2017 18th International Conference on Parallel and Distributed Computing, Applications and Technologies

Segmentation and Classification of Skin Cancer Melanoma from Skin Lesion Images

Nay Chi Lynn
Student: Department of Research & Development
University of Computer Studies, Mandalay
Mandalay, Myanmar.
naychelynn@gmail.com

Zin Mar Kyu
Student: Department of Research & Development
University of Computer Studies, Mandalay
Mandalay, Myanmar
zinmarkyu.pp@gmail.com

Abstract— **Melanoma, one type of skin cancer is considered o the most dangerous form of skin cancer occurred in humans. However it is curable if the person detects early. To minimize the diagnostic error caused by the complexity of visual interpretation and subjectivity, it is important to develop a technology for computerized image analysis. This paper presents a methodological approach for the classification of pigmented skin lesions in dermoscopic images. Firstly, the image of the skin to remove unwanted hair and noise, and then the segmentation process is performed to extract the affected area. For detecting the melanoma skin cancer, the meanshift algorithm that segments the lesion from the entire image is used in this study. Feature extraction is then performed by underlying ABCD dermatology rules. After extracting the features from the lesion, feature selection algorithm has been used to get optimized features in order to feed for classification stage. Those selected optimized features are classified using kNN, decision tree and SVM classifiers. The performance of the system was tested and compare those accuracies and get promising results.**

Keywords- Melanoma; Skin Cancer; Segmentation; Classification;

I. INTRODUCTION

Since the skin is the largest body of the body, it saves them from various infectious diseases, sunlight and many other problems and maintains body temperature. Melanoma is the most dangerous form of skin cancer. These cancers are caused by undesirable damage to DNA by skin cells (usually caused by ultraviolet light from the sun or solarium), when skin cells multiply rapidly, causing mutations (genetic defects), forming malignant tumors [15]. Malignant melanoma is one of the fastest growing cancers in the world. Melanoma is mainly caused by intense and temporary exposure to ultraviolet radiation (often resulting in sunburn) in people who are genetically prone to this disease.

As skin is the body's largest organ, it saves us from various infections, sunlight and from many other problems and helps us to maintain body temperature. Melanoma is the most dangerous form of skin cancer. These cancerous growths develop when unrepaired DNA damage to skin cells (most often caused by ultraviolet radiation from sunshine or tanning beds) triggers mutations (genetic defects) that lead the skin cells to multiply rapidly and form malignant tumors. Malignant melanoma is one of the most rapidly increasing cancers in the world. Melanoma is caused mainly by intense, occasional UV exposure (frequently leading to sunburn), especially in those who are genetically predisposed to the disease. Melanoma kills an estimated 10,130 people in the US annually. The key is early detection and the key to early detection is regular screening. Early diagnosis is particularly important since melanoma can be cured with a simple excision if detected early. There still exists a clear demand to improve the quantity and the quality of skin cancer screening. Dermatologists perform a biopsy to ascertain whether a lesion is malignant or benign. Since this procedure involves some expenses as well as morbidity, particularly in patients with multiple atypical moles, alternatively early detection techniques are being sought for rapid, convenient skin cancer screening. Patients having only benign lesions may be given a clean bill of health, whereas patients with a suspicious lesion are referred to a specialized dermatologist or oncologist. In 2016, an estimated 76,380 of these will be invasive melanomas, with about 46,870 in males and 29,510 in women. This study derive a set of computer vision algorithms to automate skin self-examination techniques, that provides a pre-screening tool for individuals in the general population to help assess their risk at ease which matters no workload.

According to the studies, there are different types of skin cancer, such as melanoma, basal cell carcinoma (BCC) and squamous cell carcinoma (SCC). Melanoma is a condition in which the formation of melanin is significantly reduced due to melanocyte dysfunction. One of the main factors affecting melanocyte cells is a sunburn caused by ultraviolet rays from the sun. To classify whether the given lesion is melanoma or not, the standard ABCD rule assess.

A- Asymmetry. One half doesn't match the appearance of the other half.

B- Border irregularity. The edges are ragged, notched, or blurred.

C- Color. The color (pigmentation) is not uniform. Shades of tan, brown, and black are present. Dashes of red, white, and blue add to a mottled appearance.

D- Diameter. The size of the mole is greater than 1/4 inch (6 mm), about the size of a pencil eraser. Any growth of a mole should be evaluated.

The paper is organized as follows: In section 1, a general outline of the introduction of the research work is given. In section 2, the dataset description of the image used in this research work. In section 3, the preprocessing methods used for segmentation of images. In section 4, the segmentation process performed in the system is described. Section 5

0-7695-6330-9/17/31.00 ©2017 IEEE
DOI 10.1109/PDCAT.2017.00028

IEEE computer society

presents the features underlying ABCD rule are extracted. Then, in section 6 the features selection and dimensionality reduction are described. The classifiers used for classification of melanoma is presented in Section 7 followed by the corresponding results and evaluation criteria are described in Section 8. Finally in section 8, the conclusion of the research work is presented.

II. Dataset Description

The International Skin Imaging Collaboration (ISIC) is an international effort to improve melanoma diagnosis. Recently, ISIC efforts to collect an accessible dataset [19] for checking the image of the skin lesion by providing publicly available database ISIC skin image data archive. Images are collected from internationally influential clinical centers, obtained from various devices used in each center. In this study work, we used randomly selected 220 train images and tested with 20 test images.

III. Preprocessing

Skin lesion images usually contain artifacts that make the segmentation process difficult. Skin characteristics, such as freckles, are easily detected by these algorithms based on color or size. Consequently, artifacts such as hair and shading, that usually also are darker than healthy skin may be mistaken as lesions during the segmentation process [3].

A. Hair Removal

The existence of hair and the presence of some artifacts, such as air bubbles and light reflections affects decreasingly in detecting the border of skin lesions and consequently yield erroneous segmentation. So, the standard Dull Razor algorithm is performed as a pre-processing stage for removing hair from images. In this research work, we used the DullRazor tool which is implemented using the following three main steps:

(1) It identifies the dark hair locations by a generalized grayscale morphological closing operation,

(2) It verifies the shape of the hair pixels as thin and long structure, and replace the verified adjacent pixels by a bilinear interpolation, and

(3) It smooth the replaced hair pixels with an adaptive median filter.

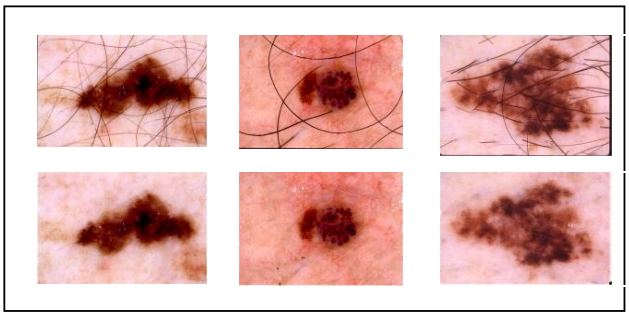

Figure 1. Example of Hair Removed Images.

B. Filtering hair removed images

Noise can stand out even after removing hair from skin lesion images. As bubbles, scratches on the skin are also form of noises. These was removed by performing filtering process. The filter used in this study is median filtering. Such noise reduction is a typical pre-processing step to remove noise from an image in order to improve the results of later processing.

IV. Image Segmentation

Image segmentation is one of the most important part in computer vision according to many studies. Segmentation methods are usually considered with the properties of a certain class of images. In recent years, several methods of segmentation have been proposed for diagnosis of images. Segmentation separates suspicious damage from normal skin area for the further extraction of signs/ features from lesion region. And with automatic segmentation, we can make the classification work easier. In this study, we utilize mean shift segmentation method for detecting the border to separate the lesion from the background skin.

A. Segmenting Lesion Images

Based on the principle that pigmented skin lesions are depigmentation of the skin and with the aim to reduce the computation cost, we used semi-supervised mean-shift algorithm for segmentation in our research work. Mean-shift is a standard algorithm used for clustering given set of data points. One advantage of mean-shift over other parametric techniques like k-means is, we do not have to specify the number of clusters. The algorithm itself finds the best number of clusters for the given data. Mean shift clustering is a powerful, non-parametric iterative method that does not require prior knowledge of the number of clusters and does not limit the shape of the clusters. It considers feature space as an empirical probability density function. If dense regions (or clusters) are present in the feature space, then they correspond to the mode (or local maxima) of the probability density function.

The basic idea is quite simple, but the result is wonderful. We can specify Mean Shift as follows: Firstly, for each data point, the mean shift determines the window around its surroundings, then it calculates the average value of the data points, and then finally it shifts the center of the window to the mean and iterates until convergence. After each iteration, the window moves to a denser region. Segmentation using mean-shift is done using the procedure followed by [7].

The steps involved in using mean-shift to do image segmentation are

1) Compute the joint histogram of the given RGB image.

The number of clusters used for each color is 10. So, there will be 1000 bin joint histogram. Two inputs are required: the spatial range and the color range. Spatial range means which pixels in what neighborhood are to be considered to compute

mean for the pixel of interest. Color range tells pixels of what color in the neighborhood are likely to be considered.

2) Compute the mean.

$$M = \text{sum}(m(i) * p(i)) / \text{sum}(p(i)) \quad (1)$$

m(i) assigns for the pixel to the mean computed whereas p(i) simply denotes the probability of pixel belonging to that m(i)th bin. The mean is computed with the neighborhood and then compute again with the new neighborhood (iterating) until the mean converges. This is done by measuring the distance between the source bin to which the pixel belongs and the new bin to which it belongs after each iteration.

3) Repeat the same procedure for all the pixels in the image.

The figure 2 shows the resultant segmentation images.

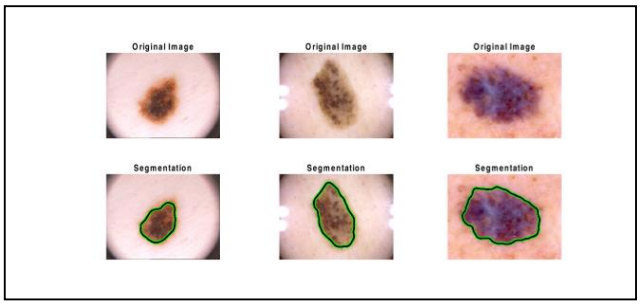

Figure 2. Examples of Segmentation Results.

V. Feature Extraction

After the image segmentation process, feature extraction process is able to perform classification (or diagnosis) of the acquired segmented pigmented skin lesions. Automated diagnostics system tries to produce the features underlying to ABCD rule. In this system, we extract the ABCD features according to [19].

A. Feature for Asymmetry

Melanoma lesions are often irregular, or not symmetrical, in shape. Benign moles are usually symmetrical [19]. The major axis M_1 of the lesion is computed aligning with its longest diameter and passing through its center; the minor axis M_2 is orthogonal to M_1 and also passes through the shape center. The asymmetry features utilized are as following:

TABLE I. Asymmetry Features

Features	Description
F_1	Solidity: the ratio between the lesion area (A) and its convex hull area
F_2	Extent: the ratio between the lesion area and its bounding box area
F_3	Equivalent diameter: $4A/(M_1\pi)$
F_4	Circularity: $4\pi A/(M_1 p)$, where p is the lesion perimeter
F_5	The ratio between the principal axes (M_2/ M_1)
F_6	The ratio between sides of the lesion bounding box
F_7	The ratio between the lesion perimeter p and its area A
F_8	$(B_1 - B_2)/A$, where B_1 and B_2 are the areas in each side of axis M_1
F_9	Similar to f_8, but makes use of the shorter axis M_2
F_{10}	B_1/B_2 with respect to the axis M_1
F_{11}	Similar to f_{10}, but makes use of the shorter axis M_2

B. Features for Border

Typically, non-cancerous moles have smooth, even borders. Melanoma lesions usually have irregular borders that are difficult to define. Pixels at the rim are analyzed to be extended. The rim is dilated by 2 pixels in order to produce a 5 pixels wide region centering at the lesion rim as in [17]. The boundary sharpness is determined by the magnitude of the gradient at each pixel with the Canny operator.

Two additional axes in addition to two principal axes are obtained by rotating by 45 degrees to these orthogonal axes. So, the lesion rim is divided in 8 symmetric regions in order to check the irregularity. For the lesion boundary irregularity, we extract the following features:

TABLE II. Border Features

Features	Description
$F_{12} - F_{14}$	Average gradient magnitude of the pixels in the lesion extended rim [17], in each one of the three I channels
$F_{15} - F_{17}$	Variance of the gradient magnitude of the pixels in the lesion extended rim, in each one of the three I channels
$F_{18} - F_{20}$	Average of the gradient magnitudes of the extended rim pixels values in each one of the three I channels
$F_{21} - F_{23}$	Variance of the gradient magnitudes of the extended rim pixels values in each one of the three I channels

C. Features for Colour Variation

The presence of more than one color (blue, black, white, red, light and brown, blue gray) or the uneven distribution of color can sometimes be a warning sign of melanoma. Benign moles are usually a single shade of brown or tan [19].So, the lesion color variability can be quantified by identifying the occurrence of those color variability within segmented lesion image.

For the features to extract for the color variation in the lesion, we used HSV color channel. The hue occurrence counter for each lesion pixel is created to count nearest typical hue to increase by 1. Typical hues counters are normalized by the lesion area A to generate more color variation features as depicted in Table 3.

TABLE III. COLOR VARIATION FEATURES

Features	Description
F_{24}–F_{27}	Average gradient magnitude of the pixels in the lesion extended rim [17], in each one of the three I channels
F_{28}–F_{39}	Variance of the gradient magnitude of the pixels in the lesion extended rim, in each one of the three I channels
F_{40} – F_{42}	Average of the gradient magnitudes of the extended rim pixels values in each one of the three I channels
F_{43} – F_{48}	Typical hues counters divided by the lesion area A to white, red, light and dark brown, blue gray and black.

D. Features for Diameter

Melanoma lesions are often greater than 6 millimeters in diameter (approximately the size of a pencil eraser). In order to capture textural variation information, the normalized luminance image L is firstly computed as follows:

$$L(x,y) = \frac{\sum_{i=1}^{3} \mathrm{I(x,y)}}{3} \quad (2)$$

Where, I(x,y) is a pixel (x,y) of the normalized color image. The textural variability in L(x,y) is computed by using T(x,y,σ):

$$\mathrm{T(x,y,\sigma)} = \frac{L(x,y)}{\mathrm{S(x,y,\sigma)}} - L(x,y) \quad (3)$$

where, S(x,y,σ) = L(x,y) * G(σ) in which the luminance image L is smoothed by using gaussian filter with standard deviation σ). To capture textural variability of different types of lesions in generic images, T(x,y,σ) for different σ values { $\sigma_1, \sigma_2, \ldots, \sigma_N$} are used in our experiment. We used the σ value as σ=1, 11/7, 15/7, 43/7, and window size for filter is 7σ*7σ suggested in [3].

$$I(\mathrm{x,y}) = (T(\mathrm{x,y}) - \min(T))/(\max(T) - \min(T)) \quad (4)$$

TABLE IV. DIAMETER FEATURES

Features	Description
F_{49}–F_{52}	Maximum, minimum, mean and variance of the pixels intensities inside the lesion segment to represent the textural variation in the Hue channel

We have extracted 52 ABCD features and then two additional features for B rule is then developed. Those are thinness ratio (TR) of the skin lesion and variance of the distance of the border lesion points from the centroid location. The equation for TR is:

$$\mathrm{TR} = \frac{4*pi*A}{\mathrm{perimeter\text{^}2}} \quad (5)$$

TABLE V. ADDITIONAL BORDER FEATURES

Features	Description
F_{53}	Thinness ration to measure the circularity of the skin lesion
F_{54}	Variance of the distance of the border lesion points from the centroid location

VI. FEATURE SELECTION AND DIMENSION REDUCTION

In order to improve classification performance, methods for selecting a subset of relevant features is performed by removing redundant, irrelevant and noisy data from large dimension feature set. We apply feature selection algorithm called RELIEF as in [15] for those extracted ABCD features. The RELIEF algorithm, one of the filter based feature selection methods, is an effective, simple, and widely used approach to feature weight estimation. The weight for a feature of a measurement vector is defined in terms of feature relevance. To those selected features dataset, we perform dimension reduction further. We eliminate those feature vectors which has very little weight values according to Relief algorithm result and then create new feature subset. Figure 3 shows the predictor important weights of feature set values.

Dimensionality reduction is achieved by the use of linear principal component analysis in this research work. It can be achieved by removing data closely related to other data in the set, or by using a combination of data to create a smaller set of functions. A linear projection requires the optimum choice of projections to minimize of a quadratic error. This is achieved by first subtracting the mean value of the data set. The covariance matrix is calculated and its eigenvectors and eigenvalues are found. The eigenvectors corresponding to the largest eigenvalues are conserved, and then the input vector is projected onto eigenvectors to obtain the transformed vector components in the m-dimensional space. Supporting a subset of basis vectors so that only M-coefficients are used, the remaining coefficients can be replaced by the constant. In our system, we compare classification to original features set, selected features set and dimension reduced features set and compare their accuracies.

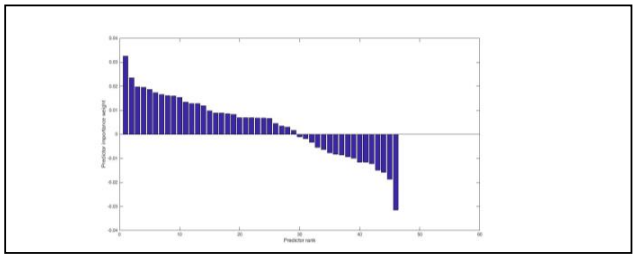

Figure 3. Bar plot of Predictor Importance Weights of Feature Set by Relief Algorithm.

VII. CLASSIFICATION

In recent years, many classification algorithms already exists to classify segmented image. In our research work, we used three classifiers: kNN classifier, Decision Tree, Support Vector Machine (SVM) classifier to classify the segmented lesion as melanoma or non-melanoma. The classification accuracy results are compared and described in next section. SVM classifier to the feature selected data subset with

dimensionality reduction by PCA provides best classification results as shown in table 7.

VIII. Evaluation Criteria

Performance evaluation was conducted using 5 fold cross-validation. In this study, evaluation matrices are used for both segmentation and classification. Segmentation is evaluated using five metrics: pixel level accuracy, pixel level sensitivity, dice coefficient, and jaccard index according to [12].

Pixel Level Accuracy

$$\text{PLA} = \frac{TP + TN}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

Pixel Level Sensitivity

$$\text{PLS} = \frac{TP}{\text{TP} + \text{FN}}$$

Pixel Level Specificity

$$\text{PLSp} = \frac{TN}{\text{TN} + \text{FP}}$$

Dice Coefficient

$$\text{DC} = \frac{2.TP}{2.\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

Jaccard Index

$$\text{JI} = \frac{TP}{\text{TP} + \text{FP} + \text{FN}}$$

Where TP, FP, TN, FN are synonyms for true positive, false positive, true negative and false negatives respectively at the pixel level.

TABLE VI. TP, FP, TN, FN At the Pixel Level

	Our Segmented Pixels	*GroundTruth Segmented Pixels*
TP	1	1
FP	1	0
TN	0	1
FN	0	0

For the classification task, the confusion matrix and the receiver operating characteristic (ROC) curve are used for evaluation.

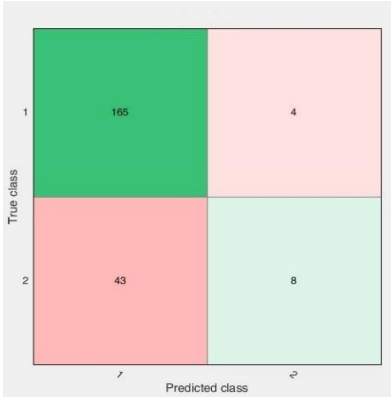

Figure 4. Top Evaluation Accuracy of Classification: Confusion Matrix

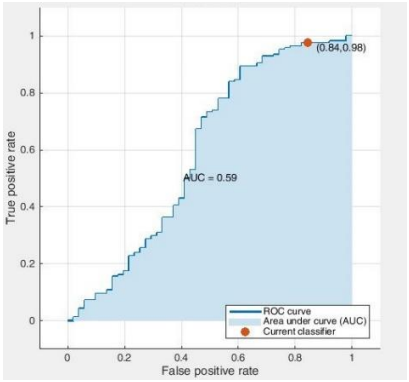

Figure 5. Top Evaluation Accuracy of Classification: ROC Curve

TABLE VII. Top Evaluation Results for Segmentation

PLA	PLS	PLSp	DC	JI
99.2%	71.4%	100.0%	98.6%	97.2%

TABLE VIII. Top Evaluation Results for Classification

Method	Original Feature Set		Selected Feature Subset by Relief	
	No Dimension Reduction	*Dimension Reduced Subset by PCA*	*No Dimension Reduction*	*Dimension Reduced Subset by PCA*
kNN	71.4%	66.8%	73.6%	76.4%
Decision Tree	77.7%	74.1%	77.3%	76.4%
SVM	76.8%	77.7%	76.4%	78.2%

IX. Conclusion

In this research work, segmentation and classification of skin lesion image is performed. With the aim to help the patient to identify the skin cancer without going to hospitals. This diagnosis research work includes both segmentation with MeanShift algorithm and classification using kNN, decision tree, SVM with the highest accuracy of 78.2% which shows promising results.

References

[1] J F Dale Addison, Stefan Wermter, Garen Z Arevian, "A Comparison of Feature Extraction and Selection Techniques"212--215", June 2003, Springer, Istanbul.

[2] Cavalcanti PG, Scharcanski J (2011), "Automated Prescreening of Pigmented Skin Lesions using Standard Cameras", Comput Med Imaging Graph 35(6):481–491.

[3] José Fernández Alcón, C˜alina Ciuhu, Warner ten Kate, Adrienne Heinrich, Natallia Uzunbajakava, Gertruud Krekels, Denny Siem, and Gerard de Haan, "Automatic Imaging System With Decision Support for Inspection of Pigmented Skin Lesions and Melanoma Diagnosis," IEEE J Sel Top Signal Process 3:14–25.

[4] T.Kanimozhi1 , Dr.A.Murthi, "Computer Aided Melanoma Skin Cancer Detection Using Artificial Neural Network Classifier", Singaporean Journal of Scientific Research(SJSR), Journal of Selected Areas in Microelectronics (JSAM), Vol.8.No.2 2016 Pp.35-42.

[5] R.P.L.DURGABAI, “Feature Selection using ReliefF Algorithm", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 10, Oct. 2014.

[6] Cavalcanti, P.G., Scharcanski, J., “Macroscopic Pigmented Skin Lesion Segmentation and Its Influence on Lesion Classification and Diagnosis,” In: Color Medical Image Analysis, pp. 15–39. Springer (2013).

[7] Sumithra, R.; Suhil, M.; Guru, D.S. “Segmentation and Classification of Skin Lesions for Disease Diagnosis,” Proc. Comput. Sci. 2015, 45, 76–85.

[8] Xu, L & Jackowski, Marcel & Goshtasby, A & Roseman, D & Bines, S & yu, Changxiong & Dhawan, Atam & Huntley, A. (1999). “Segmentation of Skin Cancer Images,” Image and Vision Computing, 17. 65-74. 10.1016/S0262-8856(98)00091-2.

[9] S. Anand, S. Mittal, O. Tuzel, P. Meer, "Semi-Supervised Kernel Mean Shift Clustering," Trans. Pattern Analysis and Machine Intelligence, vol. 36, no. 6, pp. 1201-1215, 2014.

[10] M.Chaithanya Krishna, S.Ranganayakulu, “Skin Cancer Detection and Feature Extraction through Clustering Technique,” International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 3, March 2016.

[11] Supriya Joseph, Janu R Panicker, “Skin Lesion Analysis System for Melanoma Detection with an Effective Hair Segmentation Method,” 978-1-5090-1987-8/16/$31.00, 2016 IEEE International Conference on Information Science (ICIS).

[12] David Gutman, Noel C. F. Codella, Emre Celebi, Brian Helba, Michael Marchetti, Nabin Mishra, Allan Halpern, “Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016 hosted by the International Skin Imaging Collaboration (ISIC)" in , May 2016, [online] Available: https://arxiv.org/abs/1605.01397”.

[13] Dr. S.Gopinathan, S. Nancy Arokia Rani, “The Melanoma Skin Cancer Detection and Feature Extraction through Image Processing Techniques”, International Journal of Emerging Trends Technology in Computer Science (IJETTCS), Vol. 5, no 4, pp. 112-116., July-August 2016.

[14] http://www.dermweb.com/dull_razor/

[15] https://homepages.inf.ed.ac.uk/rbf/HIPR2/median.htm

[16] https://en.wikipedia.org/wiki/Median_filter

[17] https://saravananthirumuruganathan.wordpress.com/2010/04/01/introduction-to-mean-shift-algorithm/

[18] https://surajvantigodi.wordpress.com/2013/11/24/meanshift-algorithm-for-segmenting-image/http://www.webmd.com/melanoma-skin-cancer/abcds-of-melanoma-skin-cancer

[19] International Skin Imaging Collaboration Website: http://www.isdis.net/index.php/isic-project