Skin Cancer Detection using Machine Learning Techniques

Vidya M
Department of Electronics and Communication
Ramaiah Institute of Technology
Bangalore, India

Dr. Maya V Karki
Department of Electronics and Communication
Ramaiah Institute of Technology
Bangalore, India

***Abstract*: As increasing instant of skin cancer every year with regards of malignant melanoma, the dangerous type of skin cancer. And the detection of skin cancer is difficult from the skin lesion due to artifacts, low contrast,** and similar **visualization like mole, scar etc. Hence Automatic detection of skin lesion is performed using techniques for lesion detection for accuracy, efficiency and performance criteria. The proposed algorithm applies feature extraction using ABCD rule, GLCM and HOG feature extraction for early detection of skin lesion. In the proposed work, Pre-processing is to improve the skin lesion quality and clarity to reduce artifacts, skin color, hair, etc., Segmentation was performed using Geodesic Active Contour (GAC) which segments the lesion part separately which was further useful for feature extraction. ABCD scoring method was used for extracting features of symmetry, border, color and diameter. HOG and GLCM was used for extracting textural features. The extracted features are directly passed to classifiers to classify skin lesion between benign and melanoma using different machine learning techniques such as SVM, KNN and Naïve Bayes classifier. In this project skin lesion images were downloaded from International Skin Imaging Collaboration (ISIC) in which 328 images of benign and 672 images of melanoma. The classification result obtained is 97.8 % of Accuracy and 0.94 Area** under **Curve using SVM classifiers. And additionally the Sensitivity obtained was 86.2 % and Specificity obtained was 85 % using KNN.**

Keywords: ABCD, HOG, GLCM, SVM, KNN, Navie Bayes.

I. INTRODUCTION

The largest organ of human body is skin which is the outer covering of the body. There are up to seven layers of ectodermal tissues in skin, which guards the underlying muscles, bones, ligaments and internal organs. Skin protects human body from elements and microbes entering the body, it helps in regulating body temperature, and permits the sensation of cold, heat and touch. When the part of skin is abnormal, compared to other parts of skin it is known as skin lesion. The basic and main cause of skin lesion is that infection caused in or on the skin. Skin lesion can be divided into two parts: primary (which is present by birth or developed over the lifetime) and secondary (which is caused by mishandling the primary skin lesion), this can lead to skin cancer, some form of skin cancer is diagnosed in more than three million people in the united states each year. More than 5000 skin cancer patients are admitted and over 4000 people die every year in India. Skin tumors can have categorized into three: Basal cell carcinoma (BCC), Squamous cell carcinoma (SCC) and Melanoma Tumors are considered to be cancer if tumor found to be Malignant which is highly dangerous type of skin cancer because this tumor grows rapidly and spreads to other parts of skin. Whereas benign tumor is not very dangerous type because the tumor grows but does not speared. Therefore, manual detection of skin cancer is not very suitable, since the skin lesion is inspected with the naked eye where feature cannot be observed exactly and results into mistreatment and finally death. Survival rates can be increased by early stage detection of accurate skin cancer. Hence automatic detection is more reliable for better accuracy and efficiency.

Dermoscopy techniques are developed for obtaining the clear skin lesion spot which in turn the visual effect is enhanced by removing reflection. But automatic recognition of skin lesion has a few difficult task such as artifacts, low-contrast, skin color, hairs [1], veins and similar visual of melanoma and non-melanoma [2]. All this can be reduced by pre-processing techniques. The pre-processed skin lesion image is segmented used to get the accurate position of skin lesion. There are several methods for segmentation techniques such as wavelet algorithm, basic global thresholding, region based segmentation, watershed algorithm, snakes method, Otsu method, active contours, and geodesic active contours etc. Segmentation is performed using geodesic active contour [3]. The segmented skin lesion image is later used for feature extraction, there are several methods for extracting features like CASH rule, ABCD rule, ABCDE rule, GLCM, HOG, LBP, and HLIFS etc. ABCD rule is the scoring method where asymmetry, color, border and diameter features are extracted [6]; Authors in this paper explains to take total dermoscopic score and with this value classification of melanoma and non-melanoma is performed and obtained the accuracy of 90 % using wavelet. GLCM basically extracts textural features [7], the extracted feature can be directly passed to neural network and success rate of 95.83 % is obtained. HOG extracts the shape and edge of skin lesion [8]; in this paper, the extracted feature is directly passed to SVM classifier and obtained the accuracy of 97.32 %. Classifier is the last step of skin lesion detection to classify between different classes. This process involves training and testing. Unknown pattern is fed and the idea obtained during training process will classify the

978-1-7281-6828-9/20/$31.00 ©2020 IEEE

unknown patterns. There are different type's classifiers such as SVM, KNN, Naïve Bayes, and neural network etc., Author khan [9] applied features to SVM, KNN and Naïve Bayes classifiers and obtains accuracy of 96 %, 84 % and 76 % respectively. The proposed algorithm and experimental results are explained in this paper which is shown in detail from section 2.

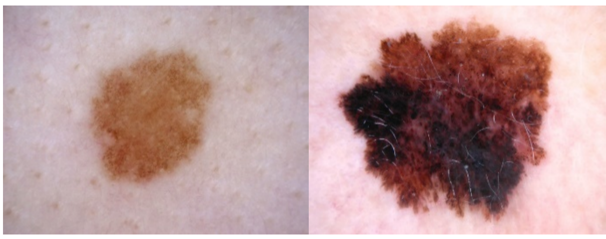
II. PROPOSED METHODOLOGY

The skin lesion detection goes through different phases as shown in Figure. 1. It involves data acquisition, pre-processing of data, segmentation, feature extraction and classification.

Fig. 1 Methodology

A. DATA ACQUISITION

The first phase of this project involves datasets of skin lesion images collected from International Skin Imaging Collaboration (ISIC). Datasets used in this project consists of Benign and Malignant Melanoma type of skin lesions. The Benign lesion images consists of 328 and Malignant Melanoma lesion images consists of 672. Skin lesion images were collected from ISIC 2017 datasets. Images are of JPEG format. Figure. 2.1 (a) And (b) shoes the examples of Benign and Malignant Melanoma images. The skin lesion images were divided into 80:20 ratios for training and testing.



(a) (b)

Fig. 2.1 An example of Skin lesion images (a) Benign image (b) Melanoma image

B. PRE-PROCESSING

The Second phase involves Pre-processing of the skin lesion datasets. Pre-processing removes unwanted things other than lesion to basically identify the lesion in further processes. Unwanted things are such as artifacts, low contrast, hairs, veins, skin colors, moles etc. Are removed by performing the following methods: (i) RGB image is converted to Grayscale, the grayscale image consists only of intensity information which is mostly used by the digital systems. (ii) The grayscale image was passed to median filtering for removing noise which improves skin lesion image, this median filtered image was used for hair detection and removal. (iii) The hair on skin lesion was detected using bottom hat filtering which extracts smallest element from an image such as hair. The detected hair was removed using region filling morphology where inward interpolation was performed on the pixel. Figure. 2.2 (a) to (e) shows the Pre-processed results. Since benign images obtained are less than the melanoma images we augment the benign skin lesion datasets, we crop the skin lesion images and then rotate the skin lesion images to 45º. By which 328 benign images were increased to overcome degradation and over fitting. Figure. 2.3 (a) To (b) shows the augmented results.

C. SEGMENTATION

The third phase involves segmentation of the pre-processed images. Accurate position of skin lesion is found using segmentation process. In this project segmentation was performed using Geodesic Active Contours (GAC). GAC basically senses the maximum variations in the overall skin lesion which is usually found at the boundaries of the skin lesion. Pre-processed skin images are binarized by applying Otsu thresholding and this binarized image is applied with GAC. Figure. 2.4 (a) to (c) shows the result of segmented images. In the proposed algorithm, the evaluation of segmentation process was mainly used are Jaccarb Index (JA) and Dice Index (DI), where JA framework obtained was compared with the existing work is listed below in Table □.

D. FEATURE EXTRACTION

Fourth phase involved is feature extraction from the segmented skin lesion. The feature extraction was performed to get the accurate information about the skin lesion which may be border, color, diameter, symmetrical, textural nature of skin lesions, with the help of these extracted information. Detection of cancerous skin is easily performed. Three types of feature extraction were performed: ABCD, GLCM and HOG.

a *ABCD (ASYMMETRY, BORDER, COLOUR, DIAMETER)*

- ABCD rule identifies the symmetrical nature, border, colour and diameter of skin lesion. Features are extracted as described below.

- Symmetrical: The skin lesion was divided the lesion into four axes as shown in Figure. 2.5 (b) to check for symmetrical nature, if the lesion found is not symmetrical i.e., Asymmetry then lesion can be said cancerous.

- Border: Differentiate the lesion part of skin to other part of skin detection of border is necessary to find exact location, shape of lesion. This was obtained by dividing the lesion into eight slices of axis as shown in Figure. 2.5 (c).

- Colour: There are six colours that can be used for predicting the cancerous or non-cancerous lesions such as black, blue-grey, dark-brown, light-brown, red and white. The results are shown in R (red), G (green), B (blue) form as shown in Figure. 2.5 (d). Each colour values is tabulated below in Table □.

- Diameter: skin lesion diameter was checked which must be less than 6mm for non-cancerous type.

b *GLCM (GRAY LEVEL CO-OCCURRENCE MATRIX)*

GLCM is usually used for textural analysis where the distributed intensity from an object is obtained. GLCM considers two pixels where one is neighbour pixel and other is reference pixel. Using GLCM features obtained are contrast, correlation, energy, entropy, homogeneity, prominence and shade. Calculation of each feature is defined below.

- Contrast: spatial frequency of texture in a skin lesion was measured using the below formula.

$$\sum_{i,j} (i-j)^2 c(i,j)$$

- Correlation: Grey level linear dependencies of a skin lesion was measured using the below formula.

$$\frac{[\sum_{i,j} c(i,j)] - \mu_x \mu_y}{\sigma_x \sigma_y}$$

- Energy: Degree of disorder of a skin lesion was measured using the below formula.

$$\sum_{i,j} c(i,j) \log(c(i,j))$$

- Homogeneity: Distribution of element in the skin lesion is measured using the below formula.

$$\sum_{i,j} \frac{c(i,j)}{1+(i-j)^2}$$

Where; c (i, j) is the normalized entry in the row I and the column j. and i is the intensity of one pixel and j is the intensity of next pixel.

c *HOG (HISTOGRAM OF ORIENTED GRADIENTS)*

Shape and edge information is extracted using HOG. Orientation Histogram is performed on lesion to find the edge intensity. There are basically two units for computing this purpose namely cell and block. The cell gradient orientation is shown in below Figure 2.6.

E. CLASSIFICATION

There are several models to distinguish between cancerous and non-cancerous skin lesions. Most widely used are Machine learning techniques for classification of lesion types. The most used are SVM, KNN, Naïve Bayes, Neural Network etc., in this work, SVM, KNN and Naïve Bayes classifier are used where extracted features are directly given to the classifier.

SVM: Most popular type of classifier is support vector machine (SVM). The basic advantage is unified framework; i.e., different type of machine learning architectures can be generated through a kernel choice. SVM classifier are applied on the features obtained to the benign and melanoma skin lesion.

KNN: K-Nearest Neighbor classifier is the simplest, easiest, quick and efficient. An image is classified based on the majority votes of its neighbors. The training and testing samples are given to the KNN classifier where each class is calculated by using nearest distance. KNN classifier are applied on the feature obtained to the benign and melanoma.

Naïve Bayes: Naïve Bayes analysis is basically based on the prior probability belief which is based on Bayes theorem. The main advantage is large data is not required, fast and conditional independence i.e. there is no dependency among the attributes. Naïve Bayes is the supervised machine learning technique.

Results of classification of proposed method is tabulated in Table □ and □ as shown below. Comparison is made for ABCD rule result where improved accuracy of 96 % is obtained by our proposed method which is tabulated in Table □ as shown below.

III. EXPERIMENTAL RESULTS

The proposed method is applied to skin lesion images collected from ISIC. The datasets consist of 328 benign images and 672 melanoma images. The classifiers are trained with different train and test ratios. 10-fold validation has been performed. With train to test ration of 80:20 results are tabulated below in Table □.

As explained in methodology three different features extractors and three classifiers models are used. Scoring feature are extracted by applying ABCD rule and texture features are extracted from GLCM and HOG methods. Sample feature extracted from these methods are shown in Figure. 2.5 (a) to (d) (ABCD) and Figure. 2.6 (a) To (b) (HOG).

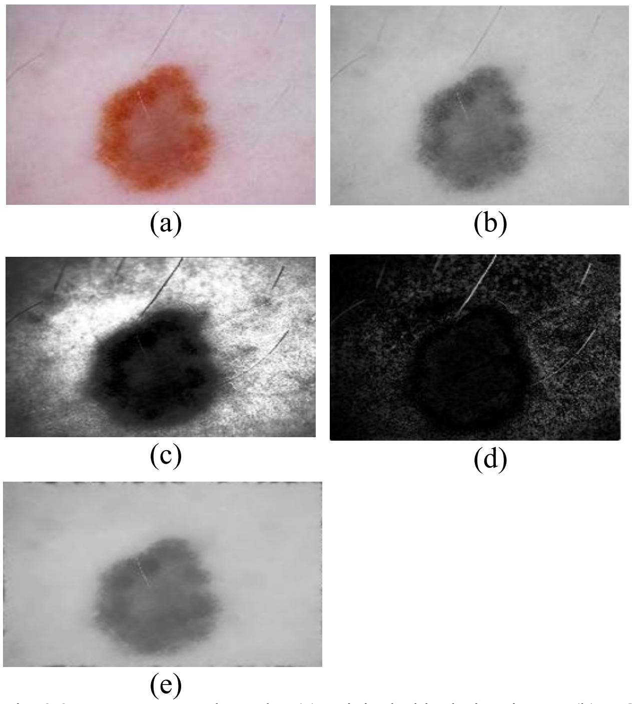


Fig. 2.2 Pre-processed results (a) original skin lesion image (b) RGB to gray converted image (c) Median filtered image (d) Hair detected (e) Hair removed.

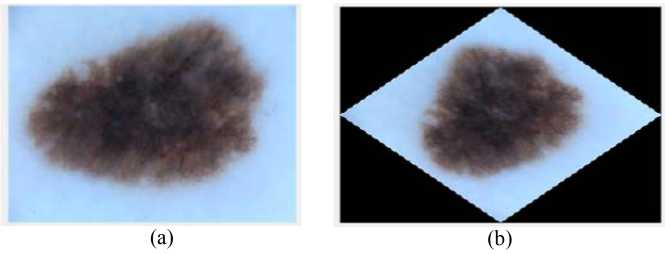


Fig. 2.3 Augmented results (a) cropped image (b) rotated image

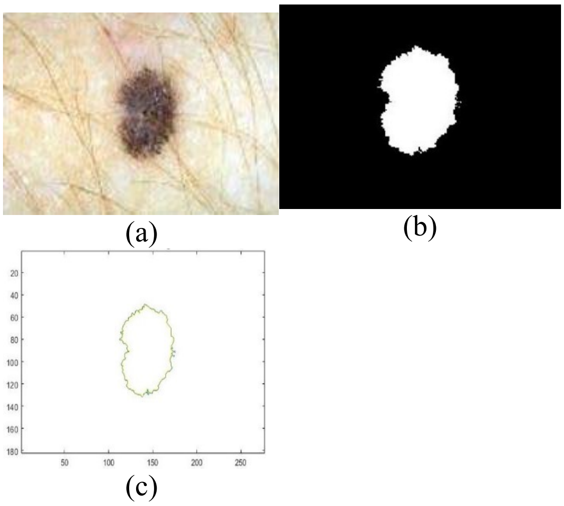


Fig. 2.4 Segmented results using GAC (a) original lesion image (b) Otsu thresolding (c) Active contour.

TABLE I Segmentation performance result.

Segmentation Method	JA	DI
LIN [5]	0.710	0.85
Proposed Method	**0.9**	**0.82**

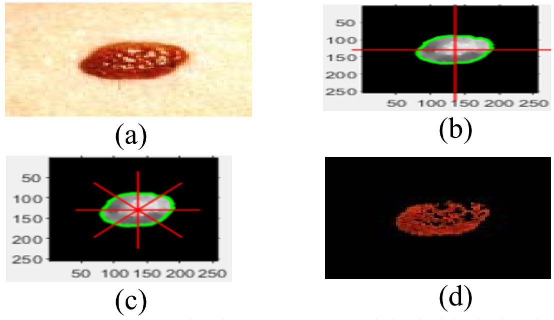


Fig. 2.5 Feature extracted using ABCD (a) original skin lesion image (b) Asymmetry of lesion (c) Border of lesion (d) Colour of lesion.

TABLE II Sample colour feature score result (c=2).

	Color count	Percentage	Score
Black	77	0.0052888	0
White	6768	0.46487	0
Red	0	0	0
Light Brown	75900	5.2132	1
Dark Brown	95013	6.526	1
Blue Gray	0	0	0

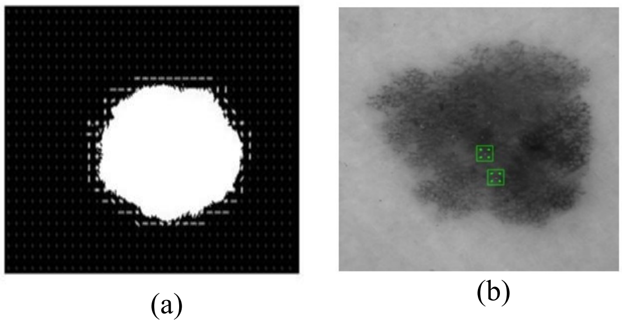


Fig. 2.6 (a) and (b) Feature extracted result using HOG (using strongest corner)

Performance of classifier models are evaluated by calculating specificity (SP) and sensitivity (SE). They are defined as

$$Specificity = \frac{TN}{TN + FP}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

Where, TP = correctly classified positive class (True Positive).
TN = correctly classified negative class (True Negative).
FP = incorrectly classified positive class (False Positive).
FN = incorrectly classified negative class (False Negative).

The classification accuracy (AC) of 97.8 % and ROC curve (AUC) of 0.94 is obtained from a SVM classifier when compared to other classifiers with HOG and GLCM features. This result is shown in Table □ where features ABCD, GLCM, HOG and hybrid features of GLCM + HOG, GLCM + HOG + ABCD and GLCM + HOG + Colour are classified using different classifiers and observed that texture feature using SVM classifier accuracy is better. Large value of Negative predicted (NP) of 20 % is obtained from SVM classifier when compared to other classifier. This result is shown in Table □ where statistical parameter was used to see the performance of classifier. The extracted features can be used in computer-assisted skin cancer tool for better results. However 139 skin lesion images are misclassified. For the SVM classifier Hinge Loss calculated was 0.3. In summary, the SVM classifier outperforms with the 97.8 % accuracy and 0.94 AUC for textural features.

TABLE III Testing results

	SVM Classifier
GLCM + HOG	**97.8 %**

TABLE IV Statistical result

Classifier's	**Specificity %**	**Sensitivity %**	**Positive Predicted Value %**	**Negative Predicted Value %**
SVM	75	65	80	20
KNN	**85**	**86.2**	87	13
Naïve Bayes	82	72	85.2	14.8

TABLE V Classification AC and AUC of proposed method

Features	**Classifier's**					
	SVM		**KNN**		**Naïve Bayes**	
	AC	**AUC**	**AC**	**AUC**	**AC**	**AUC**
GLCM	82 %	0.6	72 %	0.75	75 %	0.79
HOG	87 %	0.71	79.4 %	0.8	82 %	0.779
ABCD	96 %	0.89	81.2 %	0.81	85.1 %	0.85
HOG + GLCM	**97.8 %**	**0.94**	**95 %**	0.83	**91.2 %**	0.9
HOG + GLCM + ABCD	85 %	0.87	83 %	0.75	77 %	0.80
HOG + GLCM + Color	**92.2 %**	0.89	89.6 %	0.88	86.8 %	0.94

TABLE VI Comparison of Automatic ABCD rule implementation

	Dermoscopic images	AC
Kasmi et al. [6]	EDRA international atlas of dermoscopy (200 images)	94 %
Proposed method	ISIC 2017 (250 images)	**96 %**

IV. CONCLUSION

In this paper, hybrid feature extraction has been used to classify skin lesion either as benign or melanoma. Automatic detection of skin lesion with ABCD rule, GLCM and HOG for feature extraction and classification using machine learning techniques. GAC method was proposed for segmentation of the skin lesion. Segmentation result with 0.9 JA and 0.82 DI has been achieved. For color, symmetry and diameter of skin lesion ABCD rule, texture of skin lesion GLCM and shape, edge of skin lesion HOG was proposed for feature extraction. Different machine learning techniques such as SVM, KNN and Naïve Bayes was proposed to address the classification. The proposed method was performed on skin lesion images of ISIC datasets. Comparing all the classification methods SVM out performs with other classifiers for AC of 97.8 % and AUC of 0.94. Sensitivity and Specificity obtained are 86.2 % and 85 % respectively using KNN. From the results obtained we can observe that accuracy obtained is better after augmentation performance. This method further can be implemented on the neural network platform for better accuracy.

REFERENCES

[1] Jaisakthi, Seetharani Murugaiyan, Palaniappan Mirunalini, and Chandrabose Aravindan. "Automated skin lesion segmentation of dermoscopic images using GrabCut and k-means algorithms." *IET Computer Vision* 12, no. 8 (2018): 1088-1095.

[2] Chung, Do Hyun, and Guillermo Sapiro. "Segmenting skin lesions with partial-differential-equations-based image processing algorithms." *IEEE transactions on Medical Imaging* 19, no. 7 (2000): 763-767.

[3] Hemalatha RJ, Thamizhvani TR, Dhivya AJ, Joseph JE, Babu B, Chandrasekaran R. Active Contour Based Segmentation Techniques for Medical Image Analysis. Medical and Biological Image Analysis. 2018 Jul 4:17.

[4] SALIH, SHAKAR H., and S. H. E. R. E. E. N. AL-RAHEYM. "Comparison of Skin Lesion Image Between Segmentation Algorithms." *Journal of Theoretical and Applied Information Technology* 96, no. 18 (2018).

[5] Li, Yuexiang, and Linlin Shen. "Skin lesion analysis towards melanoma detection using deep learning network." *Sensors* 18, no. 2 (2018): 556.

[6] Kasmi, R. and Mokrani, K., 2016. Classification of malignant melanoma and benign skin lesions: implementation of automatic ABCD rule. *IET Image Processing*, *10*(6), pp.448-455.

[7] Goel, Rashi, and Saranjeet Singh. "Skin cancer detection using glcm matrix analysis and back propagation neural network classifier." *International Journal of Computer Applications* 112, no. 9 (2015).

[8] Bakheet, Samy. "An svm framework for malignant melanoma detection based on optimized hog features." *Computation* 5, no. 1 (2017): 4.

[9] Khan, Muhammad Qasim, Ayyaz Hussain, Saeed Ur Rehman, Umair Khan, Muazzam Maqsood, Kashif Mehmood, and Muazzam A. Khan. "Classification of Melanoma and Nevus in Digital Images for Diagnosis of Skin Cancer." *IEEE Access* 7 (2019): 90132-90144.

[10] Victor, Akila, and M. Ghalib. "Automatic detection and classification of skin cancer." *International Journal of Intelligent Engineering and Systems* 10, no. 3 (2017): 444-451.