2012

Automatic detection of annular-granular patterns in melanoma *in situ* dermoscopy images

Parivash Hajiyani

Department:

AUTOMATIC DETECTION OF ANNULAR-GRANULAR PATTERNS IN MELANOMA *IN SITU* DERMOSCOPY IMAGES

By

PARIVASH HAJIYANI

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

2012

Approved by

Randy H. Moss, Advisor
R. Joe Stanley
Steven Grant

ABSTRACT

Early detection of malignant melanoma greatly benefits patients, as the overall success is dependent on finding these melanomas before they reach the invasive stage. Dermoscopy is a non-invasive skin imaging technique that studies have shown can improve the diagnostic accuracy of dermatologists by as much as 30% over clinical examination. In this project machine vision and image analysis techniques are used to detect annular granular areas in dermoscopy images automatically. The proposed algorithm utilizes the luminance ratio between annular and granular areas within the darkest 30% of the lesion. All points whose luminance value are less than 30% of the histogram are considered for further processing. The method has used some preprocessing steps to remove the unwanted effect of luminance reflection, to extract hair and bubble from the lesion image and to enhance the contrast of the image. Then the lesion plane is searched to find the center and border of annular-granular areas. Statistical analysis has shown that the implemented algorithm has the highest 92 percent in correct detection of annular granular areas.

ACKNOWLEDGMENTS

I would like to express my gratitude to all those who gave me the possibility to complete this thesis and finish my study. I am deeply indebted to my advisor Dr. Moss and co-advisor Dr. Stoecker, whose help, suggestions and encouragement helped me in this research and the process of writing this thesis.

I take this opportunity to thank Dr. Stoecker for his support and medical explanations of the images, for his patience in manual markings of the features and the analysis of the results; I thank Dr. Stanley, for his valuable suggestions and constant helpful comments. I would like to thank Dr.Grant, my committee member, for his time, guidance and commitment. I also would like to show my appreciation towards Dr. Ferdowsi for his constant support, motivation and encouragement throughout the period of my graduate study. I am also obligated to my colleagues and fellow classmates for being a constant source of inspiration throughout my graduate study period.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

1.1 SKIN CANCER

Skin cancer is the most common type of cancer in the United States. The risk factors known for skin cancer include the following: lighter complexion, genetic predisposition, advanced age, increased sun exposure and high frequency of sunburns [1]. There are several types of skin cancer:

-Basal cell carcinoma: Forms in the lower part of the epidermis, which comprises the outermost layer of the skin [2]. Basal cell carcinoma accounts for more than 75% of all skin cancers. It is a slow-growing cancer that seldom spreads to other parts of body. If left untreated, however, it can spread to nearby areas and invade bone and other tissues under the skin, causing significant morbidity [3].

-Squamous cell carcinoma: Forms in the squamous cells, which are flat cells that form the most superficial layer of the skin [1]. This type is less common than basal cell carcinoma. It is more likely to grow deep below the skin and spread to distant parts of the body [3].

-Neuroendocrine carcinoma: Forms in neuroendocrine cells, which can release hormones in response to signals from the nervous system [2].These three types of skin cancer collectively comprise a vast majority of non-melanoma skin cancers. When diagnosed early, there is a low reoccurrence rate, depending upon which therapy is prescribed. The other predominant type of skin cancer is melanoma.

The word melanoma comes from the Greek words; *melas* (black) and *oma* (tumor) [4]. Melanoma forms in the melanocytes, which are cells that make pigment, especially in the skin. Malignant melanoma is the most serious type of skin cancer. A less common skin cancer, melanoma has a low survival rate, which is heavily dependent upon its invasiveness (99% 5-year survival rate for Stage I, 33% 5-year survival for Stage IV). It is a very aggressive tumor that can spread quickly throughout the body, leading to death. If caught at an early stage, melanoma can often be cured with a simple excision.

Estimated cases by the National Cancer Institute show that more than one million new cases of non-melanoma skin cancer were diagnosed and less than one thousand deaths were caused by non-melanoma skin cancer in 2010. While melanoma accounts for only 5% of skin cancers, it is important because it is the cause of 75% of all skin cancer deaths [2].

Each year, more than 68,000 Americans are diagnosed with melanoma, and another 48,000 are diagnosed with an early form of the disease that involves only the top layer of skin (melanoma *in situ*). Also, more than 2 million people are treated for basal cell or squamous cell skin cancer each year, making basal cell carcinoma several times more common than squamous cell skin cancer [2]. There are four major types of melanoma [5]:

Superficial spreading melanoma: The most common type of melanoma. It is usually flat and irregular in shape and color, with different shades of black and brown. It is most common in Caucasians [6].

Nodular Melanoma: Usually starts as a raised area that is dark blackish-blue or bluish-red. However, some do not have any color. [6]

Lentigo maligna melanoma: Usually occurs in the elderly. It is most common in sun-damaged skin on the face, neck, and arms. The abnormal skin areas are usually large, flat, and tan with areas of brown. [6]

Acral lentiginous melanoma: The least common form of melanoma. It usually occurs on palm, soles, or under nails and is more common in African-Americans [6].

Melanomas have key features to discriminate them from other skin lesions. These symptoms have been codified into the ABCDE system. The ABCDE system can help to recognize the signs of melanoma and also serve to distinguish between melanomas and moles [3, 4]:

Asymmetry: One half of the lesion is different from the other half as depicted in Figure 1.1.

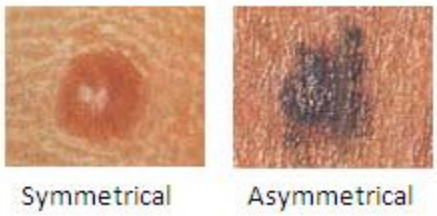


Figure 1.1: Asymmetry characteristic [20]

Borders: Irregularity of borders is present. The border may be uneven, fuzzy and have notched or scalloped edges and the edges of the growth are irregular. It shows in Figure 1.2.

Color: Color changes from one area to another (variegated), often with shades of tan, brown, or black, and sometimes white, red, or blue. A mixture of colors may appear within one lesion as shown in Figure 1.3.

Diameter: The lesion is usually larger than 6mm in diameter and grows larger than normal moles; about the size of a pencil eraser as depicted in Figure 1.4.

Evolution: The mole keeps changing appearance. The beginning of an increase in thickness of a mole, freckle, blemish, or birthmark, even if the increase is small, often signifies a lesion that is entering a dangerous phase as illustrated in Figure 1.5.

Figure 1.2: Uneven border [20]

Figure 1.3 : Color changes within the area [20]

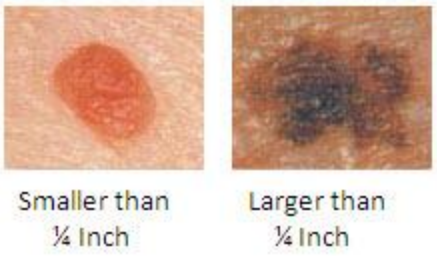


Figure 1.4 : Larger than ¼ inch in diameter [20]

Figure 1.5 : The mole keeps changing in appearance [20]

Elevation changes (invasiveness) are important because, when the thickness of melanoma exceeds 1mm, about 1/25th of an inch, the melanoma is graded at Stage IV and the chance of metastasis increases. When the thickness reaches 3mm, the cure rate is only about 50% and quickly decreases as the thickness increases further [4]. However, if caught at early stage, melanoma can often be cured with a simple excision. Early detection of malignant melanoma, therefore, can significantly reduce mortality.

1.2 DERMOSCOPY IMAGES

Early detection of malignant melanoma greatly benefits patients, as the overall success is dependent on finding these melanomas before they reach the invasive stage. Dermoscopy was introduced to improve accuracy in the diagnosis of melanoma.

Dermoscopy, also known as epiluminescence microscopy, is a non-invasive skin imaging technique that has been shown to be effective for the diagnosis of melanoma and other pigmented of skin lesions [7, 8]. The procedure requires illumination at a low angle of incidence and a liquid contact medium applied at the skin-microscope interface. Dermoscopy uses optical magnification and, either liquid immersion and low angle-of-incident lighting, or cross-polarized lighting, making surface structure more easily visible compared to a conventional clinical image. This technique allows the incident light to penetrate the top layer of skin tissue and permits a detailed examination of the pigmented structure beyond what would be visible to the naked eye. Studies have shown that dermoscopy can improve the diagnostic accuracy of dermatologists by as much as 30% over clinical examination [9]. Dermoscopy allows identification of dozens of morphological features, such as pigment network, dots/globules, streaks, blue-white areas and blotches, which aid in the diagnosis of skin lesions [10]. Due to the difficulty and subjectivity of human interpretation, automated analysis of dermoscopy images has become an important research area for automated skin cancer detection and the development of computerized image analysis techniques to minimize diagnostic error.

1.3 ANNULAR GRANULAR FEATURE IN MELANOMA IN SITU

Lentigo maligna is a type of melanoma *in situ* predominantly seen in flat, pigmented of lesions located on the face. Early melanoma on the face is often difficult to distinguish from solar lentigo and seborrheic keratosis. Early features of facial melanoma are:

-An annular pattern with light brown, dark brown and gray color. They may be asymmetrical follicular openings, focal structure-less areas, or thickened or broad pigment around follicles.

-A reticular pattern with light brown, dark brown or gray color. They may be focal structure-less areas.

The presence of black or blue-gray dots/granules, either localized or diffused, is a clue to diagnosis of melanoma on the face. Localized black or blue-gray dots/granules may be seen in association with either the annular or reticular pattern. Black or black-gray dots/granules coalesce to form clumps of gray pigment. The clumps of gray pigment gradually extend into streaks and ultimately take shape as rhomboidal structures. The formation of pigment continues to expand around hair follicles and gradually obscures the follicular opening [4].

The annular granular feature is often characterized by three regions: 1) an inner ring that can be either a dark or light region and is referred to here as a "don't care" region, 2) a second ring that has high luminance and is referred to as the “annular” region, and 3) an outer ring that has low luminance and is referred to as the “granular” region [4]. Figure 1.6 graphically depicts the three regions of an annular-granular area in a melanoma *in situ*. Figure 1.7 shows dermoscopy images where annular-granular regions in a melanoma, identified by a dermatologist, are shown in black circles.

This project is an extension of the work performed by Nishat Nepal [4]. The proposed algorithm utilizes the luminance ratio between annular and granular areas. Dermoscopy images have been degraded by luminance reflection in them that affects the

contrast and resolution of the images. Thus, a contrast enhancement procedure is added to the algorithm. A hair and bubble removal function created by Thomas Szalapski is used as a pre-processing step in this project. To improve the efficiency of the algorithm, the two largest convex-hull within the darkest 30% areas of a lesion are considered for further processing. Darkest 30% of the lesion is calculated according to histogram of the luminance and Red, Green, Blue planes and all points whose value are less than 30% of the histogram of the related plane are considered for further processing.

The method for finding the center of an annular-granular melanoma feature is changed from Nepal's thesis and adaptive thresholds are added to the algorithm to remove as many false positive points as possible. False positive points are considered as annular-granular regions by the algorithm, when, in actuality, while they are not.

Machine vision and image analysis techniques are used to detect annular granular features in dermoscopy images automatically. The rest of the thesis is organized as follows: Section 2 concentrates on the algorithm of automatic detection of annular-granular regions and explains the entire operation and steps needed to extract the annular-granular areas from dermoscopy skin lesion images, Section 3 discusses some statistical tests examined based on results from the algorithm, and Section 4 presents the results achieved by this automatic annular-granular detection and determine its success on different images.

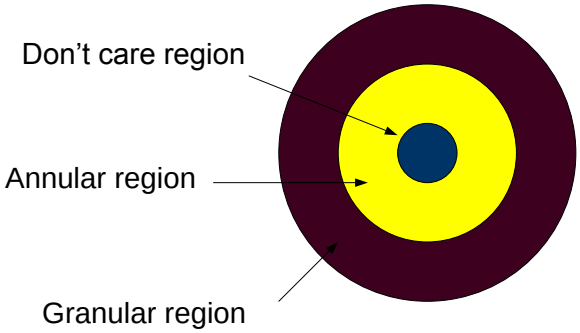


Figure 1.6: Three regions of annular granular feature

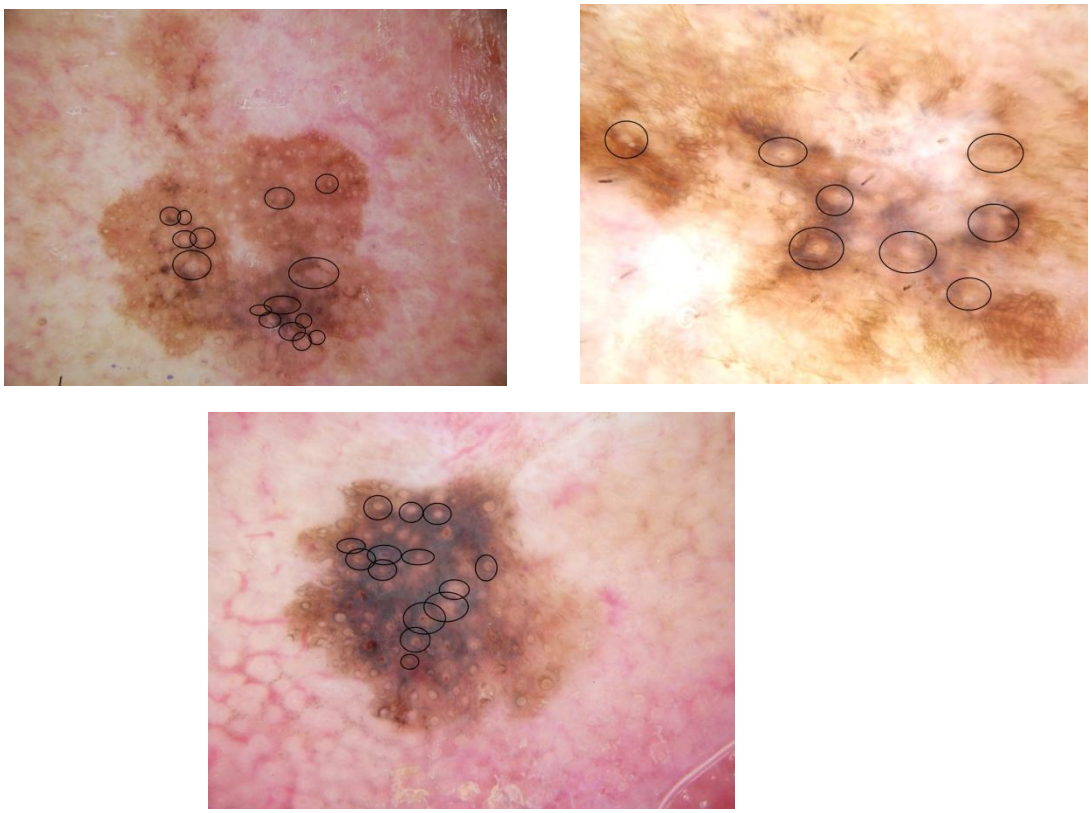


Figure 1.7: Annular-Granular melanoma in dermoscopy images

The features to be analyzed have been saved in a series of Excel files for each image to be interpreted by a software program called Statistical Analysis Software (SAS Institute, Cary, NC). SAS analysis chooses the best features to discriminate annular-granular features using a logistic regression model. The results obtained in this work, as well as suggestions for future work, are explained in Section 4.

A set of 20 dermoscopy images are used in this research to attempt automatic identification of annular-granular features for melanoma *in situ* diagnosis. As mentioned before, these images have some effects of luminance reflection in them that degrade the image effectiveness. To resolve the unwanted reflection in the images, pre-processing steps are needed before the determination of candidate points for annular-granular areas can be made. In the following chapter, the pre-processing steps are discussed and the results have been shown in sample images. The remainder of the processing algorithm is discussed below.

2. AUTOMATIC DETECTION ALGORITHM FOR ANNULAR GRANULAR AREAS

2.1 OVERVIEW

Dermoscopy is an imaging technique used in diagnosis of melanoma and other types of pigmented skin lesion. Automatic skin cancer detection via analysis of dermoscopy images is an important research area due to the importance of early diagnosis of melanoma. Automation using machine vision is crucial because of the difficulty and subjectivity of human interpretation. In addition, accurate color information in dermoscopy images is very important for melanoma diagnosis, since inappropriate white balance or brightness in the images adversely affect the diagnostic performance.

Demarcating the inner annular area is the first step in the automatic detection of an annular-granular region. The detection algorithm is based on luminance comparison between the annular and granular areas. The difference between the average luminance of the annular region, L_{Anul} and the average luminance of the granular region, L_{Granul} is used to determine the pixels in the annular region (Equation 2.1). If this difference is greater than zero, the pixel is a candidate for the annular region.

To implement this method on the dermoscopy images, some pre-processing steps are needed. These steps include contrast enhancement, hair and bubble removal and finding the darkest 30% area of the lesion, as annular-granular regions are found particularly in this area.

$$L_{Rsult} = L_{Anul} - L_{Granul}$$
$$L_{Anul} = \frac{\sum \text{pixel value of inner annular ring}}{\text{Area of Annular ring}}$$
$$L_{Granul} = \frac{\sum \text{pixel value of outer granular ring}}{\text{Area of Granular ring}} \quad (2.1)$$

2.2 FIRST STEP - CONTRAST ENHANCEMENT

Dermoscopy images have insufficient contrast, making the determination of a threshold between light area and dark area of the image difficult. A pre-processing step is therefore needed to enhance the contrast of dermoscopy images. Contrast enhancement techniques are widely used for image processing to achieve wider dynamic range. Histogram-based techniques are often used for image enhancement. One of the advantages of histogram-based techniques is the simplicity in implementation. Histogram-based techniques are based on two primary methods: histogram stretch and histogram equalization, which try to increase the dynamic range of the image.

In this project, the histogram stretch method is used to increase the image contrast and achieve the enhancement needed for processing. Because annular-granular regions do not exist in the darkest and lightest areas of a lesion, the lowest 5% and the highest 5% of each Red, Green, and Blue plane histogram are omitted. The remaining histogram region is stretched by a factor of 150*(highest luminance - lowest luminance). Figure 2.1 shows the result of this contrast enhancement method for a dermoscopy image where Figure 2.1 a shows the original image and Figure 2.1 b) shows the resultant image using the histogram stretch method.

a)

b)

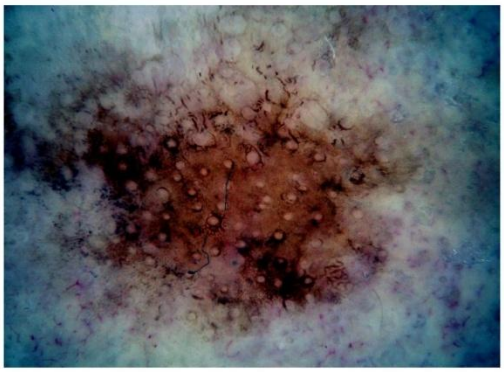


Figure 2.1 Histogram enhancement , a) Original image , b) Resultant image: Histogram stretch method

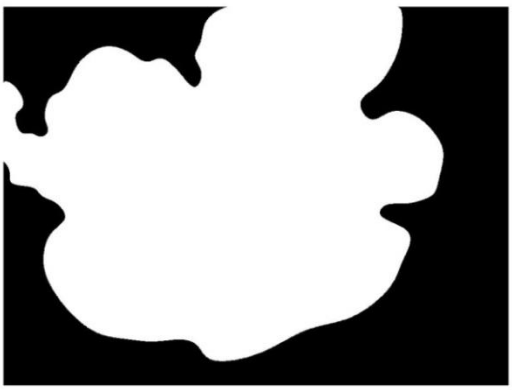
2.3 SECOND STEP - HAIR AND BUBBLE REMOVAL

The existence of hair and bubble in a dermoscopy image can cause a significant error in the detection processing, thus a method is needed to exclude these from the skin lesion mask. To solve this problem, a function written by Thomas Szalapski, is used that returns a mask identifying the location of the hair and bubble in the lesion. The returned

mask is a binary mask with pixels having a value of one where hairs and bubbles are present.

Using the exclusive-or logical operation between the lesion mask and the produced hair-bubble mask, a new mask is generated for further processing. This new mask omits hair and bubble areas.Figure 2.2 (c) shows the outcome lesion mask for a sample dermoscopy image.

a)



b)

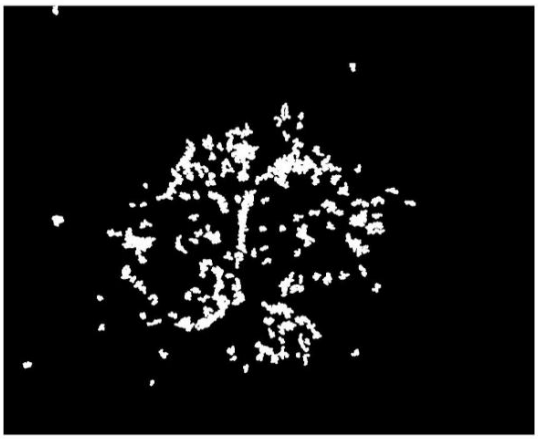


Figure 2.2: Mask output a) Lesion mask b) Hair-Bubble mask c) Resultant mask

c)

Figure 2.2: Mask outputs a) Lesion mask b) Hair-Bubble mask c) Resultant mask.(cont.)

2.4 THIRD STEP - FINDING DARKEST 30 PERCENT OF THE LESION

Most of the annular-granular regions lie in the darkest part of lesion; it is therefore more convenient to locate the darkest part of the lesion for further processing. In this algorithm, the darkest 30% of the image is considered for further processing. The digital image is separated into its red (R), green (G), blue (B), and luminance (L) planes. The luminance value (L) at pixel location (x, y) is defined as:

$$L(x, y) = 0.2989R(x, y) + 0.5870G(x, y) + 0.1140B(x, y) \quad (2\text{-}2)$$

Each plane is filtered by a median filter with a 51×51 window size to remove noise, thereby making the image smoother. The points laying in the darkest 30% the histogram of each plane are then identified. These points are passed to a function in

MATLAB which uses the BWLABEL command to label and classify the points. Those connected points with the same label are bounded by the smallest convex polygon called the convex hull. The two largest convex polygons among the four planes are considered as the darkest 30% lesion mask. Figure 2.3 shows the final darkest 30% mask area related to a sample image. Finally, the logical OR operation is used to create one lesion mask from these two depicted masks.

a) b)

Figure 2.3: Two largest polygons a) First largest polygon from the RGBI planes b) Second largest polygons from the RGBI planes

2.5 FOURTH STEP - OBTAINING ANNULAR AREA CANDIDATE POINTS

Annular-granular melanoma areas consist of three regions: 1) the “don't care” area, 2) the annular ring and 3) the granular ring. Based on previous work and observations, it was estimated that the typical radius of annular-granular regions could be approximated to 3 pixels for the “don't care” region, 4 pixels for the annular region and 3 pixels for the granular region. However, some melanoma images have larger annular-granular regions, also. For these regions, a larger set of radii was used: 5 pixels radius for

the “don't care” region, 12 pixels for the annular region and 5 pixels radius for the granular region.

Some images have both thicker annular-granular regions and less thick annular-granular regions, which are better determined by using the smaller radius set. For these images, both sets of radii are used to find the candidate points. The radius set is chosen according to the size of the darkest 30% lesion mask.

According to the lesion mask size, two sets of radii approximated for the annular-granular regions. These radius sets were determined by examining different dermoscopy images [4]. The small radius set considers a radius of 3 pixels for the “don't care” region, 4 pixels for the annular region, and 3 pixels for the granular region. The large radius set consists of 5 pixels for the “don't care” region, 12 pixels for the annular region and 5 pixels for the granular region.

Table 2.1 shows the relationship between the lesion mask size and the set of radius used to determine the candidate points.

Table 2.1: Lesion mask size versus radius set used for processing

Lesion mask area	Radius set chosen
1 to 56000 Pixels	Small radius set
56001 to 13000 pixels	Large and small radius set
Greater than 13000	Large radius set

Candidate points for the annular region were obtained by using an appropriate set of radii based on the size of the darkest 30% lesion mask. Each pixel that lies in the darkest 30% lesion mask is considered as a potential annular center pixel, and the sum of the luminance of the pixels that fell in the inner annular ring was calculated. Similarly, the sum of luminance of the pixels in its outer granular ring was calculated. Those pixels whose difference between the luminance of the annular region and the granular region (Equation 2-1) according to the radius set are greater than zero ($L_{Rsult} > 0$) will be considered as candidate points for annular region centers.

After finding the candidate center points for annular regions, it became apparent that some of these points were redundant and needed to be removed. First, we sorted the identified points based on maximum resultant intensity. Resultant intensity, L_{Rsult}, is the difference of the luminance between the annular region and the granular region. The points with maximum resultant intensity are kept for the second step. We next considered the points with maximum resultant intensity and calculated the distance of the points with other selected points then removed the points existing within the distance of annular-granular outer radius. The remaining points constitute the annular region.

2.6 FIFTH STEP – FINDING ANNULAR REGION BORDERS

The candidate annular-granular center points were obtained according to the radius set and lesion mask size. In this step, new radii for the annular region and the granular region are estimated according to the data extracted from the previous step. To find the border of the annular region, the standard deviation within a specified window size around each candidate point was calculated to determine a threshold for the annular

region. By examining different sizes of window for calculation of the standard deviation within an annular region, it was found that the 31×31 window produced the best result.

For the subsequent step, all points lying within a distance of 30 pixels in the north direction were considered and the first drop in luminance was observed. If the drop was: 1) greater than the calculated standard deviation around the center point, and 2) maintained for at least 5 pixels, then this point was considered as the start of the granular region.

This process was repeated for the other cardinal directions: north, south, east, and west using a 61×61 square window.

For each candidate point, at most 4 points were considered for the beginning of the granular area. Only those that had fallen in luminance in at least 3 directions were considered for further processing; all others were omitted. The remaining points, with their respective border points, were used to update the annular region radius.

2.7 SIXTH STEP - FINDING THE CENTER OF THE ANNULAR REGION

The annular border points, plus the candidate annular center point, were used to improve the estimate for the annular-granular center location. To determine the center of annular-granular area, it was assumed that the set of four border points lay on the perimeter of a circle. Thus, finding a circle equation that goes through the border points can map the center of an annular-granular region. In the case that four border points were found, we can designate X to be the location of the center of annular-granular area that lies halfway on the line that goes through the east and west border points. We can similarly designate Y to be the location of the center of annular-granular area as the

halfway point on the line that goes though the north and south border points. The coordinate (X, Y) would then form the center of the annular region. In the case that 3 border points are found, the center of the circle that goes through these three points is found. This method is described below. Figure **2.4** shows a picture of the circle that has only 3 border points and the calculation of its center

Suppose that P1, P2 and P3 are three border points of one annular candidate point. Consider the line *a*, which goes through P1 and P2, and the line *b*, which goes through P2 and P3. The slopes of the lines *a* and *b* are used to find the center of the circle. The circle center is the intersection of two bisector lines which are perpendicular to lines *a* and *b*. The slopes of the lines *a* and *b* can be found by the following equations:

$$m_a = \frac{y_2 - y_1}{x_2 - x_1} \; and \;\; m_b = \frac{y_3 - y_2}{x_3 - x_2} \tag{2-3}$$

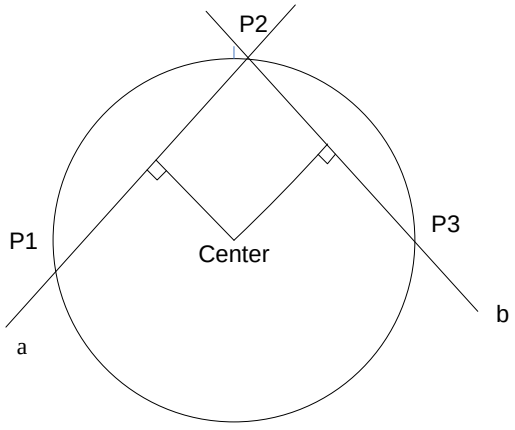


Figure 2.4: Drawing a circle using three border points

So the X coordinate of the circle center is defined by equation(2-4),and the Y coordinate of the center is defined by substituting the X value in one of the following line equations(2-5):

$$x = \frac{m_a m_b (y_1 - y_3) + m_b (x_1 + x_2) - m_a (x_2 + x_3)}{2(m_b - m_a)} \tag{2-4}$$

$$\begin{aligned} y_a' &= -\frac{1}{m_a}\left(x - \frac{x_1 + x_2}{2} \right) + \frac{y_1 + y_2}{2} \\ y_b' &= -\frac{1}{m_b}\left(x - \frac{x_2 + x_3}{2} \right) + \frac{y_2 + y_3}{2} \end{aligned} \tag{2-5}$$

By this method, the center of the annular-granular region was determined and the circle radius is the annular region radius.

After the new centers were obtained, the candidate points that lay within a small distance, less than outer radius, converged into one pixel. To further remove points that lay within a small distance less than outer radius, a process similar to Section 2-5 was used. The sum of the luminance of the inner annular ring was calculated and weighted by its area; the same process was performed for the granular ring. The difference of these two weighted sums was then calculated. The center points having a small difference and that lay within an update radius distance of other center points were omitted.

Consequently the “don't care” regions, the annular radii and granular radii were calculated for the remaining points. The area of the “don't care” region was considered to be 20% of the area created by update radius:

$$A_{don'tcare} = 0.2\left(\pi R_{update}^2\right) \tag{2-6}$$

As a result the radius of the don't care region was calculated to be $R_{dont'care} = \sqrt{0.2}R_{update}$. The area of the outer ring was taken to be 150% of the total area of the circle having a radius of R_{update}. The area of the annular-granular region can be written $A_{gr} = 1.5\left(\pi R_{update}^2\right)$. Hence the new radius of the granular ring was calculated to be $R_{gr} = \sqrt{1.5}R_{update}$. At the end of these processing steps, the number of false positive points was still high and additional thresholding was required to remove these unwanted points.

2.8 SEVENTH STEP - REMOVING FALSE POSITIVE CENTER POINTS

After removing the points that lay within the update radius of other center points, there were still points that lay in the dark area and they were not annular-granular center points. To solve this, an appropriate threshold should be set to remove extra candidate points that exist in the dark region.

The histogram of the image has an approximately normal distribution; this property was used to determine a suitable threshold. According to the normal distribution property, 68 percent of the points lay within the one standard deviation distance from the mean value **(**Figure 2.5).

Since annular-granular regions do not lie in the very dark and the very light areas of the histogram, a proper threshold can be set using the mean value of the lesion area and a determined fraction of the standard deviation of the lesion area. By comparing

different threshold values, the final threshold for removing extra false positive points was set to:

$$\text{Threshold} = \text{Average\_lesion luminance plane} + \ldots \\ (-0.8) \times \text{standard deviation of lesion luminance plane} \quad (2\text{-}7)$$

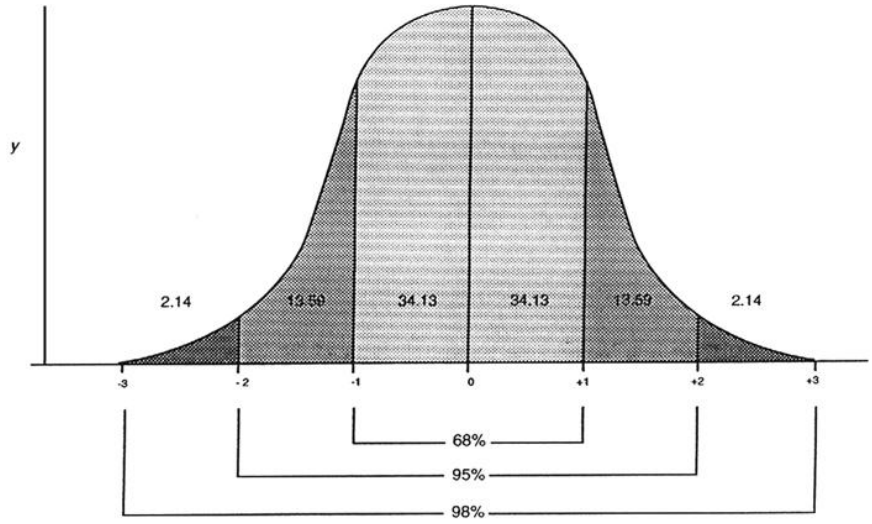


Figure 2.5: Normal distribution probability density function [19]

By using this threshold value, more false positive center points lying in the dark area of the lesion were removed.

2.9 EIGHTH STEP - CALCULATING NEW ANNULAR BORDER POINTS

The remaining center points were used to find the annular border points. The Principal Component Transform (PCT) was used to create an image with higher contrast and the resulting image was used to find the border points. A PCT searches for c k-

dimensional orthogonal vectors that can best be used to represent the data, where $c \leq k$. The original data are thus projected onto a much smaller space. PCT transform was used to find the plane of maximum possible variance.

The basic procedure is as follows:

The input data are normalized, so that each attribute falls within the same range. This ensures that attributes with large domains will not dominate attributes with smaller domains.

PCA computes c orthogonal vectors that provide a basis for the normalized input data. These are unit vectors so that each points in a direction perpendicular to the others. These vectors are called the principal components. The input data are a linear combination of the principal components.

The principal components are sorted in order of decreasing "significance", or strength. The principal components essentially serve as a new set of axes for the data, providing important information about variance. The axes are sorted such that the first axis shows the most variance among the data, the second axis shows the next highest variance, and so on.

Since the components are sorted according to decreasing order of "significance"; the size of the data can be reduced by not calculating the features that comprise the weaker components (those with low variance). Using the strongest principal components, it should be possible to reconstruct a reliable approximation of the original data [8].

The remaining center points, plus a neighborhood of 25×25 windows around each of them, were used to calculate the PCT coefficients. The PCT was then used on the R, G and B planes of the selected set of points and the PCT coefficient were calculated. The

largest first three eigenvalues of the correlation matrix of the selected pixels set were considered as PCT coefficients. These PCT coefficients are normalized, such that their summation is equal to one.

After calculating the PCT transform of the image, the algorithm looked for a drop in the gray level of the pixels around each center point. The threshold used for finding the drop is 15 gray level points in the luminance plane. The algorithm searches in eight directions from each center point outwards for this drop. The points that met the threshold were considered to be the annular border points. Figure 2.6 shows the PCT transform of the image and the identified border and center points. The yellow plus signs are the border of the annular areas and the red plus signs are the centers of the annular–granular regions.

a)

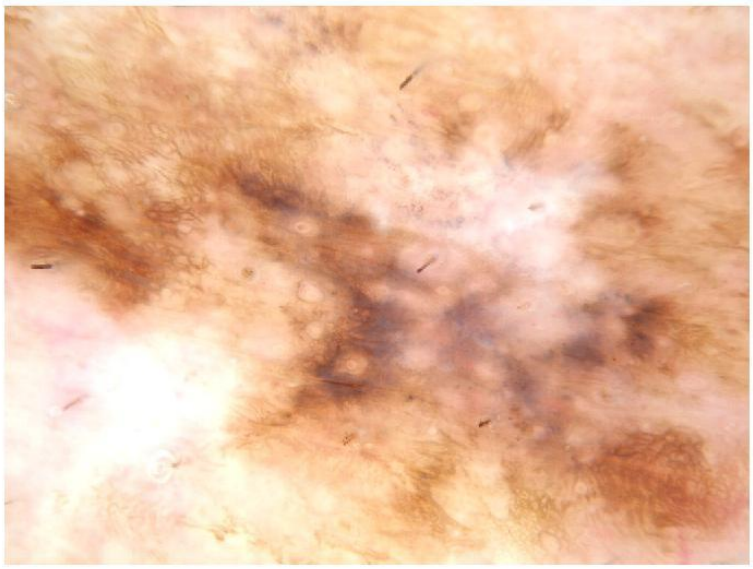
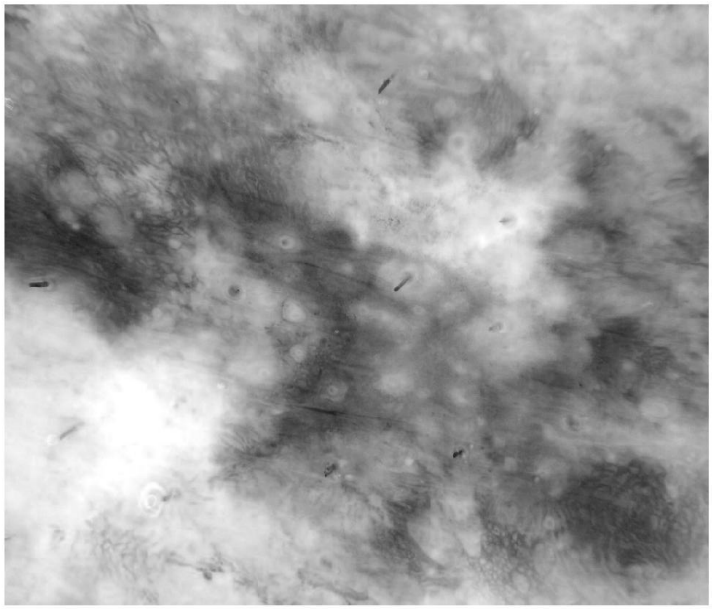


Figure 2.6: Output Images a) Original image b) PCT transform image c) Detection algorithm result

b)



c)

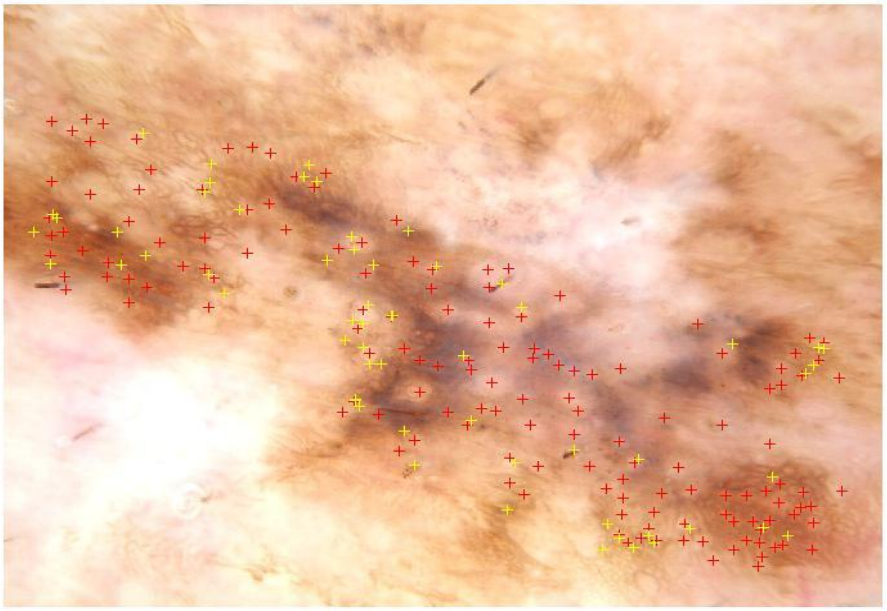


Figure 2.6: Output Images a) Original image b) PCT transform image c) Detection algorithm result.(cont.)

3. STATISTICAL STUDY OF THE ALGORITHM RESULT

3.1 OVERVIEW

The proposed algorithm has been tested on a set of twenty melanoma *in situ* lesions in dermoscopy images to detect annular granular regions. Annular granular regions were defined manually for six images in this set, then a list of features has been considered for statistical testing. The features measured in these six images are as follows:

1) Average values of the red, green, and blue planes of the lesion area (features c1—c3)
2) Variance of the red plane within the lesion area, since red plane has more variation within the lesion area compared to the blue and green planes, making it more useful for annular granular detection (feature c4)
3) Red, green and blue values of the automatically determined annular granular melanoma centers (red-cross points) determined by the algorithm. (Features c5—c7)
4) Number of yellow border points found for each center (red-cross point). (Feature c8)
5) Red, green and blue values of each automatically determined yellow border point. (Features c9—c32)
6) Shortest distance between yellow border point and corresponding red center point. (Feature c33)
7) Average of the distance between yellow border points and corresponding red center point. (Feature c34)
8) Standard deviations of the distance between yellow border points and corresponding red center point. (Feature c35)

9) Computed new radii for annular granular regions. (Feature c36)
10) Means of red, green and blue values in the annular areas. (Features c37-c39).
11) Standard deviations of red, green and blue values in the annular areas. (Feature c40-c42)
12) Means of red, green and blue values in the granular areas. (Features c43-c45)
13) Standard deviations of red, green and blue values in the granular areas. (Features c46-48)
14) Whether or not the red-cross center point is defined manually as annular granular by a dermatologist

These features for each of the six selected images were analyzed by the Statistical Analysis Software (SAS) using a logistic regression model.

Logistic regression, also called a logit model, is a statistical method used for prediction of the probability of occurrence of an event based on the fitting of the data to the logistic function. There are many cases when that a binary or ordinal response is needed based on the observed variables. In this kind of problem, logistic regression analysis is a useful tool to investigate the relationship between the discrete response and a set of variables.

The binary response can get two values: zero or one. The response probability of output Y=1 can be written as p=P(Y=1|x) condition on the explanatory data. In other words, predicting the likelihood that Y=1 given certain values of x can be estimated by logistic regression. Logistic regression is used to transform an S-shape curve into an approximately straight line, thus changing the range from zero to one to $-\infty$ to $+\infty$. The linear logistic model has the following equation:

$$\text{logit}(p) \equiv \log(\frac{p}{1-p}) = \alpha + \beta' x \qquad (3\text{-}1)$$

Where α is the intercept parameter and β' is the vector of slope parameter. So the logit function is defined as the natural logarithm of the odds of event A, where *p* is the probability of event A. The odds of event is defined by $odds = \frac{p}{1-p}$.

The logistic regression model is part of a category called generalized linear models, where a function $g = g(\mu)$ of the mean of the response variable is assumed to be linearly related to explanatory variables [11, 12]. The mean value depends on the stochastic behavior of the response and the observation variables are assumed to be fixed, thus the function *g* provides the link between the random stochastic component and the systematic component of the binary or ordinal response. For this reason, $g(\mu)$ is known as the link function.

The LOGISTIC procedure of SAS software has been used to fit the linear logistic model for binary response data by the method of maximum likelihood. The maximum likelihood estimation is carried out with the Fisher scoring algorithm. This procedure provides four variable selection methods: forward selection, backward elimination, stepwise selection and best subset selection. The method used in our logistic regression model is the forward stepwise selection [11, 12].

There are several steps in evaluating the appropriateness, adequacy and usefulness of the model. First, the importance of each explanatory variable can be determined by statistical tests of the significance of the coefficients. Then, the overall goodness of fit of

the model is evaluated, as well as the ability of the model to discriminate between the two output responses.

3.2 CONVERGENCE CRITERION

Convergence criterion was calculated by the likelihood ratio test. The likelihood ratio test is a statistical test used to compare the fit of two models, one of which is the special case of the other. This test shows how likely the data under one model than the other are closer to the output response. It compares the likelihood ratio when the output is zero, L_0, with the likelihood ratio of the output when it is one, L_1. The likelihood ratio can be used to decide whether to keep the model or reject it in favor of the alternative one. The statistics test is calculated as follows:

$$-2\times \ln\ \textit{likelihoodratio}\ = -2\times \ln\left({L_0}\big/{L_1} \right) = -2\times\ \ln L_0 - \ln L_1 \tag{3-2}$$

If the distribution of likelihood ratio can be determined, then it can be directly used for decision making and choosing the better one compared to the alternative model. However in most cases, determining the likelihood distribution corresponding to a specific hypothesis is not that easy. Large sample distribution of likelihood ratio states that the likelihood ratio distribution will be an asymptotically chi-squared distribution with degrees of freedom equal to the difference in dimensionality of L_1 and L_0 [13, 14]. In the stepwise method of the regression model, there is one difference in dimensionality of L_1

and L_0 likelihood ratio. Thus, the result is compared with a chi-square distribution X^2 with one degree of freedom.

3.3 MODEL FITNESS

Logistic regression procedure uses some criterion to choose the best features for creating the most fitted statistical modeling. These criterions are discussed as follows.

3.3.1 AIC Criterion. The Akaike Information Criterion is a measure of relative goodness of fit of a statistical model [15,16]. Developed by Akaike around 1974 under the name of 'an information criterion' (AIC), it is grounded in information theory and measures the information lost when a given model is used to describe reality [17]. It can return a tradeoff between bias and variance in the model and loosely evaluate accuracy versus complexity of a model. AIC values provide a means for model selection based on the statistical likelihood function.

The Kullback-Leibler distance is the amount of information lost when model *i* is used, thus the best model in a set is the one which has the minimum losses in the amount of information, the one which has minimum Kullback-Leibler distance [18]. Since AIC is the estimate of the amount of information loss, minimizing AIC means finding the optimal model, the model that maximizes the information extracted from data. The K-L distance provided by AIC criterion has an unknown bias so one cannot figure out how far the model is from the true one. But it can be a good criterion to rank the models and find the best one. In the general case AIC is

$$AIC = 2k - 2\ln(L) \tag{3-3}$$

Where *k* is the number of parameters in the statistical model and *L* is the maximum value of the likelihood function of the estimated model. Thus in a set of extracted models for the data, the preferred model is the one with minimum AIC value.

3.3.2 SC Criterion. The Bayesian Information Criterion (BIC) or Schwartz criterion (SBC or SC) is a criterion for model selection among a finite set of models .Based on the likelihood function, BIC is an asymptotic result derived under the assumption that observation distribution belongs to an exponential family. It can be formulized as:

$$BIC = k\ln(n) - 2\ln(L) \tag{3-4}$$

Where *k* is the number of parameters in the statistical model and *L* is the maximum value of the likelihood function of the estimated model and *n* is the number of observed data.

3.4 RESULT

The selected features of the image set were analyzed by the SAS software and the logistic regression procedure performed on this data set. The final result of the procedure is as follows:

X^2	$Pr > X^2$	Feature
16.9463	<.0001	Var Red
3.5161	.0608	R of 3rd RGB
11.7402	.0006	Red value of red cross
3.4100	.0648	Average Distance
8.9708	.0027	G of 5^{th} RGB

Highest Percent Correct is 92.5% (7.7% Sensitivity, 100% Specificity)

4. ALGORITHM RESULT

4.1 SIMULATION RESULT

Output images from analysis based on the algorithm and their PCT transforms are rendered. The red-cross points in the images are the center of annular-granular regions defined by the algorithm and the yellow points are border points related to each center point. Figure 4.1 shows an image where the red-crosses depict the points determined by the algorithm as annular-granular region centers. Figure 4.2 shows the PCT transform of the image from Figure 4.1.

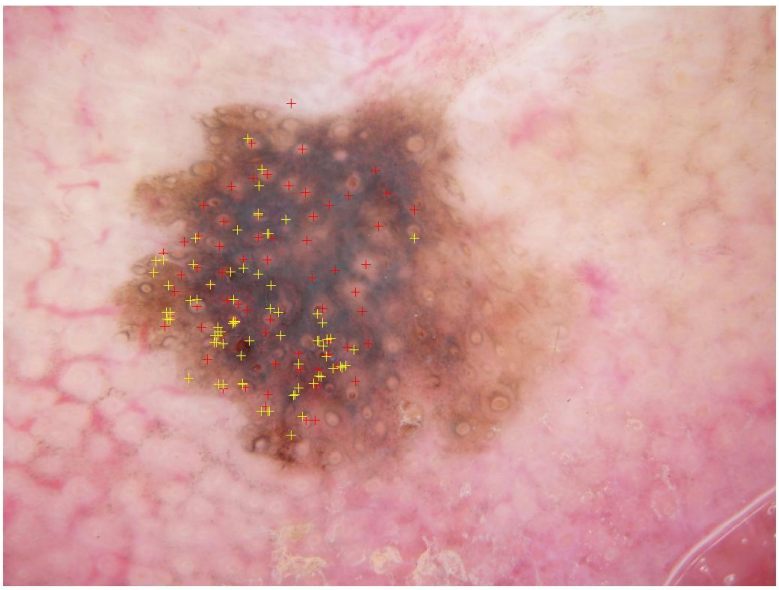


Figure 4.1: Red-cross center points detected in an annular granular region

In Figure 4.3, annular-granular center points determined by the algorithm are illustrated by red–cross points and the border by yellow points. Figure 4.4 shows the PCT transform of Figure 4.3.

In Figure 4.5, the annular-granular areas have been detected by red-cross points and the image PCT transform is shown in Figure 4.6.

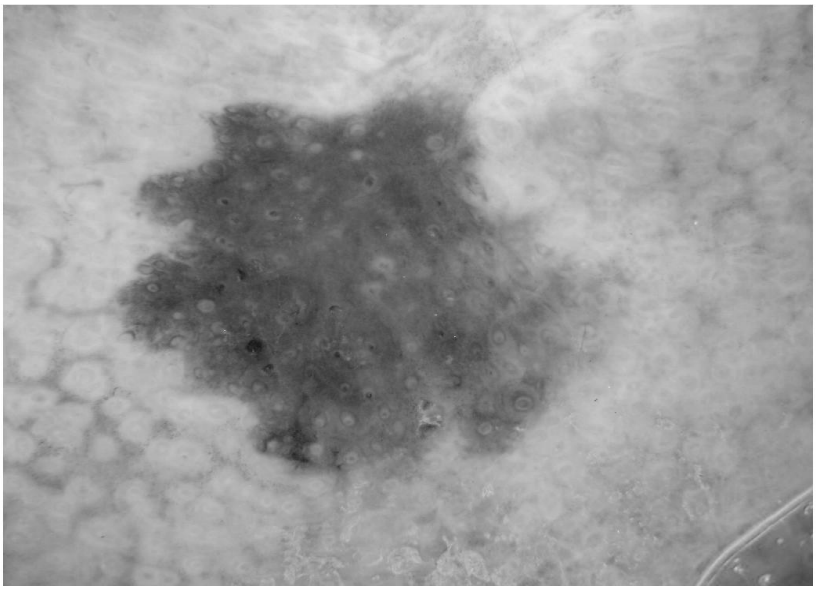


Figure 4.2: PCT image of Figure 4-1

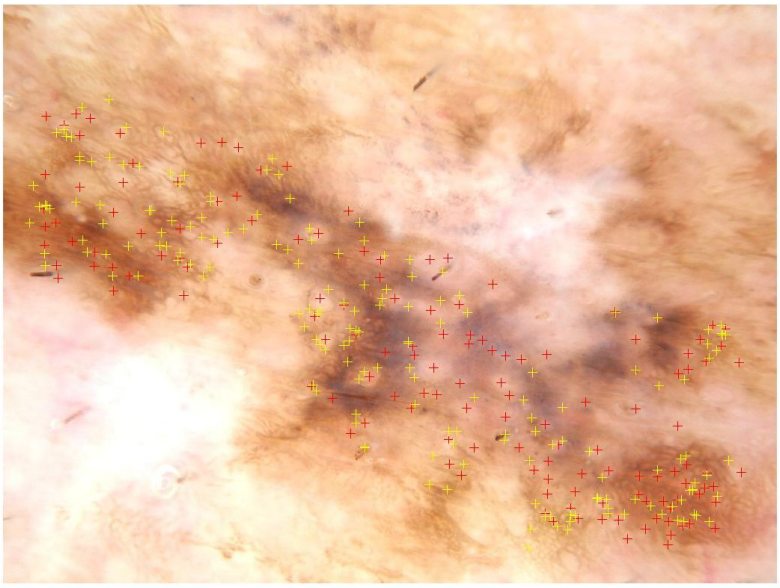


Figure 4.3: Red-crosses are center points detected in a melanoma

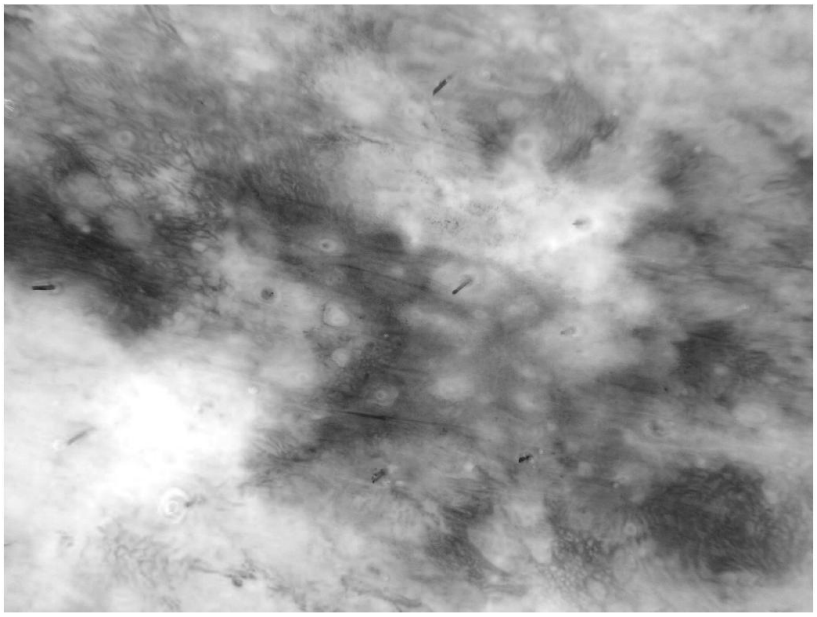


Figure 4.4: PCT image of Figure 4.3

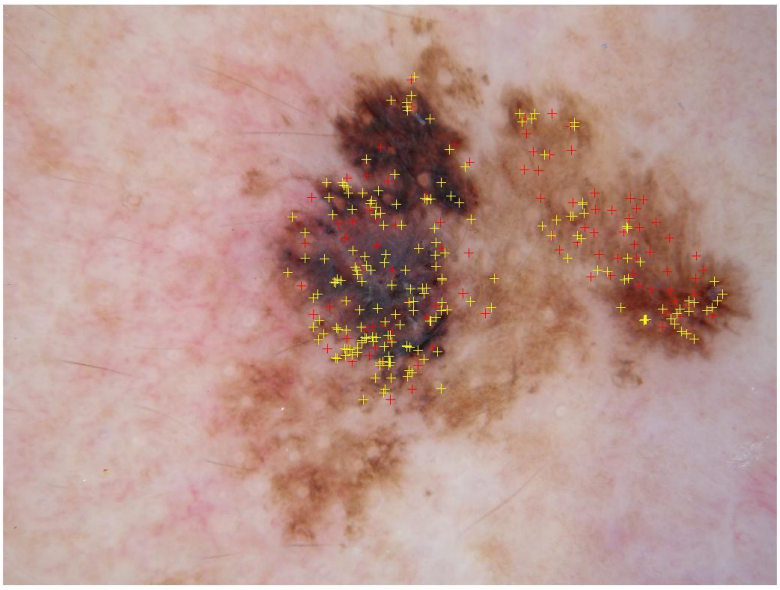


Figure 4.5: Red-crosses are center points detected in a melanoma

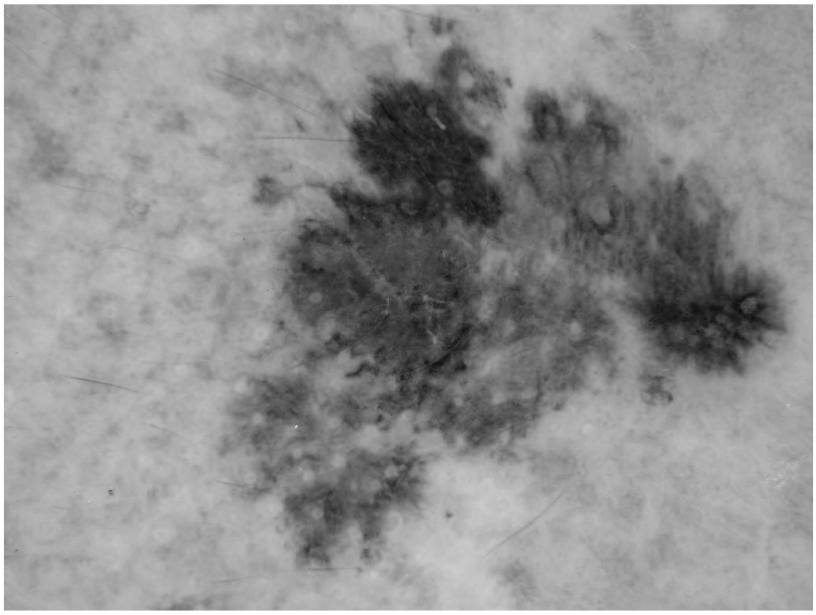


Figure 4.6: PCT image of Figure 4-5

4.2 SAS ANALYSIS

Annular-granular areas were defined in the following the six images manually and a set of features mentioned in section 3-1 were measured and analyzed by the SAS software. The six images analyzed were:

" bb082107rol196", " bm021108rab555" ,"dm041309rab1034",

"ev060507rol117", " hm031708rab605" , " wt022007tre6"

Then the measured features were populated into a spreadsheet file analyzed by the SAS software and the logistic regression procedure with forward stepwise selection was run on this set of features. There were 483 observations in the spreadsheet of which 39 observations were true annular granular regions and 444 were false annular-granular regions.

A Receiver Operating Characteristic curve, or ROC curve, plots sensitivity versus one minus specificity as a function of threshold. Sensitivity relates to the test's ability to identify positive results and is defined as $Sensitivity = \frac{\text{number of true positives}}{\text{number of true positives + number of false negatives}}$. Specificity relates to the test's ability to identify negative results and is defined as $Specificity = \frac{\text{number of true negatives}}{\text{number of true negatives + number of false positives}}$. For the features defined above, SAS analysis (forward stepwise logistic regression) produces an area under the ROC curve of 84.7%.

The classification table produced by this same SAS procedure is shown in Table 4.1.Because the data contain so many more false annular-granular regions (444) than true regions (39), the highest percent correct (92.5%) occurs at probability levels of 0.620

through 0.720 where sensitivities are quite low (10.3% to 7.7%) and specificities are high (99.8% to 100%). At a probability level of 0.620, for example, only 10.3% of the true annular-granular regions would be detected, but 99.8% of the undetected regions are actually not annular granular-regions. Since malignant melanomas generally do not have annular-granular regions, this threshold could be chosen so the user could be fairly certain that melanomas were not being incorrectly eliminated.

The sensitivity and specificity are approximately equal at a probability level of 0.060 (sensitivity of 69.2% and specificity of 69.6%). At this threshold, the percent correct is 69.6%.

Table 4.2 shows the features SAS selected by forward stepwise logistic regression to be the best model for the data. The linear logistic coefficients α and β were estimated by the SAS software and are indicated in Table 4.2.In this table c1 through c44 are the features extracted from images mentioned in Section 3. The Estimate column in Table 4.2 represents β values related to each feature and the α intercept value.

Table 4.1 Logistic Procedure –Classification Table

Prob Level	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.000	8.1	100.0	0.0	91.9	.
0.020	47.6	89.7	43.9	87.7	2.0
0.040	61.7	82.1	59.9	84.8	2.6
0.060	69.6	69.2	69.6	83.3	3.7
0.080	74.7	66.7	75.5	80.7	3.7
0.100	76.4	56.4	78.2	81.5	4.7
0.120	78.5	51.3	80.9	81.0	5.0
0.140	81.2	51.3	83.8	78.3	4.9
0.160	82.8	48.7	85.8	76.8	5.0
0.180	84.7	48.7	87.8	74.0	4.9
0.200	85.9	43.6	89.6	73.0	5.2
0.220	86.3	38.5	90.5	73.7	5.6
0.240	87.0	33.3	91.7	74.0	6.0
0.260	88.2	33.3	93.0	70.5	5.9
0.280	88.6	33.3	93.5	69.0	5.9
0.300	89.0	30.8	94.1	68.4	6.1
0.320	89.2	25.6	94.8	69.7	6.4
0.340	90.1	25.6	95.7	65.5	6.4
0.360	91.5	25.6	97.3	54.5	6.3
0.380	91.5	25.6	97.3	54.5	6.3

Table 4.1 Logistic Procedure –Classification Table.(cont.)

Prob Level	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.400	91.3	20.5	97.5	57.9	6.7
0.420	91.3	17.9	97.7	58.8	6.9
0.440	90.9	12.8	97.7	66.7	7.3
0. 460	90.7	10.3	97.7	71.4	7.5
0.480	90.9	10.3	98.0	69.2	7.4
0.500	91.1	10.3	98.2	66.7	7.4
0.520	91.1	10.3	98.2	66.7	7.4
0.540	91.5	10.3	98.6	60.0	7.4
0.560	91.7	10.3	98.9	55.6	7.4
0.580	92.1	10.3	99.3	42.9	7.4
0.600	92.1	10.3	99.3	42.9	7.4
0.620	92.5	10.3	99.8	20.0	7.3
0.640	92.5	7.7	100.0	0.0	7.5
0.660	92.5	7.7	100.0	0.0	7.5
0.680	92.5	7.7	100.0	0.0	7.5
0.700	92.5	7.7	100.0	0.0	7.5
0.720	92.5	7.7	100.0	0.0	7.5
0.740	92.1	2.6	100.0	0.0	7.9
0.760	92.1	2.6	100.0	0.0	7.9
0.780	92.1	2.6	100.0	0.0	7.9
0.800	92.1	2.6	100.0	0.0	7.9

Table 4.1 Logistic Procedure –Classification Table.(cont.)

Prob Level	Correct	Sensi- tivity	Speci- ficity	False POS	False NEG
0.820	92.1	2.6	100.0	0.0	7.9
0.840	92.1	2.6	100.0	0.0	7.9
0.860	92.1	2.6	100.0	0.0	7.9
0.880	92.1	2.6	100.0	0.0	7.9
0.900	91.9	0.0	100.0	.	8.1

Table 4.2 Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Error	Chi-Square	Pr > ChiSq
Intercept	1	4.7857	3.3062	2.0952	0.1478
c1	1	-0.0817	0.0424	3.7168	0.0539
c3	1	0.0892	0.0403	4.8901	0.0270
c5	1	-0.00305	0.00183	2.7701	0.0960
c7	1	0.0783	0.0222	12.4859	0.0004
c10	1	-0.4931	0.1798	7.5236	0.0061
c11	1	0.00234	0.00238	0.9648	0.3260
c17	1	0.00413	0.00229	3.2499	0.0714
c23	1	0.1226	0.0436	7.9120	0.0049
c24	1	-0.2710	0.1253	4.6758	0.0306
c25	1	0.1398	0.0872	2.5687	0.1090
c26	1	0.00292	0.00242	1.4577	0.2273
c29	1	-0.1117	0.0560	3. 0	.0461

Table 4.2 Analysis of Maximum Likelihood Estimates

c30	1	0.2198	0.1307	2.8280	0.0926
c31	1	-0.1020	0.0846	1.4530	0.2280
c32	1	-0.0506	0.0378	1.7884	0.1811
c33	1	0.2141	0.1101	3.7822	0.0518
c34	1	-0.1650	0.0800	4.2603	0.0390
c36	1	0.0280	0.0131	4.5546	0.0328
c38	1	-0.0462	0.0355	1.6920	0.1933
c39	1	-0.0774	0.0261	8.8127	0.0030
c41	1	-0.0353	0.0217	2.6436	0.1040
c42	1	-0.0800	0.0542	2.1804	0.1398
c44	1	0.1561	0.0592	6.9607	0.0083

5. CONCLUSION

In this method, the detection of annular-granular regions within lentigo maligna melanoma *in situ* are based on the luminance difference between the annular area and the granular area. Thus, higher contrast images can lead to more accurate results. The better the image contrast, the more reliable result in detection of melanoma *in situ*.

Detecting blotch and very dark areas in the lesion can aid in determining the most useful darkest percent of lesion. Using proper thresholding according to the image helps us to remove more false positive points. SAS analysis result shows that the highest percent correct in detection annular granular melanoma feature by this algorithm is 92.5%.

BIBLIOGRAPHY

[1] http://www.mdanderson.org/patient-and-cancer-information/cancer-formation/cancer-types/skin-cancer/prevention/index.html

[2] http://www.cancer.gov/cancertopics/wyntk/skin

[3] http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0001827/

[4] Nishat Nepal, "Detection of Annular Granular Melanoma in Situ", M.S. Thesis, Electrical and Computer Engineering Department, Missouri University of Science and Technology, 2007.

[5] http://seer.cancer.gov/statfacts/html/melan.html

[6] http://www.skincarephysicians.com/skincancernet/four_types.html#Superficial Spreading.

[7] G. Argenziano, H.P. Soyer, and V. De Giorgi, Dermoscopy: A tutorial, Milan, Italy: EDRA, Medical Publishing & New Media, 2002.

[8] Sayali Deshpande, "Identification of Radial Streaming and Pseudopods in Dermoscopy Images", M.S. Thesis, Department of Electrical and Computer Engineering, Missouri University of Science and Technology, 2007.

[9] Howard Zhou, Mei Chen, Le Zou, Richard Gass, Laura Ferris, Laura Drogowski, James M.Rehg, " Spatially Constrained Segmentation of Dermoscopy Images", 5th IEEE International Symposium on Biomedical Imaging (ISBI), 800-803, May, 2008..

[10] S.W. Menzies, K.A. Crotty, C. Ingwar, and W.H. McCarthy, An atlas of surface microscopy of pigmented skin lesions: Dermoscopy, Sydney, Australia: McGraw-Hill, 2003.

[11] http://www2.stat.unibo.it/manualisas/stat/chap39.pdf

[12] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1065119/

[13] http://luna.cas.usf.edu/~mbrannic/files/regression/Logistic.html

[14] http://courses.washington.edu/b515/l13.pdf

[15] http://www.garfield.library.upenn.edu/classics1981/A1981MS54100001.pdf

[16] Akaike, H. " Information theory and an extension of the maximum likelihood principle. " Second International Symposium on Information Theory. Ed. B N Petrov & F Csaki. Akademiai Kiado, 1973. 267-281.

[17] DeLeeuw, J. " Introduction to Akaike (1973) Information Theroy and an Extension of the Maximum Likelihood Principle. " Breakthroughs in Statistics Volume I Foundations and Basic Theory. Ed. Samuel Kotz & Norman L Johnson. Springer-Verlag, 1992. 599-609. Print.

[18] S. Kullback and R. A. Leibler," On Information and Sufficiency" , The Annals of Mathematical Statistics, Vol. 22, No. 1 (Mar., 1951), pp. 79-86, published by: Institute of Mathematical Statistics

[19] http://personalwm.com/2010/05/11/investment-risk-in-greater-detail/

[20] http://www.skincancer.org/skin-cancer-information/melanoma

VITA

Parivash Hajiyani was born in Iran. She completed her primary and secondary education in Iran. She graduated with a Bachelor's degree in Electrical Engineering from Amirkabir University-Tehran Polytechnic, Iran in August 2001. Prior to pursuing her M.S. degree, she worked with Iran Telecommunications Research Center as a researcher. She obtained her Master's degree in Electrical Engineering from Tarbiat Modares University, Iran.

She enrolled for her second Master of Science in Electrical Engineering at Missouri University of Science and Technology (formerly University of Missouri-Rolla) in August 2010. She worked as a Graduate Research Assistant with the DERMVIS group of cancer research from August 2010 to May 2012. She completed her second Master of Science degree in Electrical Engineering in August 2012.