Summer 2016

Thresholding techniques for discrimination of seborrheic keratosis from melanoma in dermoscopy images

Pramada Kishtagari

Department:

THRESHOLDING TECHNIQUES FOR DISCRIMINATION OF SEBORRHEIC KERATOSIS FROM MELANOMA IN DERMOSCOPY IMAGES

by

PRAMADA KISHTAGARI

A THESIS

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

IN

ELECTRICAL ENGINEERING

2016

Approved by:

Dr. Randy H. Moss, Advisor

Dr. William V. Stoecker

Dr. R. Joe Stanley

Dr. Bijaya Shrestha

ABSTRACT

Malignant melanoma is a very deadly form of skin cancer which has claimed many lives over the past few years. If detected early this can be cured, hence early detection of malignant melanoma is essential. Unfortunately melanoma is mimicked by seborrheic keratosis, a benign skin cancer. Identifying malignant melanoma as seborrheic keratosis using clinical diagnosis can prove fatal to the patient. To prevent such errors, dermoscopy, a common non-invasive skin imaging technique, is used which improves the diagnosis of these pigmented lesions by visualizing the morphological structures. This study proposes an automatic method by applying image processing techniques to aid in dermoscopy. The purpose of this study is to differentiate melanoma from seborrheic keratosis by applying thresholding techniques to the dermoscopy images. The algorithm consists of absolute thresholding of the red chromaticity plane and adaptive thresholding of the green and blue planes to detect inflamed keratin plugs in the images. The parameters for thresholding are obtained from histogram analysis. The images obtained after applying this technique are then processed to extract different features such as color and texture features. The information obtained from the feature extraction is given to a classifier to differentiate melanoma from seborrheic keratosis. The proposed algorithm is applied on a dataset consisting of 369 melanomas and 256 seborrheic keratoses. This method yielded 94.0% accuracy with 98.6% of melanomas correctly identified.

Keywords— Melanoma, Seborrheic Keratosis, Dermoscopy, Inflamed Keratin Plugs, Adaptive Thresholding, Absolute Thresholding, Histogram, Feature Extraction, Classifier

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Randy Moss for his constant guidance and encouragement in successfully accomplishing my degree; I would like to express my gratitude for his patience in helping me understand the technical concepts required for the research. I thank Dr.William Stoecker for his continuous inputs and ideas throughout the research, for explaining the research from the medical perspective and for being available all the time for guiding and analyzing the work in spite of his busy schedule. I would also like to express my gratefulness to Dr.Ronald J.Stanley and Dr.Bijaya Shrestha for their valuable contributions and suggestions in my research.

I would like to thank my fellow researcher Kavya SudhiReddy for her contribution and support throughout the research. I would like to thank Dr.Stanley for the relative color code. I would also like to thank Nabin Mishra and Jason Hagerty for their contribution and sharing their knowledge to understand the research from the engineering perspective.

I would like to thank my parents for their support and blessings through my life and for making me a better person today. I would like to express my gratitude to my friends for their constant encouragement throughout this journey.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

Malignant melanoma is the most dangerous form of skin cancer. Melanoma is a dark colored structure that begins in pigment cells called melanocytes which produce skin color. Melanoma exists as a mole which spreads and has an irregular border. It occurs mostly in white-skinned people, but can occur in people with all skin colors. Even though melanoma consists of only 1-2 % of skin cancer cases, deaths due to this type of skin cancer is around 75%. According to 2016 estimates, around 10,130 people in the U.S. are expected to die of melanoma [1]. Hence the early detection and correct identification of melanoma is essential.

Deeply pigmented seborrheic keratosis can mimic malignant melanoma, and these may be confused by laymen or by non-dermatologist physicians [2]. Unfortunately, “in the clinical diagnosis, seborrheic keratosis is one of the most common non melanoma diagnoses in retrospective studies, instead of the lesion being histologically confirmed melanomas” [3]. Seborrheic keratosis usually appears as a pale, black or brown growth on the back, shoulders chest or face, but can appear anywhere on the skin. It has a waxy appearance and is not painful**.** It is benign or non-malignant and doesn’t require treatment but is sometimes removed due to cosmetic reasons.

Studies show that the “accuracy of clinical diagnosis of melanoma by dermatologists varies between 49% and 81%, with approximately one third of melanomas being misdiagnosed as seborrheic keratosis” [4]. In another study, “out of the 9204 lesions which were diagnosed clinically as seborrheic keratosis in the differential diagnosis, 61 (0.66%) revealed melanoma on histological examination"[3].

Therefore distinguishing seborrheic keratosis (SK) from malignant melanoma (MM) clinically is difficult. Melanoma can be cured if detected early, but if it is identified as seborrheic keratosis wrongly, as seborrheic keratosis is a benign mimic of malignant melanoma, it can become fatal to the patient. Hence the primary focus in this study is to find a reliable method to distinguish these two diagnoses. Dermoscopy, a non-invasive skin imaging technique is used, which improves the diagnosis of these pigmented lesions by visualizing the morphological structures [5]. This study proposes an automatic method to differentiate melanoma from seborrheic keratosis by applying image processing techniques to aid in dermoscopic images.

Different methods have been applied in the past to classify melanoma lesions from seborrheic keratosis. In this research, thresholding techniques are applied on the lesions to discriminate them accurately. The model used in this research is depicted in the flowchart in Figure1.1.

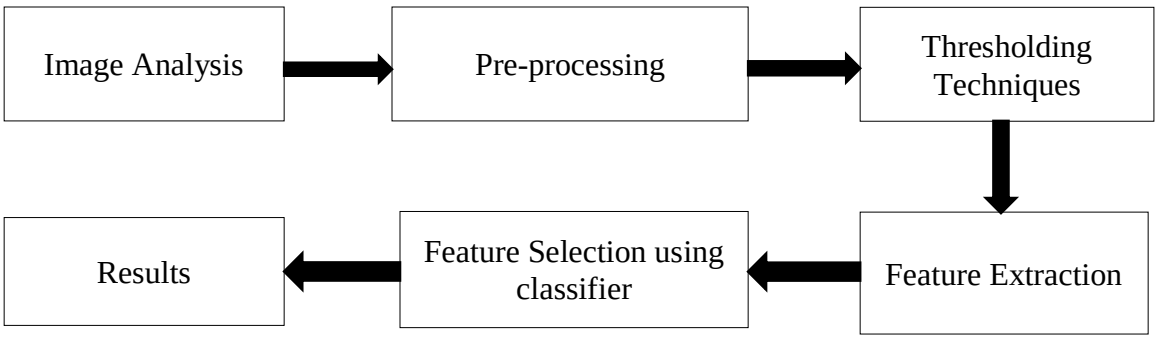

Figure 1.1: Flowchart of the model

The image analysis, preprocessing and thresholding techniques are discussed in Section 2. Section 3 describes the feature extraction step and discusses in detail the different features to be extracted from the image. The model used for selecting the features using a classifier to differentiate seborrheic keratosis from melanoma with high accuracy is described in Section 4. The results obtained from this model are discussed in Section 5.

2. METHODS

2.1. METHODOLOGY OVERVIEW

The following is the strategy involved in differentiating seborrheic keratosis from melanoma [6], although steps are not necessarily performed in the order listed.

1. Obtain a data set of melanoma and seborrheic keratosis images
2. Determine the lesion boundary of each image
3. Remove unwanted hair
4. Perform color and texture histogram analysis, generating a cumulative histogram of features over the images for the marked inflamed keratin plugs
5. Identify color characteristics of melanomas and seborrheic keratosis from cumulative histogram
6. Determine the threshold or cut off of selected color characteristics to identify inflamed keratin plugs for discriminating melanomas and seborrheic keratosis
7. Extract features for each image after applying thresholding technique.
8. Compute and select the features and its parameters that best separate the melanomas and seborrheic keratosis / best identifies the structures / identifies the structures with different intensities using forward stepwise technique.
9. Perform classification of the above acquired data using logistic regression analysis.

The above algorithm is explained in detail in the following sections.

2.2. INTRODUCTION TO INFLAMED KERATIN PLUGS

The classical description of seborrheic keratosis includes dark structures that clog the follicular openings, often described as keratin plugs. In practice, these are often in

areas of inflammation, resulting in an admixture to varying degrees with blood and serum (serosanguinous fluid), resulting in blood-tinged keratin plugs. For convenience in this paper, we refer to these structures as inflamed keratin plugs. Inflamed keratin plugs are about 0.1 mm in size occurring mostly in round or oval shapes. These plugs have color ranging from dark/light red to dark/light brown to light yellow. Inflamed keratin plugs are especially useful in discriminating seborrheic keratosis from melanoma, since melanoma of any type, invasive, in-situ or metastatic, doesn't contain inflamed keratin plugs. Since most inflamed keratin plugs are red, brown or yellow in color, the most important factor to check is the intensity of the pixels in the red plane, green plane and blue plane. An example of above description is attached below where Figure 2.1 describes the inflamed keratin plugs in seborrheic keratosis, Figure 2.2 describes no sign of inflamed keratin plugs in melanoma.

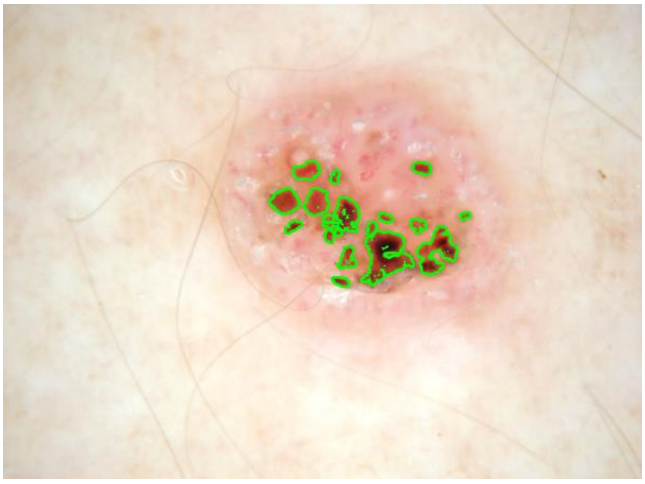

Figure 2.1: Seborrheic keratosis showing inflamed keratin plugs in green outline

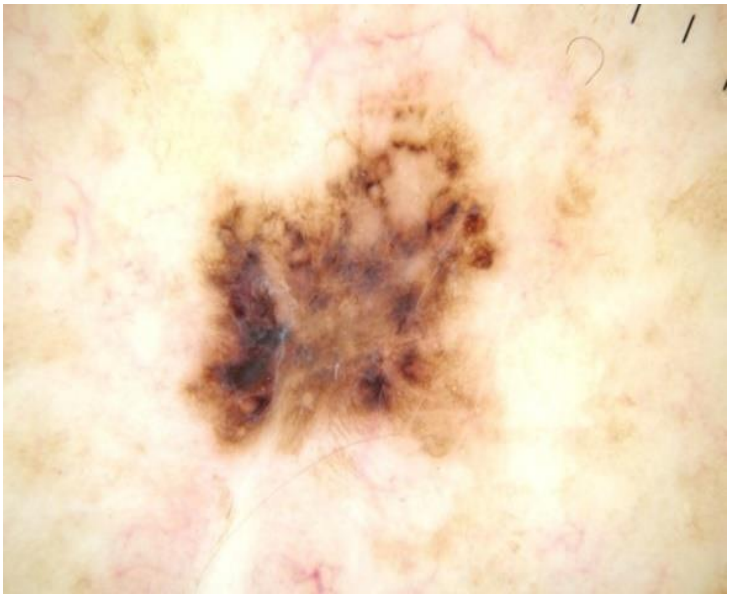

Figure 2.2: Melanoma showing no signs of inflamed keratin plugs

2.3. MATERIALS AND PREPROCESSING

2.3.1. Dataset and Equipment. For this study contact non-polarized dermoscopy images were used. To obtain these images a 3Gen DermLite DL2 dermatoscope (3Gen LLC, San Juan Capistrano, CA) was used. To visualize the lesion composed of superficial structures and deeper pigmentation, the device has 32 bright white LED lights with 10X magnification and a gel interface contact. The dataset for this study consists of contact dermoscopy images with 1024×768 resolution. A total of 625 images with 369 melanomas and 256 seborrheic keratoses were analyzed. These images were obtained from six private practice clinics namely Skin and Cancer Associates (Plantation, FL), The Dermatology Center (Rolla, MO), Columbia Dermatology (Columbia, MO); and Sheard and Drugge (Stamford, CT), from January 2007 to February 2010 per protocol for NH R44 CA-101639-

02A2 [7]. The study was reviewed and approved by the Phelps County Regional Medical Center Institutional Review Board. For the above dataset, histopathology results were obtained for all melanomas. All seborrheic keratoses which are not biopsied were determined to be seborrheic keratoses both clinically and dermoscopically [8].

2.3.2. Finding the Boundary of the Skin Lesion. All the clinical images with 1024 × 768 resolution have surrounding skin which is not part of the lesion. To delineate the surrounding skin from the lesion each image was segmented manually to detect the lesion boundary. For this purpose, borders of skin lesions are drawn manually usually a program developed here at Missouri University of Science and Technology called Winshow by choosing points along the border which are then joined by a second-order b-spline function. These generated borders are then confirmed by a dermatologist (W.V.S.) with 20 years of experience in dermoscopy [9].

2.3.3. Hair Mask. Most of the images contain hair. These can mimic the inflamed keratin plugs we want to identify in the lesion. To prevent this, a hair mask was generated from an automated hair detection algorithm (R. Kaur et al., in press, IEEE Trans Biomed Eng.) which detected long and thin hair areas. The hair mask pixels are represented as one, the mask is inverted making the hair pixels zero, this is multiplied to the border mask image (border mask consists of the lesion with the border represented as one and rest of the image represented as zero) to remove hair for the images containing hair before analysis occurs.

2.4. IMAGE SEGMENTATION USING THRESHOLDING

2.4.1. Finding Region of Interest (Inflamed Keratin Plugs). In order to find inflamed keratin plugs which are red in color and small in size we first look at the characteristics of the inflamed keratin plugs such as the red, green, and blue values of the pixels within inflamed keratin plugs. Inflamed keratin plugs, as the name suggests, contain small dense red structures. Some structures range from red to dark brown or yellow. Hence, the color which differentiates these structures primarily is red. We therefore started by looking at red chromaticity, but, for completeness, we looked at green and blue chromaticity, as well. The inflamed keratin plugs were manually marked (using the Winshow program) and confirmed by a dermatologist (W.V.S.) on a subset of 62 seborrheic keratosis images taken from the image set described earlier. The first column of Figure 2.3 shows the average red, green, and blue chromaticity with the red curves representing the areas of the manually marked inflamed keratin plugs and the green curves representing the remainder of the lesion area. As can be seen from the top left curve in the figure, in general, the red chromaticity of the inflamed keratin plugs is greater than the red chromaticity of the rest of the lesion. The second column of the figure shows the number of manually-marked plugs that are found (1) and the number missed (0) when an optimal threshold is set for red chromaticity (top), green chromaticity (middle) and blue chromaticity (bottom). Clearly red chromaticity works better than green or blue. The threshold used for red chromaticity is 0.547.

Various other features within the marked ROIs and within the remainder of the lesion such as average of red, green, blue, hue, saturation, value, their minimum values, maximum values, variance and standard deviation were also calculated. The relative values

(differences of pixel values inside the structure to that of surrounding skin) of red chromaticity, blue chromaticity and green chromaticity were also extracted.

Figure 2.4 shows the average red (top), green (middle), and blue (bottom) values with the red curves representing the areas of the manually marked inflamed keratin plugs and the green curves representing the remainder of the lesion area for the same set of 62 seborrheic keratosis images. Note that red chromaticity separates the inflamed keratin plugs much better than just the value of the red color plane. Also, note that the green and blue color planes separate the inflamed keratin plugs better than the corresponding chromaticities, with the inflamed keratin plugs having lower values of green and blue, in general, than the rest of the lesion. We therefore developed an adaptive algorithm for the green and blue planes to try to find some of the inflamed keratin plugs missed by the red chromaticity threshold. In this algorithm, a constant times the standard deviation of the given plane (green or blue) of the entire lesion is subtracted from the mean of the rest of that plane for the entire lesion. This number is then used as a threshold for the given plane, with those pixels below the threshold retained as inflamed keratin plug areas.

The middle column in Figure 2.3 is calculated by representing 1 to the pixel which are above the threshold value of 0.547 and representing them as 0 if the respective pixel value is below the threshold value. Based on these calculations 1 represents the number of inflamed keratin plugs correctly classified as inflamed keratin plugs and 0 represents the number of inflamed keratin plugs wrongly identified as rest of the lesion. The threshold condition for middle column in Figure 2.4 is an adaptive threshold. It is calculated by average of the lesion minus constant times the standard deviation of the lesion in the respective plane.

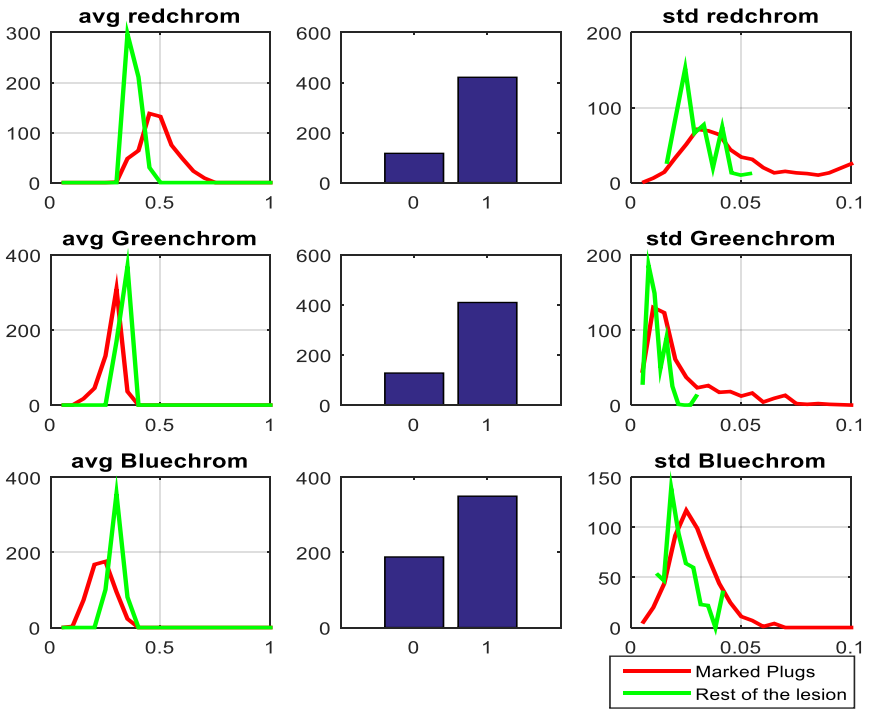

Figure 2.3: Histogram plot of average, standard deviation of RGB chromaticity

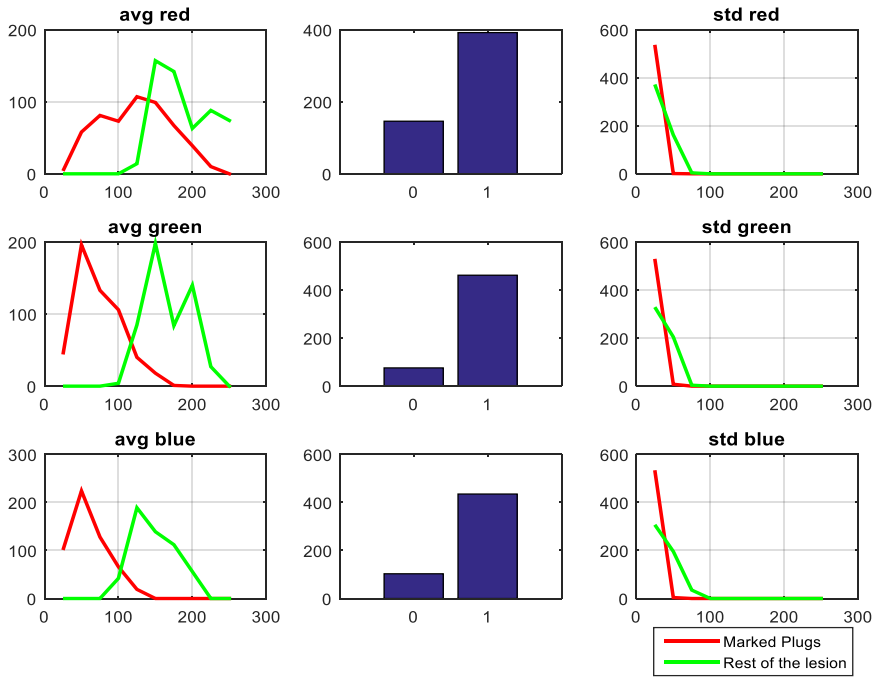

Figure 2.4: Histogram plot of average, standard deviation of RGB planes

2.4.2. Thresholding Algorithm. Summarizing the above, the algorithm to segment inflamed keratin plugs is as follows:

1. The color image is converted into R,G,and B image components.
2. The red chromaticity plane is extracted using the equation R_{chrom}= R/(R+G+B).
3. An absolute threshold (0.547) is applied to the red chromaticity plane of the image to create a binary image.
4. The green plane is thresholded with an adaptive threshold value to create a binary image.
5. The blue plane is thresholded with an adaptive threshold value to create a binary image.
6. The intersection (common areas) of the green and blue thresholded images is obtained by multiplying the two binary images.
7. The union of the resulting mask (binary image) and the red chromaticity mask is then found to incorporate all kinds of plugs – dark brown to red to yellow.

   The algorithm is discussed in detail in the following.

   Step 1: Absolute threshold of red chromaticity plane

The red chromaticity of the entire image is calculated using the formula Red_{chrom} = R/(R+G+B). From the histogram, after testing with various threshold values 0.547 was found to be the best value to use as the threshold value. This can be justified as most of the inflamed keratin plugs have almost the same red chromaticity value. All pixel values ≥ 0.547 on the unit scale are included in the KeratinPlug_Redchrom mask. The absolute chromaticity threshold proved superior to the relative chromaticity threshold, i.e., computing the relative color for each component, then replacing each of the three

components in the chromaticity equation above by the corresponding relative component, i.e. $R_{rel} = R_{rel}/(R_{rel}+G_{rel}+B_{rel})$, where R_{rel} is the red value of a pixel in the lesion minus the average red value of the surrounding skin, with G_{rel} and B_{rel} defined similarly. The use of relative color gave inferior results to the threshold described above for red chromaticity. Even though this step detected the inflamed keratin plugs which are red in color, plugs which are dark brown and yellow remain undetected. To improve the detection of dark brown and yellow inflamed keratin plugs, the green and blue planes are considered.

Step 2: Adaptive threshold of green plane

As the pixel values of plugs in the green plane is not as consistent as the red chromaticity plane, adaptive thresholding is used. In the histogram on the left in the second row of Figure 2.2, the red curve indicates the inflamed keratin plugs and the green curve indicates the pixel values of the rest of the image. For the adaptive threshold, the average green value of the lesion is calculated and the standard deviation of the lesion (in the green plane) is multiplied by a constant and subtracted. This value is considered as the threshold value for green plane. The constant k=0.05 was found to be optimal based on the 62-image set of seborrheic keratosis images described earlier.

Summarizing, Green threshold = Average green value – 0.05×standard deviation green value. A mask KeratinPlug_Green is created by using the above value as the threshold for the green plane and keeping those pixels below that threshold.

Step 3: Adaptive threshold of blue plane

As the pixel value of plugs in the blue plane is also not as consistent as the red chromaticity plane, adaptive thresholding is used. The average blue value of the lesion is calculated and a constant times the standard deviation of the blue value of the lesion is subtracted to create

a threshold. The constant k=0.05 is used for the blue plane as well. Again, this constant was found based on the same 62-image set.

Blue threshold = Average blue value – 0.05× standard deviation blue value

The mask KeratinPlug Blue is created by using the above value as the threshold applied to the blue plane and keeping those pixels below that threshold. However, the green threshold and blue threshold operations both resulted in some pixels which are not part of inflamed keratin plugs. Both of them had pixels from inflamed keratin plugs in common; hence an AND operation was used in order to find just the inflamed keratin plugs and eliminate areas which are not inflamed keratin plugs.

The output mask of this was ORed with the output mask of the red chromaticity threshold for better results of inflamed keratin plugs.

Hence the final equation is:

KeratinPlug Mask = KeratinPlug_Redchrom + (KeratinPlug_Green × KeratinPlug _Blue), where + represents the OR or union operation and × represents the AND or intersection operation. Figures 2.5 through 2.9 illustrate these operations on three sample images.

Once the final inflamed keratin plug mask from the above algorithm is obtained, several features are calculated to differentiate seborrheic keratoses from melanomas.

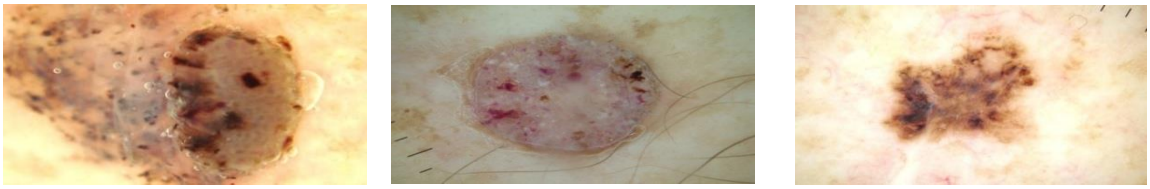

Figure 2.5: Original Images

Figure 2.6: Red chromaticity mask

Figure 2.7: Green mask

Figure 2.8: Blue mask

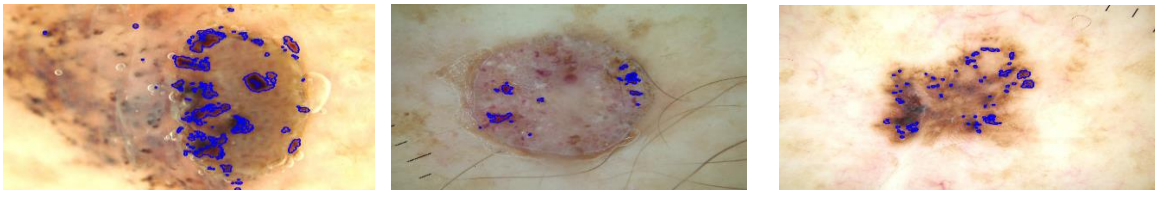

Figure 2.9: Final overlay

(a) (b) (c)

(a) , (b) Seborrheic Keratosis (c) Melanoma

3. FEATURE EXTRACTION

The process of extracting certain characteristics and generating a set of features from an image is known as feature extraction. The purpose of this study in computer aided diagnosis is to extract various features from a given skin image which characterize the skin lesion as benign or malignant [10]. There are various methods to extract features from the skin lesion images which analyze different components like texture, shape, color and dermoscopic features such as globules or blotches. The following features were used in this study to differentiate seborrheic keratosis from melanoma accurately:

1.) Color Features
2.) Texture Features
3.) Relative Color Features
4.) Blotch Features
5.) Demographic Features
6.) Global Features/Lesion Features

3.1. COLOR FEATURES

Color is an important feature used in representing an image and it is the most intuitive feature upon perception of an image. The key components of color feature extraction are the color space, color moments and similarity measurement. The features extracted are maximum value and minimum value of color planes and color moments. Color moments are the measures which describe the distribution of color in an image. The mean, variance and standard deviation of a color plane are considered as color moments [11]. These features are calculated for the regions in the image obtained after segmentation

is done, i.e., from the inflamed keratin plug mask. From the final inflamed keratin plug mask, two color spaces RGB and HSV are extracted. Since HSV consists of two color components (hue and saturation) and a brightness component (value), this color space is used for extraction of features. “H indicates the wavelength of the color if it would be monochromatic. S indicates the amount of white color mixed with monochromatic color” [12]. The transformation of the RGB planes to the chromaticity planes contributes to a normalized color representation which distinguishes illumination of the image from the color. The red, green and blue chromaticity are calculated. For each color space, the color moments and the minimum and maximum color pixel values over the inflamed keratin plug mask are calculated.

The color planes used in extracting features are shown in Table 3.1. The original image is represented as ‘Image’ in the table for better understanding. The image set consists of images with the RGB color model. Hence there is no particular conversion required for the image to get the RGB color model. The red, green and blue plane components are extracted from the original image by giving different coordinates.

Table 3.2 shows the features that are extracted from each color plane for each color component. Hence we have 9 color components and five features for each component, giving a total of 45 color features generated.

All the features listed below are calculated for the inflamed keratin plugs identified by the algorithm described in section 2.The features which are extracted and selected by the model are described in Appendix. Table 3.1 describes the various planes used in extracting the features whereas Table 3.2 describes the features extracted from those planes.

Table 3.1: Different color planes used to extract color features [13]

Color Plane	Conversion of Image
RGB Plane	**Red Plane(R) = Image(:,:,1)** **Green Plane(G) = Image(:,:,2)** **Blue Plane(B) = Image(:,:,3)**
HSV Plane (converting RGB plane to HSV plane) [12]	V = [max$(\boldsymbol{R},\boldsymbol{G},\boldsymbol{B})$] and X=[$\mathbf{min}(\boldsymbol{R},\boldsymbol{G},\boldsymbol{B})$] S = $\begin{cases}\frac{V-X}{V} & ; \quad \boldsymbol{if\ S=0\ return};\end{cases}$ r = $\frac{V-R}{V-X}$, g = $\frac{V-G}{V-X}$, b = $\frac{V-B}{V-X}$ $\boldsymbol{if\ R=V\ then\ H}$ $= (\boldsymbol{if\ G=X\ then\ 5+b\ else\ 1-g})$ $\boldsymbol{if\ G=V\ then\ H} = (\boldsymbol{if\ B=X\ then\ 1+r\ else\ 3-b})$ $\boldsymbol{else\ H} = (\boldsymbol{if\ R=X\ then\ 3+g\ else\ 5-r})$
Chromaticity Plane (converting RGB plane to chromaticity planes)	**Red Chromaticity = R/(R+G+B)** **Green Chromaticity = G/(R+G+B)** **Blue Chromaticity = B/(R+G+B)**

Table 3.2: Statistical measures to be calculated from different planes

Features Extracted	Description	Equations
Average	mean value of the pixels in all the plugs over the lesion	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$
Standard deviation	describes the contrast or spread in the data	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$
Variance	Square of standard deviation	σ^2
Minimum value of plug	smallest value of all the pixels in all the plugs over the lesion	Minimum(x_i) i=1,2…N
Maximum value of plug	largest value of all the pixels in all the plugs over the lesion	Maximum(x_i) i=1,2…N

3.2. TEXTURE FEATURES

The appearance or consistency of a surface or an object is defined as texture. It is related to the smoothness or roughness of an object. Numerous approaches have been proposed to determine image texture. A classical method to measure texture characteristics is to use a set of second-order statistical features which are derived based on the joint

probability distribution of local gray levels using a second-order histogram technique [14]. This technique, also known as the gray level co-occurrence matrix (GLCM) method, is derived from the statistics of pixel pairs and their respective gray levels. The GLCM is a two-dimensional array of data of how often specific pixel pairs with specific spatial relationships occur in an image. The features derived from the GLCM depend on two parameters: distance and angle. The distance is measured as the pixel distance between the pixel pairs in a given direction. The angle indicates the angle between the pixel pairs. The default angles correspond to horizontal (0°), vertical (90°), and diagonal directions (45°, 135°). Based on these parameters, four basic statistical features are obtained. They are defined as shown in Table 3.3.p(i,j) in Table 3.3 is defined as probability of occurrence of pixels having i,j adjacent to each other.

The features shown in Table 3.3 are calculated for the red plane, green plane, blue plane and intensity or gray-scale plane. The intensity plane is obtained by the following equation applied on each pixel in the whole image.

Intensity plane = (Red plane + Green plane + Blue plane) / 3

The six features in Table 3.4 are calculated for the lesion mask using the histogram technique as described in [16]. Here r_i is a random variable indicating the gray level, $p(r_i)$ is the probability of occurrence of gray level r_i. In other words, $p(r_i)$ is the histogram of the intensity levels in a region. L is the number of possible gray levels.

These ten GLCM features are given to the model and the features selected by the model are described in Appendix. The features in Table 3.3 are calculated using gray level co-occurrence matrix while as features in Table 3.4 are calculated using histograms.

Table 3.3: GLCM features [15]

Features	Description	Equations
Contrast	"Measures the local variations in the gray-level co-occurrence matrix"	$\sum_{i,j} \lvert i-j \rvert^2 p(i,j)$
Energy	"Provides the sum of squared elements in the GLCM. Also known as uniformity or the angular second moment"	$\sum_{i,j} \{(p(i,j)\}^2$
Correlation	"Measures the joint probability occurrence of the specified pixel pairs"	$\frac{\sum_{i,j}(ij)p(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y}$
Homogeneity	"Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal"	$\sum_{i,j} \frac{p(i,j)}{1+\lvert i-j \rvert}$

Table 3.4: Features calculated on lesion mask [16]

Feature	Description	Equations
Mean	Average of pixels in the plugs with their probability	$\mu = \sum_{i=0}^{L-1} r_i p(r_i)$
Variance	Average contrast	$\sigma^2 = \sum_{i=0}^{L-1} (r_i - \mu)^2 p(r_i)$
Smoothness	S measures the relative smoothness of the gray level in a region	$S = 1 - \frac{1}{(1 + \sigma^2)}$
Skewness	Measures the skewness of a histogram	$\kappa = \sum_{i=0}^{L-1} (r_i - \mu)^3 p(r_i)$
Entropy	Measure of randomness	$e = -\sum_{i=0}^{L-1} P(r_i) log_2 p(r_i)$
Uniformity	Energy or angular second moment	$U = \sum_{i=0}^{L-1} p^2(r_i)$

3.3. RELATIVE COLOR FEATURES

Absolute color measures in some cases do not give similar results under different conditions such as lightning, cameras, and specific color conditions whereas relative color measures the color with respect to the background making these features robust [10]. Hence

relative color features are considered for the classification of melanoma and seborrheic keratosis.

Relative color is defined as differences of pixel values inside the structure to that of the average color of the surrounding skin. The skin color surrounding the lesion is determined by first finding the surrounding pixels. The pixels are determined/detected using the Euclidean distance method which calculates the distance of each pixel present outside the lesion from the lesion boundary. The pixels which are less than the distance D (D is increased from 1 until the total area outside the lesion reaches 4×Area of the lesion) and satisfies RGB constraints, such as R>G and G> B, to assure the pixel falls into the surrounding skin color category [17].

From the surrounding skin color mask, the following parameters are calculated: Average color of the skin for three color components (R, G and B) and their chromaticity are discussed in detail in Table 3.5. From the original image, the average value for R, G and B color components are calculated for each pixel in the lesion from the border mask to be subtracted to get the relative parameters. The respective relative features are calculated for the average of the R, G and B color components and the chromaticities as discussed below. The general definition of calculating relative for each pixel used here is difference between the respective lesion pixel value and the average surrounding skin color:

Relative = Lesion color – average skin color

In Table 3.5 R_{les},G_{les},B_{les} represents each pixel in the red, green and blue plane of the lesion respectively, R_{skin},G_{skin},B_{skin} represents each pixel in the red, green and blue plane of the skin respectively, N represents the number of pixels within the lesion and M represents the number of pixels in the surrounding skin.

Table 3.5: Relative color features

Features	Description	Equation
Average lesion color (R, G and B)	Mean of all the pixel values present in the lesion	$\mu_r = \frac{\sum_{i=1}^{N} R_{les}}{N}$ $\mu_g = \frac{\sum_{i=1}^{N} G_{les}}{N}$ $\mu_b = \frac{\sum_{i=1}^{N} B_{les}}{N}$
Lesion chromaticity (R,G and B)	Chromaticity of the lesion for three planes	Red chromaticity of lesion $Red_{chrom_{les}} = \frac{R_{les}}{(R_{les}+G_{les}+B_{les})}$ Green chromaticity of lesion $Green_{chrom_{les}} = \frac{G_{les}}{(R_{les}+G_{les}+B_{les})}$ Blue chromaticity of lesion $Blue_{chrom_{les}} = \frac{B_{les}}{(R_{les}+G_{les}+B_{les})}$
Average skin color (R, G and B)	Mean of all the surrounding skin pixels	$\mu_{rskin} = \frac{\sum_{i=1}^{M} R_{skin}}{M}$ $\mu_{gskin} = \frac{\sum_{i=1}^{M} G_{skin}}{M}$ $\mu_{bskin} = \frac{\sum_{i=1}^{M} B_{skin}}{M}$

Table 3.5: Relative color features (contd)

Features	Description	Equation
Relative color (R, G and B)	Difference between the pixel value in the lesion and average surrounding skin	Relative red color $R_{rel} = (R_{les} - \mu_{rskin})$ Relative green color $G_{rel} = (G_{les} - \mu_{gskin})$ Relative blue color $B_{rel} = (B_{les} - \mu_{bskin})$
Relative chromaticity	Chromaticity of the relative image for three planes	Red chromaticity of relative color image $Red_{chrom_{rel}} = \frac{R_{rel}}{(R_{rel}+G_{rel}+B_{rel})}$ Green chromaticity of relative color image $Green_{chrom_{rel}} = \frac{G_{rel}}{(R_{rel}+G_{rel}+B_{rel})}$ Blue chromaticity of relative color image $Blue_{chrom_{rel}} = \frac{B_{rel}}{(R_{rel}+G_{rel}+B_{rel})}$

3.4. DEMOGRAPHIC FEATURES

Melanoma and seborrheic keratosis have different characteristics such as the type and size of the lesion structures and its prevalence in humans of particular ages, as discussed in the introduction part of this paper. Hence, demographic features are useful in discriminating melanoma from seborrheic keratosis. Demographic features used in this study include patient age, gender, patient concern about the lesion, patient noticed change in the lesion, personal history of melanoma, family history of melanoma, lesion size, lesion location on the body (quantized), and a binary location about the lesion clinic (within 23.5 degrees of the equator or not). These features are taken from the database obtained from different clinics as discussed in the dataset section of this paper.

3.5. GLOBAL FEATURES / LESION FEATURES

The same color and texture features as discussed in Sections 3.1 and 3.2 are calculated for the whole lesion instead of just the detected inflamed keratin plugs.

3.6. BLOTCH FEATURES

Characteristics of the detected plugs in each image can be helpful in detecting the types of lesion. Hence features such as the number of plugs found in the image, eccentricity or relative size are defined.

From the plug mask, the following are calculated [18].

1. Area of the largest plug within the lesion ($Area_{max}$)
2. Number of plugs detected within the lesion (N)
3. Eccentricity measures the degree to which the largest plug is eccentric in location.

$$E = \text{Distance} / \sqrt{Area}$$

4. Relative size (R) = $\frac{\sum_{i=1}^{N} B_i}{\text{Area}}$,
5. Relative size index of largest plug = $\frac{Area_{max}}{Area}$
6. Irregularity of largest plug = $\frac{Perimeter_{max}}{Area_{max}}$

where B_i = Area of the i^{th} plug,

Distance represents the Euclidean distance between the largest plug centroid and the lesion centroid,

Area = Area of the lesion, and

$Perimeter_{max}$ = Perimeter of the largest plug

4. FEATURE SELECTION USING A CLASSIFIER

Feature selection is a process of selecting the various features that enhance lesion discrimination. In this research, the goal is to differentiate seborrheic keratoses from melanomas with maximum accuracy. SAS performs a logistic regression with forward-stepwise feature selection. So logistic regression is the prediction/machine-learning/classification algorithm and the stepwise option performs the feature selection. This study implements the stepwise logistic regression model in SAS. The model is discussed below in detail.

4.1. SAS

Statistical Analysis System,or SAS is a software tool developed by the SAS institute used for advanced analytics, business intelligence, data management and statistical analysis [19]. In this research SAS was utilized for both feature selection and creation of a classification model. A logistic regression model was chosen for classification and feature selection was performed using a forward stepwise selection method.

4.2. STEPWISE LOGISTIC REGRESSION MODEL

The stepwise logistic regression model was used to both automatically determine the most significant features and best model using the selected feature when evaluated using the receiver operating characteristic value. Logistic regression is a linear model defined as

$$f(x) = \frac{1}{\left(1+e^{-z(x)}\right)}$$

where z(x) is defined as

$$z(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_N x_N$$

and x is the feature vector. One could simply use the feature vector in its entirety and determine the betas that are used to define the model. While this is an accepted method, there is a direct relationship between the size of the feature vector and the complexity of the resultant model. A forward stepwise selection method was used in an effort to reduce the complexity of the model and as a guide to indicate what type of features (shape, color or texture for example) the research should focus on.

The forward stepwise selection method is an iterative method that consists of two basic stages, selection and evaluation. The most less complex model is first created by determining the intercept β_0.The data is then analysed to determine the most significant feature based on that feature's *p*-value and add that feature to the model. In the evaluation stage, the new model determines whether the addition of the newly added feature, as well as all previously selected features, are significant to the model. If any of the features are deemed insignificant, they are removed from the model and the process repeats. This continue until the best model is selected.

The forward logistic regression model implemented in SAS is controlled by two parameters, SLENTRY and SLSTAY. SLENTRY represents the minimum *p*-value required for a feature to enter the model and SLSTAY represents the minimum *p*-value for a feature to stay in the model. The values of SLENTRY and SLSTAY do not necessarily have to be the same value, setting up scenarios that a feature must be really significant to enter the model (a numerically lower SLENTRY value), but once entered in the model, it

can stay in the model even though its significance, when compared to the other feature in the model, may be lower (a numerically high SLSTAY value). The model was run for different values of SLENTRY and SLSTAY ranging from 0.1 to 0.5.The model yielded best results for SLENTRY and SLSTAY = 0.25.The model f(x) gives the probability of the outcome. The extracted features and products of some features (to model interactions between features) are input to the above model and the significant features are chosen using stepwise logistic regression.The lesions are classified as melanoma and seborrheic keratosis depending on the value of f(x) in the final model.

5. RESULTS

The SAS logistic regression is performed on the data set of 625 images containing 256 seborrheic keratoses and 369 melanomas. A total of 133 features are analyzed and operated upon by the model. The model gave the best results with 71 features including the interactions between several features. All the features are described in Appendix with selected features highlighted. The most significant features as ranked by high chi-square value, are the location of the lesion, standard deviation of the green plane of the plugs, size of the lesion and interaction between size of the lesion and location of the lesion, standard deviation of the blue plane of the plugs, variance of red chromaticity plane of plugs, minimum Hue value of plugs, relative red chromaticity of plugs, minimum and maximum S value of plugs. Demographic features play an important role in maximizing the accuracy.

In this research melanomas are defined as type 1 and seborrheic keratoses are defined as type 0. Sensitivity is a parameter ranging 0-1 which is defined as the number of melanomas which are correctly classified as melanoma among 369 melanoma images, i.e. the proportion of melanomas correctly identified. Specificity is a parameter which is defined as number of lesions which are correctly classified as seborrheic keratosis among 256 seborrheic keratosis images, i.e. the proportion of seborrheic keratoses correction identified.

To analyze the obtained results, a Receiver Operating Characteristic (ROC) is plotted as the graph of sensitivity or true positive fraction (TP/ (TP+FN)) vs. 1 – specificity (true negative fraction (TN/ (TN+FP)) [19]. The pro logistic parameter ‘c’ which is defined as area under the curve is 0.94 in this experiment. Hence in this model, a lesion is correctly classified as melanoma with 94.0% accuracy (estimated area under the ROC curve) with

sensitivity (Melanoma = 1) being 0.986 and 1 - specificity = 0.429 at one point on the curve. Thus the final model does well in distinguishing melanoma from seborrheic keratosis mimicking melanoma.

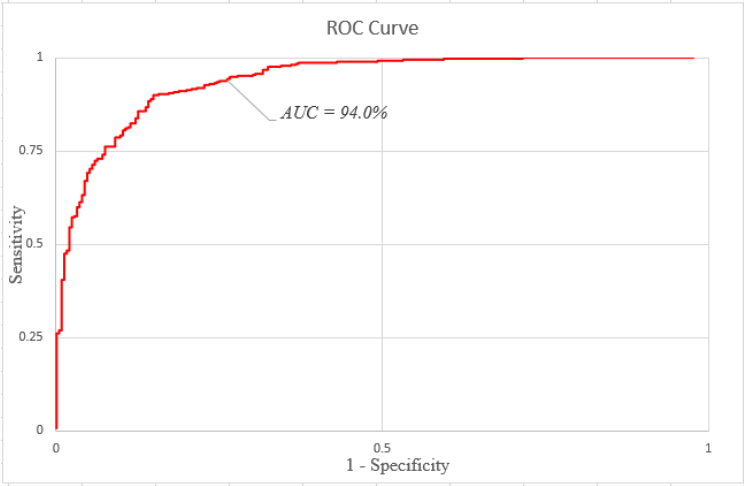

Figure 5.1: ROC curve with AUC = 94.0%

6. FUTURE WORK

To further improve the above model, characteristics which are unique to either melanoma or seborrheic keratosis must be identified. Structures known as cloudy milia-like cysts (MLCs) are present, almost without exception, only in seborrheic keratoses. Hence, cloudy MLCs can be useful in maximizing the number of true positives. If these structures are detected, their features can be added to the logistic regression model to further increase the accuracy.

APPENDIX

The table below provides the feature number and the feature description, the features from 1 to 133 are given to the model. The final model consists of interactions between these features along with most significant features. The features selected by the model are represented in bold.

Table: Features given to the SAS, along with the interactions between the features selected by SAS

Plug features	Plug features	Plug features	Rest of the Lesion features	Rest of the Lesion features
1.Average of red plane	**32. Average of S plane**	63. Entropy	85. Average of red plane	**116. Maximum of S plane**
2. Standard deviation of red plane	33. Minimum of S plane	**64. Average lesion color of red plane**	86. Standard deviation of red plane	**117. Average of V plane**
3. Maximum of red plane	**34. Maximum of S plane**	65. Average lesion color of green plane	87. Maximum of red plane	118. Minimum of V plane

Table: Features given to the SAS, along with the interactions between the features selected by SAS (contd)

Plug features	Plug features	Plug features	Rest of the Lesion features	Rest of the Lesion features
4. Minimum of red plane	35. Average of V plane	66. Average lesion color of blue plane	88. Minimum of red plane	**119. Maximum of V plane**
5. Variance of red plane	36. Minimum of V plane	67. Average skin color of red plane	89. Variance of red plane	120. Average contrast of red plane
6. Average of green plane	37. Maximum of V plane	68. Average skin color of green plane	90. Average of green plane	**121. Average correlation of red plane**
7. Standard deviation of green plane	38. Average contrast of red plane	**69. Average skin color of blue plane**	**91. Standard deviation of green plane**	**122. Average energy of red plane**
8. Maximum of green plane	39. Average correlation of red plane	**70. Relative color of red plane**	92. Maximum of green plane	123. Average homogeneity of red plane
9. Minimum of green plane	**40. Average energy of red plane**	71. Relative color of green plane	93. Minimum of green plane	124. Average contrast of green plane

Table: Features given to the SAS, along with the interactions between the features selected by SAS (contd)

Plug features	Plug features	Plug features	Rest of the Lesion features	Rest of the Lesion features
10. Variance of green plane	**41. Average homogeneity of red plane**	72. Relative color of blue plane	94. Variance of green plane	125. Average correlation of green plane
11. Average of blue plane	**42. Average contrast of green plane**	**73. Relative color of red chromaticity plane**	95. Average of blue plane	**126. Average energy of green plane**
12. Standard deviation of blue plane	43. Average correlation of green plane	**74. Relative color of green chromaticity plane**	**96. Standard deviation of blue plane**	**127. Average homogeneity of green plane**
13. Maximum of blue plane	**44. Average energy of green plane**	75. Relative color of blue chromaticity plane	**97. Maximum of blue plane**	128. Average contrast of blue plane
14. Minimum of blue plane	45. Average homogeneity of green plane	**76. Irregularity of the plug**	**98. Minimum of blue plane**	129. Average correlation of blue plane

Table: Features given to the SAS, along with the interactions between the features selected by SAS (contd)

Plug features	Plug features	Plug features	Rest of the Lesion features	Rest of the Lesion features
15. Variance of green plane	**46. Average contrast of blue plane**	**77. Area of the largest plug**	**99. Variance of blue plane**	**130. Average energy of blue plane**
16. Average of red chromaticity plane	47. Average correlation of blue plane	78. Number of plugs	**100. Average of red chromaticity plane**	**131. Average homogeneity of blue plane**
17. Standard deviation of red chromaticity plane	48. Average energy of blue plane	79. Eccentricity	**101. Standard deviation of red chromaticity plane**	132. Average contrast of intensity plane
18. Variance of red chromaticity plane	49. Average homogeneity of blue plane	80. Relative size	102. Variance of red chromaticity plane	**133. Average correlation of intensity plane**

Table: Features given to the SAS, along with the interactions between the features selected by SAS (contd)

Plug features	Plug features	Plug features	Rest of the Lesion features	Rest of the Lesion features
19. Average of green chromaticity plane	50. Average contrast of intensity plane	81. Personal history of melanoma	103. Average of green chromaticity plane	**134. product of feature no. 26 and 29**
20. Standard deviation of green chromaticity plane	51. Average correlation of intensity plane	82. Patient age	**104. Standard deviation of green chromaticity plane**	**135. product of feature no. 20 and 96**
21. Variance of green chromaticity plane	**52. Average energy of intensity plane**	**83. Gender**	**105. Variance of green chromaticity plane**	**136. product of feature no. 20 and 100**
22. Average of blue chromaticity plane	53. Average homogeneity of intensity plane	84. Patient concern about the lesion	106. Average of blue chromaticity plane	**137. product of feature no. 131 and 29**

Table: Features given to the SAS, along with the interactions between the features selected by SAS (contd)

Plug features	Plug features	Plug features	Rest of the Lesion features	Rest of the Lesion features
23. Standard deviation of blue chromaticity plane	**54. Energy**		**107. Standard deviation of blue chromaticity plane**	**138. product of feature no. 97 and 100**
24. Variance of blue chromaticity plane	**55. Lesion location**		108. Variance of blue chromaticity plane	**139. product of feature no. 26 and 100**
25. Average of H plane	56. Family history of melanoma		**109. Average of H plane**	**140. product of feature no. 26 and 96**
26. Lesion size	**57. Relative green chromaticity of the lesion**		**110. Minimum of H plane**	**141. product of feature no. 26 and 96**

Table: Features given to the SAS, along with the interactions between the features selected by SAS (contd)

Plug features	Plug features	Plug features	Rest of the Lesion features	Rest of the Lesion features
27. Patient noticed change in the lesion	58. Relative blue chromaticity of the lesion		111. Maximum of H plane	**142. product of feature no. 29 and 97**
28. Relative red chromaticity of the lesion	**59. Variance**		112. Average energy of intensity plane	**144. product of feature no. 20 and 26**
29. Binary location of the clinic	60. Smoothness		113. Average homogeneity of intensity plane	**145. product of feature no. 20, 26 and 100**
Plug features	Plug features	Plug features	Rest of the Lesion features	Rest of the Lesion features
30. Minimum of H plane	61. Uniformity		114. Average of S plane	**145. product of feature no. 29, 97 and 100**
31. Maximum of H plane	62. Skewness		**115. Minimum of S plane**	**145. product of feature no. 26, 29 and 100**

BIBLIOGRAPHY

[1] R.L. Siegel, et al., "Cancer statistics, 2016," *CA Cancer J Clinicians,* vol. 66, no. 1, pp. 7–30, 2016.

[2] C. Hafner and T. Vogt, "Seborrheic keratosis," *JDDG J der Deutschen Dermatologischen Gesellschaft,* vol. 6, no. 8, pp. 664-677, 2008.

[3] L. Izikson, et al., "Prevalence of melanoma clinically resembling seborrheic keratosis: Analysis of 9204 cases," *Archives of Dermatol,* vol. 138, no. 12, pp. 1562-1566, 2002.

[4] S. V. Deshabhoina, et al., "Melanoma and seborrheic keratosis differentiation using texture features," *Skin Research and Technology,* vol. 9, no. 4, pp. 348-356, 2003.

[5] M. Anantha, et al., "Detection of pigment network in dermatoscopy images using texture analysis," *Comput Med Imaging Graph,* vol. 28, no. 5, pp. 225-34, Jul 2004.

[6] Y. Faziloglu, et al., "Colour histogram analysis for melanoma discrimination in clinical images," *Skin Research and Technology,* vol. 9, no. 2, pp. 147-156, 2003.

[7] S. M. Stricklin, et al., "Melanoma in situ in a private practice setting 2005–2009: location, lesion size, lack of concern," *J Am Acad Dermatol,* vol. 67, no. 3, pp. e105-9, Sep 2012.

[8] S. M. Stricklin, et al., "Cloudy and starry milia-like cysts: how well do they distinguish seborrheic keratoses from malignant melanomas?," *Journal of the European Academy of Dermatology and Venereology,* vol. 25, no. 10, pp. 1222-1224, 2011.

[9] S. E. Umbaugh, *Digital Image Processing and Analysis: Human and Computer Vision Applications with CVIPtools, Second Edition*: CRC Press, Inc., 2010.

[10] M. E. Celebi, et al., *Dermoscopy Image Analysis,* CRC Press, Inc., 2016

[11] S. R. Kodituwakku, S. Selvarajah, "Comparison of color features for image retrieval," *Indian Journal of Computer Science and Engineering,* vol. 1, no. 3, pp. 207-211, 2010.

[12] V. Arvis, et al., "Generalization of the cooccurrence matrix for colour images: application to colour texture classification," *Image Anal Stereol,* vol. 23, no.1, pp. 63-72, 2004.

[13] A. R. Smith, "Color gamut transform pairs," *SIGGRAPH Comput. Graph.,* vol. 12, no.3, pp. 12-19, 1978.

[14] H. B. Kekre., et al., "Image retrieval using texture features extracted from GLCM LBG and KPE,"*International Journal of Computer Theory and Engineering,* vol. 2, no. 5, pp. 695-700, 2010

[15] http://www.mathworks.com/help/images/gray-level-co-occurrence-matrix-glcm.html

[16] R. M. Haralick, et al., "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. SMC-3, no. 6, pp. 610-621, 1973.

[17] J. Chen, et al., "Colour analysis of skin lesion regions for melanoma discrimination in clinical images," *Skin Research and Technology,* vol. 9, no. 2, pp. 94-104, 2003.

[18] W. V. Stoecker, et al., "Detection of asymmetric blotches (asymmetric structureless areas) in dermoscopy images of malignant melanoma using relative color," *Skin Res Technol,* vol. 11, no. 3, pp. 179-84, Aug 2005.

[19] R. J. Stanley, et al., "A relative color approach to color discrimination for malignant melanoma detection in dermoscopy images," *Skin Research and Technology,* vol. 13, no. 1, pp. 62-72, 2007.

VITA

Pramada Kishtagari was born in Hyderabad, India. In May 2014, she received a Bachelor of Science degree in Electronics and Telematics from G.Narayanamma Institute of Technology and Science affiliated to Jawaharlal Nehru Technological University, Hyderabad, India. She was admitted into Missouri University of Science and Technology in August 2014 where she worked as a research assistant under Dr. Randy H. Moss on the project 'Thresholding Techniques for discrimination of melanoma from seborrheic keratosis. In June 2016 she received her MS degree in Electrical Engineering from Missouri University of Science and Technology, Rolla, Missouri, USA.