research Technology, vol. 11, no. 1, pp. 17–26, 2005.

[9]   C. Serrano and B. Acha. "Pattern analysis of dermoscopic images based on Markov random fields." Pattern Recognition, vol. 42, no. 6,pp. 1052-1057, 2009.

[10]   Glaister, Jeffrey, Alexander Wong, and David A. Clausi. "Segmentation of skin lesions from digital images using joint statistical texture distinctiveness", IEEE transactions on biomedical engineering, vol. 61, no. 4, pp. 1220-1230, 2014.

[11]   C. Scharfenberger,et al. "Statistical Textural Distinctiveness for Salient Region Detection in Natural Images," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2013, pp. 979–986

[12]   R. Nock and F. Nielsen, "Statistical Region Merging," IEEE Transactions on Pattern Analysis and Machine Intelligence., vol. 26, no. 11, pp. 1452–1458, Nov 2004.

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

[14]   A, R. Ali, et al. "A machine learning approach to automatic detection of irregularity in skin lesion border using dermoscopic images," PeerJ Computer Science, 6, e268, 2020.

[15]   Y. Filali, A. Ennouni, and M. A. Sabri, "Multiscale approach for skin lesion analysis and classification," International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). DOI: 10.1109/ATSIP.2017.8075545. 22-24 May 2017. Fez, Morocco.

[16]   Derment, (2014). [Online]. Available : http://www.dermnet.com