

# Detection of skin ulcer at early stages using Otsu's segmentation and Naïve Bayes Classifier

**Abstract**—Skin Ulcers are likely to be caused by the increase in UV radiation which occurs as a result of ozone depletion with the culture of the sun. Major other causes of skin cancer are burns, scars, sores, radiation or certain chemicals like arsenic. Moreover, tattooing is considered as a fashion in the younger generations, unaware of the fact that it can cause a skin cancer. However, if detected early, all forms of skin cancers are curable. This relies heavily on classifying skin lesion at an early stage. Skin lesion classification involves data from patient concerning both, their individual features and wound origin to be collected. Skin ulcer images and medical diagnosis about its grade can be stored, thereby submitting these data to the data mining procedures in order to detect some relations between them. Detection of Stage I skin ulcers become more difficult by unaided visual inspection. Therefore, patients are more prone towards developing Stage –II and Stage –III skin ulcers. Research has been done to develop a low-cost, non-contact, imaging-based Stage-I skin ulcer detection system for use by support staff in assisted living and skilled nursing facilities to increase their ulcer detection rate. Also, study of necessary algorithms makes it easier to detect skin ulcer and heal the patient rather in early stages which gives a visual aid in detection of skin lesion at early stage but it does not help the dermatologists to exactly classify the skin lesion into malignant and non-malignant class. To overcome this problem further classification of skin lesion is required. This paper presents implementation of skin lesion detection by classifying the skin cells into various classes of malignant and non-malignant cells. Initially, skin lesion is segmented using Otsu's thresholding technique and affected region is found followed by feature extraction and classification. Studies show that Naive Bayes classifier classifies the lesion giving highest accuracy among all the other classifiers. Therefore, for further classification Naive Bayes Classifier is used which can help the medical experts in classifying the skin lesion into various classes of malignant and non-malignant cells. The experiments are conducted on different skin types. The experiments results showed that Naive Bayes classifiers gave high accuracy and the proposed method for detection of skin ulcers can effectively improve the detection of skin ulcers. Further, result analysis using various skin lesions are also presented in the paper.

**Keywords**—Naïve Bayes classifier; Otsu's thresholding; segmentation; texture;

Patel Jigna J  
Computer Engineering Department  
Sarvajanik College of Engineering and Technology  
Surat, India

Bhatt Bhumika  
Computer Engineering Department  
Sarvajanik College of Engineering and Technology  
Surat, India

## I. INTRODUCTION

Skin cancer can cause death. It is increasing problem around the world. Skin cancer result in 80,000 deaths a year as of 2010; 49,000 of which are due to melanoma and 31,000 of which are due to non-melanoma skin cancers. Skin cancer is caused mainly due to the increase in UV radiation caused by the ozone depletion with the culture of the sun. Other factors resulting skin cancer are burns, scars, sores, radiation or certain chemicals like arsenic. However, if detected earlier, all forms of skin cancer is almost 100% curable. This relies heavily on effectively classifying skin lesion at an early stage [11]

Skin cancers (skin neoplasm) are named after the type of skin cell from which they arise. Basal cell cancer originates from the lowest layer of the epidermis, and is the most common but least dangerous skin cancer. Squamous cell cancer originates from the middle layer, and is less common but more likely to spread and, if untreated, become fatal. Melanoma, which originates in the pigment-producing cells (melanocytes), is the least common, but most aggressive, most likely to spread and, if untreated, become fatal [12]

Main objective in classifying skin lesion is to differentiate between non-ulcerous and ulcerous tissue of the skin and classify the skin lesion accordingly. The skin is the body's largest organ and weighs about 2 kg in a 70 kg person. For an adult, the surface is between 1.5 and 2.2 m<sup>2</sup> [15]. Moreover, by visual inspection it becomes difficult for the dermatologist to detect initial stage of ulcers like erythema which can result in most fatal Stage-II and Stage-III ulcers. Therefore, classifying the skin lesion correctly with use of various algorithms can be helpful to the dermatologists and lab experts to give correct result which can prevent the patients from developing more fatal ulcers. Also, by this classification process the skin ulcers can be classified into the various class i.e. melanocytic classes and non-melanocytic classes.

### A. Related Work

Different researchers has carried out various research work related to skin ulcer detection amongst which few has been presented in this section.

Prabhu Jude Rajendran et al proposed a low-cost, non-contact imaging-based Stage I Pressure Ulcer detection system for use by support staff in assisted living and skilled nursing facilities to increase the ulcer detection rate over a wide range of skin colors using CLAHE algorithm to enhance the image and hence aid in visual detection of skin lesion [1].

R.Guadagnin, R. de S. Neves et al used Isodata and K-means classification procedures which classified the skin lesion based on clustering [2].Do Hyun Chung et al proposed partial-differential equations (PDE)-based system for detecting the boundary of skin lesions in digital clinical skin images [3].

Scott E Umbaugh et al has addressed the problem of segmentation of a digital image based on feature called variegated coloring for identification [4].Aswin.R.B et al has described a technique to differentiate between benign and malignant melanoma using Artificial Neural Network based Classification methodology.The classifier classifies the given datasets into benign and melanoma cells [15].

### B. Classification Methods

For perfectly classifying any skin lesion the methodology to be followed is segmentation, feature extraction and classification. For classification purpose supervised learning algorithms gives the highest accuracy. Classification gives separate class labels or ordering to objects into classes. Classification can be applied in many fields such as medical, astronomy, commerce, biology, media, etc. Decision Tree, Naïve Bayes, k-Nearest Neighbor, Neural Networks, Support Vector Machine, and Genetic Algorithm are generally used for classification purpose.

## II. PROPOSED METHODOLOGY

The project proposed over here is structured in two main phases. In the first phase segmentation, pre-processing and feature extraction is followed. Secondly, the implementation of a classifier is carried out.

### A. Segmentation

Image segmentation is important in many computer vision and image processing applications. The regions of the image are divided into objects of interest is necessary before any processing can be done at a level higher than that of the pixel. Identification of real objects, pseudo –objects, shadows or actually finding anything of interest within the image requires some form of segmentation. The segmentation technique used in proposed methodology is Otsu's segmentation technique Segmentation techniques are divided into three main groups:

- 1) Region growing
- 2) Clustering methods, and
- 3) Edge detection

Clustering methods are separated from region growing methods because, in this context, region growing methods are restricted to methods that primarily use the spatial domain, the two dimensional row and column image space, while the clustering techniques could be applied to any domain (spatial domain, color space, feature space, etc.)

**Region Growing:**In region growing the main aim is to find regions that represent objects or meaningful parts of objects. Some of the techniques used are local, in which small areas of the image are processed; others are global, where the entire image is considered during processing.Methods that can combine local and global techniques, such as split and merge, are referred to as state space techniques and use graph structures to represent the regions and their edges (boundaries).

**Clustering Technique:** Clustering techniques are image segmentation methods whereby individual elements are placed into groups; these groups are based on some measure of similarity within the group.One method of image segmentation based on clustering, in widespread use, is the method of taking the space of interest and splitting the space into regions by setting limits on each of the dimensions for each separate region. Recursive region splitting is a clustering method that has become a standard technique. This method uses a thresholding-of-histograms technique to segment the image. A set of histograms is calculated for a specific set of features, and then each of these histograms is searched for distinct peaks. The best peak is selected and the image is split into regions based on this thresholding of the histogram.

**Edge Detection:** Edge detection, as a method of image segmentation, is performed by finding the boundaries between objects, thus defining the objects themselves, indirectly. This method is usually implemented by first marking points that may be a part of an edge. These points are merged into line segments, and the line segments are then merged into object boundaries. The most common method of finding edges in a digitized image is to apply a spatial differentiation operator to small blocks of pixels, local neighborhoods, within the digitized image [4].

The first stage involved in this project is pre-processing followed by image segmentation using Otsu's Segmentation Technique. Pre-processing of raw image data is required to enhance the image. CLAHE Algorithm has been followed for the same. Otsu's thresholding is clustering based image segmentation technique or reduction of a gray level image to a binary image. This algorithm assumes that the image to be threshold has two classes of pixels or bi-modal histogram and then it calculates the optimum threshold that separates those two classes in such a way that it has minimum variance. The basic principle for

segmenting the normal skin class with the lesion skin class is to use a threshold for RGB or texture or a combination of both to determine whether the pixels belongs to normal skin class or lesion skin class. The first step involves loading of image and manually cropping the image. The image is resized accordingly, followed by improving the contrast of each individual channel of the RGB image using CLAHE algorithm. Finally, the image is segmented using OTSU thresholding technique.

### B. Feature Extraction

To classify an image, some features must be first extracted. Various features such as color, geometric, texture and islands of color are present in the digitized image. The "ABCD rule" is a widely accepted tool used by Dermatologists to support early detection of skin ulcer. ABCD stands for Asymmetry, Border irregularity, Color variegation, and Diameter. Here lesions are evaluated for asymmetry in color, shape, and topography. Border Irregularity refers to how the borders vary irregularly with no definite shape or definite separation of the skin lesion from the surrounding normal skin. Variegation refers to a spotty or disorganized display of colors namely brown, black, blue-grey, pink, and white. However uniformly black lesions also acts as indicators of melanoma. The diameter of the lesion must be compared with that of the patient's other lesions. The diameter of melanoma is usually greater than 6mm.

**Asymmetry:** One-half of the lesions do not match the other.

**Border** : The borders of the lesion are irregular, ragged, notched, or blurred.

**Color** : The color of the lesion varies over with different shades of brown or black, red, white or blue

**Diameter** : The area is larger than 6 mm or is growing[11].

Various features are extracted to classify the image. For the purpose of this project three categories of features are extracted and implemented as follows:

1. Color features.
2. Texture features.
3. Shape features.

The color features extracted from the skin lesion are the mean and variance of the RGB and HSV color space. A total of twelve color features were extracted which contains three mean of R, G, and B values of the skin spot. Similarly, the variance of R, G and B value is estimated. Also the image is then converted to HSV color space and mean and variance of H,S and V are obtained. In all twelve color features were extracted.Two texture features were extracted, one of which was a transformation applied to each and

every pixel of green values which is 8 times its red value, subtracted by the sum of all its neighboring G values and stored in array v1. Similarly, for each and every pixel of red values, a transformation is applied which is 8 times its green value, subtracted by the sum of all its neighboring G values and stored in array v2.The mean of v1 and v2 are the two texture features.

Finally, three shape features namely area, eccentricity and perimeter were extracted. Area is the actual number of pixels in that region. Perimeter is the distance around the boundary of region. Eccentricity is also known as elongation.

### C. Classification

For the purpose of classification of skin ulcers many classifiers can be used. In pattern recognition, the  $k$ -nearest neighbors algorithm ( $k$ -NN) is non-parametric for classification and regression, that predicts objects' "values" or class memberships based on the  $k$  closest training examples in the feature space.The  $k$ -nearest neighbor algorithm is amongst the simplest of all learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its  $k$  nearest neighbors. If  $k = 1$ , then the object is simply assigned to the class of that single nearest neighbor.

Decision tree is another classifier that can be used to classify a skin lesion. Decision tree learning uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. In data mining, a decision tree describes data but not decisions; rather the resulting classification tree can be an input for decision making. Decision Trees are considered to be one of the most popular approaches for representing classifiers. Many researchers from various disciplines have dealt with the issue of growing a decision tree from available data e.g. statistics, machine learning, pattern recognition, and Data Mining.

The classifier implemented is Naïve Bayes Classifier. Bayesian analysis is based on a prior probability belief. Naïve Bayes classifiers assume that a variable value on a given class is independent of the values of other variable. Naive Bayes classifier is a simplistic probabilistic classifier which is based on Bayes theorem and on strong independent assumptions. The reason for selecting Naïve Bayes classifier is that Bayesian classifiers assume that there are no dependencies amongst attributes and this assumption is called class conditional independence. It simplifies the computations involved. It is supervised learning technique.

The advantages of Naive Bayes:

- 1) It uses a very intuitive technique. Bayes classifiers, unlike neural networks, do not have several free parameters that must be set. Due to this the design of the classifier becomes simple.
- 2) Since the classifier returns probabilities, it is simpler to apply these results to a wide variety of tasks than if an arbitrary scale was used.

- 3) It does not require large amounts of data before learning can begin.
- 4) Naive Bayes classifiers are computationally fast when making decisions.

*D. Pseudocode for the proposed algorithm*

1. Load the image.
2. Manually crop the image  
Segmentation begins
3. Divide the RGB image into R channel, G channel, and B channel.
4. Contrast stretching of individual channel of the image.
5. Concatenate the channel to make RGB image.
6. Perform CLAHE on individual contrast stretched image.
7. Concatenate the entire CLAHE operated channel.
8. Take the green channel of the contrast stretched image and convert the image pixels to double precision.
9. Perform Otsu's thresholding on the double precision image.
10. Store the pixels of the green channel of the image into an array.
11. If pixels value is less than some threshold  $t$ , then store it as 1 else 0.
12. Fill the holes in the binary image by creating a disk shaped structuring element and performing erosion followed by dilation.
13. Remove all the connected components which are having pixels less than 200 forming another object in the image.
14. Suppresses all the structures that are lighter than their surroundings and connected to the border.  
Feature Extraction
15. Convert the image pixel to double precision.
16. Extract mean and variance of R GB color space and HSV color space.
17. Convert the image into HSV color space and extract 2 texture features.
18. Extract 3 shape features namely area, eccentricity and perimeter.  
Classification
19. Read the excel file in which the features of trained images are stored.
20. Classify the test image against the trained images and predict the output.
21. Now, the image is classified into one of the malignant or non-malignant class.

Table I –Pseudo code for proposed algorithm

III. RESULT ANALYSIS

The experiments were conducted using Matlab R2012a. A total of 183 images were taken for training dataset among which 28 images belongs to Class-1, 31 images belongs to Class-2, 26 images belongs to Class-3, 12

images belongs to Class-4, 30 images belongs to Class-5, 32 images belongs to Class-6 and 24 images belongs to Class-7, respectively. Similarly, 41 test images were taken for the experiment among which 4 images belongs to Class-1, 5 to Class-2, 6 to Class-3, 3 images to Class-4, 8 images to Class-5, 5 images to Class-6 and 10 images to Class-7 and were classified correctly.

Result of Squamous cell carcinoma is shown below



Fig1. Original SCC skin lesion

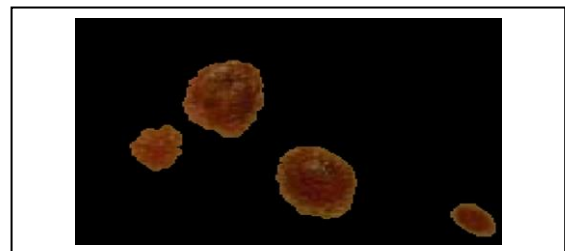


Fig 2. Segmentation of SCC skin lesion

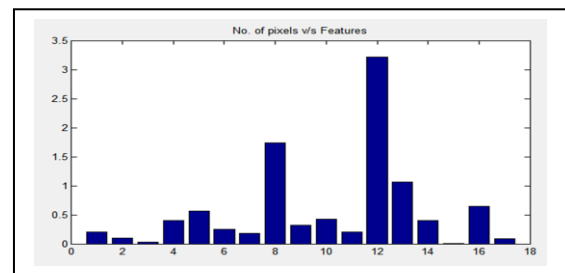


Fig 3. Features extraction

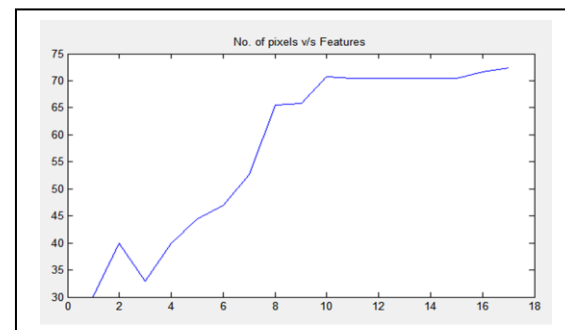


Fig 4. Accuracy graph

Confusion matrix is created to measure the performance of the classifier. Confusion matrix comprises of the

actual and predicted classifications performed by the classifier. The column of the matrix represents the instances in a predicted class and row represents the instances in an actual class. After classification of all the images correctly, accuracy was calculated. The graph below shows the accuracy for 7 class classification using 17 different features. Highest accuracy of 72.5 % was achieved.

#### IV. CONCLUSION AND FUTURE WORK

From the experiment conducted, it is found that the proposed method can be very useful for the dermatologists and lab experts to detect and classify correctly the skin lesion into various ulcerous classes and non-ulcerous classes. Moreover, the classifier used gives the highest accuracy which is very much necessary for skin lesion classification. Future work consists of classification of skin lesion using different features from different color space. Further various other classifiers can be utilized for classification purpose.

#### Reference

- [1] Prabhu Jude Rajendran, Leachtenauer Jon, Kell Steve, Turner Beverly, Newcomer Chris, Lyder Courtney, AlwanMajd, ,2006—Improving the Detection of Stage I Pressure Ulcers by Enhancing Digital Color Images, 28th Annual International Conference of the IEEE In Engineering in Medicine and Biology Society, 2006., pp.5206-5209.
- [2] Do Hyun Chung, G. Sapiro, July'2000 —Segmenting Skin Lesions with Partial-Differential-Equations-Based Image Processing Algorithms, Medical Imaging, IEEE Transaction on (Volume:19, Issue: 7), pp. 763 –767.
- [3] Yan Pingkun, Zhang Wuxia, TurkbeyBaris, L. Choyke Peter, LiXuelong, 2013, —Global structure constrained local shape priorestimation for medical image segmentation, Elsevier, Pages1017–1026.
- [4] R.Guadagnin, R. de S. Neves, S.F.Silva, E.F.Rocha Jr., L.A.Santana, and E.Ferneda, June 2009,—Basic ideas to Support PressureUlcersTreatment by Means of Image Mining, Springer, pp 239-244
- [5] MaglosiannisIlias, N. DoukasCharalampos, September 2009,—Overview of Advanced Computer Vision Systems for Skin Lesions Characterization, pp.721 – 733.
- [6] Shinkar Anita and DevalePrakash, September'2007 Contrast Enhancement Technique for Medical Images, pp.303-306, IEEE'
- [7] S.E.Umbaugh, R.H. Moss, W.V. Stoecker, , August 2002,—Automatic color segmentation of images with application to detection of Variegated coloring in skin tumors. I, Engineering in Medicine and Biology Magazine, IEEE (Volume:8, Issue: 4), pp.43 – 50.
- [8] S. NevesRinaldo, F. Silva Simônia, F.Rocha JrEdvar, A. SantanaLevy, Guadagnin Renato and FernedaEdilson, —A Proposal for Automatic Inference of Pressure Ulcers Grade based on Wound Images and Patient Data, Hospital de Apoio de Brasília (HAB), SGAN, Lote 14, Asa Norte 70.620-000 Brasília
- [9] Chalana Vikram and Kim Yongmin, , Oct. 1997, IA Methodology for Evaluation of Boundary Detection Algorithms on Medical Images, Medical Imaging, IEEE Transactions on (Volume:16, Issue: 5), pp. 642 – 652.
- [10] Hriday Ravindranath, 2006, —Skin spot classification using 3D data, School of Informatics, University of Edinburgh, S0568335.
- [11] M. Mary Synthuja Jain Preetha, L. Padma Suresh, M. John Bosco, 2012 —Image Segmentation Using Seeded Region Growing, 01/2012; DOI:10.1109/ICCEET.2012.6203897, ICCEET'.
- [12] Sharma N, Ray AK, Sharma S, Shukla KK, Pradhan S, Aggarwal LM, , July'2008, Image Segmentation and classification of medical images using texture-primitive features: Application of BAM-type artificial neural network, J Med Phys., 33(3):119-26. doi:10.4103/09716203.42763
- [13] HE Jin-qi, Zhang Li-ping, Dia Hongliang, 2010, —An enhancing Method based on locally distributed Histogram for medical image, IEEE
- [14] Yi Dingrong, Kong Lighua, Sprigle Stephan, 2010, —Detecting early stage pressure ulcer on dark skin using multispectral imager —, SPIE Proc. SPIE 7560, 75600U.
- [15] K. Al-Mohair Hani, Junita Mohamed Saleh, Azmin Suandi Shahrel, Dec'2012, —Human Skin Color detection: A review on neural network perspective, International Journal of Innovative Computing, pp.8115-8131
- [16] Nidhal Khadhair El Abbadi and Abbas Hussein Miry, “Automatic segmentation of skin lesions using histogram thresholding”, J. Computer Science'2014
- [17] Ahmad Ashari, Iman Paryudi, A Min Tjoa, “Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool”, IJACSA'2013
- [18] Wen-Yu Chang<sup>1</sup>, Adam Huang<sup>2</sup>, Chung-Yi Yang<sup>3</sup>, Chien-Hung Lee<sup>4</sup>, Yin-Chun Chen<sup>5</sup>, Tian-Yau Wu<sup>6</sup>, Gwo-Shing Chen<sup>7</sup>, “Computer-Aided Diagnosis of Skin Lesions Using Conventional Digital Photography: A Reliability and Feasibility Study”, PLoS ONE'Nov'2013
- [19] <http://www.dermafoto.com/erythema>
- [20] <http://imagezone.org/images>