Masters Theses   Student Theses and Dissertations

Fall 2012

Identification of starry milia-like cysts in seborrheic keratoses and malignant melanomas using image processing techniques

Viswanaath Subramanian

IDENTIFICATION OF STARRY MILIA-LIKE CYSTS IN SEBORRHEIC

KERATOSES AND MALIGNANT MELANOMAS USING IMAGE PROCESSING

TECHNIQUES

by

VISWANAATH SUBRAMANIAN

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

Dr. Randy Moss, Advisor
Dr. Kurt Kosbar
Dr. Sahra Sedigh

ABSTRACT

Seborrheic keratoses are the most common benign tumors found in older individuals. Seborrheic keratoses are normally benign in nature, but sometimes in a large lesion area of seborrheic keratoses, secondary tumors, or squamous cell carcinoma in situ or malignant melanoma may occur. It has been found that seborrheic keratoses have a few features that distinguish them from melanomas. Milia-like cysts, comedo-like openings and network-like structures are the most prominent features found in seborrheic keratoses. Comedo-like openings are pigmented structures and are mostly oval yellowish, brown or black in appearance. They can also be observed in melanocytic lesions and melanomas. Milia-like cysts (MLCs) are structures which are mostly white or yellowish in color, vary in size and are mostly seen in seborrheic keratoses and in certain congenital melanocytic nevi. MLCs are very useful in classifying a skin lesion as a seborrheic keratosis and hence are also useful in differentiating between seborrheic keratoses and melanomas. MLCs are classified into two categories depending on the size and shape of the MLC. The two types of MLCs are starry MLCs and cloudy MLCs. This research project gives a method to automatically identify starry milia-like cysts in any image. The aim of the research is to identify starry milia-like cysts accurately and to distinguish starry MLCs from competing structures like skin pores, bubbles from dermoscopy fluids and keratin scales.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

1.1. SEBORRHEIC KERATOSES

Seborrheic Keratoses (SKs) which have also been known as basal cell papillomas, seborrheic warts or senile keratoses are a very common form of skin lesion [1]. Seborrheic keratoses are the most common benign tumors found in older individuals. There have been no specific etiologic factors that have been identified for the occurrence of seborrheic keratoses. They are more found in both sun exposed and non-sun exposed areas. In spite of various advancements in the understanding of many common skin diseases, very little is documented about the true epidemiology of these lesions and even less has been studied and reported about their etiology [1]. It has been found that seborrheic keratoses usually appear after the age of 50 in people living in temperate climates but the people in tropical regions, with greater sun exposure, may develop seborrheic keratoses much earlier [1].

The frequency of occurrence of seborrheic keratoses appears to increase with the age of an individual. In one of the studies conducted by Tindall and Smith in 1963 [2], it was found that 88% of the people from a population of individuals older than 64 years in North Carolina had at least one seborrheic keratosis. Another important fact that was highlighted by the study was that 38% of the white women, 54% of the white men and 61% of the black men and women were found to have more than 10 seborrheic keratoses. Another similar study was conducted by Young in 1965 [3]. Young examined 222

residents of the Orthodox Jewish Home for Aged in New York and seborrheic keratoses were found in 29.3% of the men and 37.9% of the women.

Seborrheic keratoses are not only common in the United States of America but are commonly found in different parts of the world. In 2000, Memon et al. [4] found out that 8.3% of the males and 16.7% of the females in a British population less than 40 years of age had at least one seborrheic keratosis. In a study done in Australia on 100 people composed of hospital staff and non-dermatologic day patients [5], 12% of the 34 people aged 15-25 years, 79% of the 24 people aged 26-50 years, 100% of the 25 people aged 51-75 years and 100% of the 17 people older than 75 years had seborrheic keratoses. The median number of seborrheic keratoses per person increased with age as represented in Table 1.1.

Table 1.1. Median number of seborrheic keratoses per person regards to age group [5]

Serial no.	Age group in years	Median number of seborrheic keratoses per person
1	15-25	6
2	26-50	5
3	51-75	23
4	Greater than 75	69

Seborrheic keratoses are found to be less common in populations having dark skin compared to those having white skin. But black individuals develop dermatosis papulosa nigra which is a variant of seborrheic keratoses that is normally found on the face, upper

cheeks and lateral orbital areas. This variant generally occurs earlier than common seborrheic keratoses. It was found that there was no apparent gender difference in the frequency of occurrence of seborrheic keratoses [6].

Seborrheic keratoses are normally benign in nature but secondary tumors, or squamous cell carcinoma in situ or malignant melanoma may sometimes occur within the lesion. Seborrheic keratoses can itch, grow and sometimes even bleed. Scratching seborrheic keratoses or trying to take them off the skin can result in inflammation at the lesion area, known as an irritated or inflamed seborrheic keratosis, which can be difficult to distinguish from squamous cell carcinoma. Due to the presence of many seborrheic keratoses in a certain area of the skin, dysplastic nevus or malignant melanoma may not be detected [6].

1.2. MALIGNANT MELANOMA

Malignant melanoma (melanoma, invasive melanoma, cutaneous melanoma) occurs in every age group. Although rare in young children, it is currently the most common cancer in young adults in the age group of 25-29. For women in the age group of 30-35, breast cancer is the only cancer that exceeds the occurrence of melanoma. When all the forms of cancer are considered, melanoma is the sixth most common cancer in males and seventh in females [7]. The incidence has been found to be increasing at an alarming rate in the past few decades. Right now in the United States of America, incidence of melanoma exceeds that of new HIV infections. On average, there is one death almost every hour in the U.S., due to melanoma [7].

In 2005, an estimated 59,580 cases of invasive melanoma and 46,170 melanoma in situ (MIS) were diagnosed, a total of 105750 [8]. In 2011, an estimated 70,230 cases of invasive melanoma and 53,360 cases of MIS were diagnosed, a total of 123,590 cases. [9]

Melanoma poses a huge challenge because of the rising incidence. From the above statistics, melanoma is increasing at an alarming rate of almost 3% per year between 2005 and 2011. Since melanoma can be treated if diagnosed at an early stage, it is very essential that efforts should be taken to make the general public aware about melanoma. It is estimated that 70,230 men and women (40,010 men and 30,220 women) will be diagnosed with invasive melanoma out of which 8,790 men and women will die in 2011 alone [10].

Table 1.2 shows the incidence rates in the U.S. The incidence rate was found to be 20.8 per 100,000 men and women per year considering all age groups. These incidence rates based on ethnicity and gender are the cases that were diagnosed between 2004 and 2008 from 17 SEER geographic areas [10].

Table 1.2. Incidence rates by race and gender [10]

Race/Ethnicity	Male (Incidence per 100,000 men)	Female (Incidence per 100,000 women)
All races	26.7	16.7
White	30.9	19.7
Black	1.2	0.9
Asian / Pacific Islander	1.6	1.3
American Indian	3.9	3.7
Hispanic	4.8	4.8

Table 1.3 shows the mortality rates in the U.S. The mortality rate was found to be 2.7 per 100,000 men and women per year considering all age groups. The death rates are based on patients who died between 2004 and 2008 in United States [10].

Of all the skin cancers, melanoma is the least common but is collectively the most deadly, lagging only to Merkel cell carcinoma with respect to the fatality rate. According to the American Cancer society, of all the skin cancer cases only 4% are accounted by melanoma, but it is the most deadly form of skin cancer since it causes 79% of all skin cancer related deaths. Race is one of the primary risk factors for developing melanoma, with fair-skinned people at greater risk than darker-skinned people. In the United States, white Americans are 20 times more likely to develop melanoma than African Americans. Worldwide, white populations have the highest risk of developing melanoma, and Asian populations the lowest risk [11].

Table 1.3. Death rates by race and gender [10]

Race/Ethnicity	Male (Incidence per 100,000 men)	Female (Incidence per 100,000 women)
All races	4.0	1.7
White	4.6	2.0
Black	0.5	0.4
Asian / Pacific Islander	0.4	0.3
American Indian	1.4	0.8
Hispanic	1.0	0.6

1.2.1. Etiolojgy of Malignant Melanoma. Exposure to ultraviolet radiation is one of the most critical factors in the development of most melanomas. The ultraviolet radiations having wavelengths in the range of 290-320 nm (ultraviolet B) and 321-400 nm (ultraviolet A) are potentially carcinogenic and may have an influence in the development of melanoma. Another very interesting factor is that melanoma is not directly proportional to the amount of sun exposure because it is found that melanomas are more common in white collared people than in those who work outdoors [12].

The greatest risk factor in the development of melanoma that is induced by sun exposure is from the blistering sun burns that are found in those areas of the body which are not very much exposed to the sun [12]. Some of the factors which elevate the risk of malignant melanoma are as follows: [12]

- Having more than 50 nevi which are 2mm or greater in diameter
- Dysplastic nevi
- Changing mole
- Family history of melanoma
- Previous occurrence of melanoma
- High exposure to sun

1.2.2. Epidemiology of Malignant Melanoma. Since malignant melanoma is the most deadly form of skin cancer, it is very important that it is detected in its early stages. In the U.K., it has been estimated that the lifetime risk of developing malignant melanoma is 1 in 61 for men and 1 in 60 for women. This study was done in 2008 using AMP method [13].

Various studies have shown that malignant melanoma is more prevalent in white people as compared to other ethnicities. Also the possibility of occurrence of malignant melanoma increase with age. These findings are reflected in Table 1.4 and Table 1.5:

Table 1.4. Average number of new cases of malignant melanoma per year and age-specific incidence rates per 100,000 population, UK, 2006-2008 [14]

Age Range	Male Cases	Female Cases	Male Rates	Female Rates
0 to 04	0	1	0	0
05 to 09	1	1	0	0.1
10 to 14	3	4	0.2	0.2
15 to 19	21	37	1	1.9
20 to 24	53	123	2.5	6.1
25 to 29	105	206	5.2	10.4
30 to 34	138	266	7.1	13.5
35 to 39	242	412	10.8	18.1
40 to 44	319	481	13.7	20.3
45 to 49	327	471	15.6	21.9
50 to 54	384	481	20.8	25.4
55 to 59	491	548	26.5	28.7
60 to 64	648	606	38.4	34.3
65 to 69	620	481	47.6	34.1
70 to 74	630	486	57.1	38.5
75 to 79	499	464	58	41.7
80 to 84	410	405	71.8	45.8
85+	297	429	74	48.2
All Ages	5189	5901	17.3	19

Table 1.5. World age-standardized incidence rates of malignant melanoma per 100,000 population, world regions, 2008 estimates [15]

World Region	Female	Male
Australia / New Zealand	32.1	41.8
Northern America	12.5	15.8
Northern Europe	13.4	12.3
Western Europe	12	10.6
Southern Europe	6.5	6.5
Southern Africa	4.2	7.7
Central and Eastern Europe	4.3	4.4
World	2.7	3.1
South America	1.8	2.1
Eastern Africa	2	1.7
Middle Africa	1.4	1.7
Central America	1.5	1.5
Western Asia	1.2	1.5
Western Africa	1.1	1.2
Caribbean	0.6	0.7
South-Eastern Asia	0.5	0.6
Northern Africa	0.3	0.4
Eastern Asia	0.3	0.3
South-Central Asia	0.2	0.2

1.3. SEBORRHEIC KERATOSES AND MALIGNANT MELANOMA

A study was conducted by Izikson, et al. [16] to estimate the prevalence of melanoma clinically imitating seborrheic keratoses. The analysis was done on 9204 cases containing a diagnosis of seborrheic keratoses between 1992 and 2001. Those cases having a final histological diagnosis of melanoma were considered for further clinic-pathological analysis. The results were based on the histological diagnosis which was correlated with the clinical diagnosis. As previously stated in the seborrheic keratoses section, the results from this study confirmed that melanomas can mimic seborrheic keratoses. It was confirmed that out of 9204 cases, melanoma was identified in 61 cases, i.e. 0.66%. These 61 cases were identified from the submitted histological examinations which were clinically diagnosed that included seborrheic keratosis. Out of the 61 cases identified as melanoma, in 31 cases melanoma was in the clinical differential diagnosis, whereas 17 cases had a differential diagnosis of seborrheic keratoses vs. melanocytic nevus, 7 cases had a differential diagnosis of seborrheic keratoses vs. basal cell carcinoma, 3 cases had a differential diagnosis of seborrheic keratoses vs. squamous proliferation. In the remaining 3 cases, seborrheic keratosis was the only clinical diagnosis [16].

1.4. MILIA-LIKE CYSTS

1.4.1. Overview. A virtual meeting on the internet was organized on dermoscopy to investigate the usefulness of various features and diagnostic algorithms in the diagnostic significance of melanocytic and non-melanocytic lesions. For the

dermoscopic classification of skin lesions, a two-step procedure was used. The first-step algorithm was used basically to differentiate between melanocytic and non-melanocytic lesions with the help of various dermoscopic criteria. The second step concentrates on differentiating between benign melanocytic lesions and melanomas using the following four diagnostic algorithms [17]:

1. Modified pattern analysis
2. ABCD rule of dermoscopy
3. Menzies method
4. 7-point checklist

The term "milia-like cysts" was introduced for the first time in this particular study. Milia-like cysts were considered as one of the dermoscopic criteria to differentiate between melanocytic and non-melanocytic lesions and multiple milia-like cysts are used as criteria for the diagnostic significance of seborrheic keratosis [17].

1.4.2. Dermoscopy of Pigmented Seborrheic Keratosis. Skin lesions having seborrheic keratoses were obtained from two different sites and a morphological study was conducted on the dermoscopy of pigmented seborrheic keratoses. The first site was from a private practice in Plantation, Fla and the second site was the Department of Dermatology at the University Hospital Geneva in Switzerland. From these two sites, a total of 203 pigmented seborrheic keratoses were collected with complete documentation. The objective of the study was to measure the various morphological features present in seborrheic keratoses and to examine their frequency. From the entire 203 pigmented seborrheic keratoses, 15 morphological features were identified. Out of the 15 features, milia-like cysts and comedo-like openings were the most frequently found features (135

and 144 respectively). Some other features like networks and network-like structures were also present in 46% of the skin lesions [18].

1.4.3. Classification of Milia-like Cysts. It has been found that milia-like cysts are very common in seborrheic keratoses. But varied appearances of milia-like cysts can make it difficult to differentiate between benign seborrheic keratoses and malignant melanoma which is the deadliest form of skin cancer. Milia-like cysts (MLCs) are features or structures which vary in size but are mostly white or yellowish in color and are round in shape. MLCs are mostly seen in seborrheic keratoses, but are also seen in certain congenital melanocytic nevi. MLCs are useful in classifying a skin lesion as a seborrheic keratosis and hence are useful in distinguishing between melanomas and benign lesions. The MLCs were divided into two types based on the appearance and size. Small MLCs whose diameter is often less than 1/3 mm and which have an appearance like a star in a misty night with bright centers and having sharp borders are one of the types of MLCs. These are the most common type of MLC. They are referred to as 'starry' MLCs. MLCs which have a diameter greater than 1/3 mm are considered as large MLCs and they appear less frequently as compared to starry MLCs. The large sized MLCs are normally oval in shape and lack significant irregularities and have a fluffy border. These MLCs are referred as 'cloudy' MLCs. In this study, the odds ratio, sensitivity and specificity were calculated for starry and cloudy MLCs. For starry MLCs, the study showed an odds ratio of 8.0 (4.7-13.7) with a high sensitivity of 90.5% and 45.7% specificity for a diagnosis of seborrheic keratoses vs. all melanoma [19]. Normally the odds ratio denotes the odds of a malignant diagnosis for a given sign, but here the odds

ratio denotes the odds of a benign diagnosis for the given sign. The definition of

sensitivity and specificity are given as follows:

$$sensitivity = \frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ negatives}$$

$$specificity = \frac{number\ of\ true\ negatives}{number\ of\ true\ negatives + number\ of\ false\ positives}$$

2. PRE-PROCESSING OF SEBORRHEIC KERATOSES AND MALIGNANT MELANOMA LESIONS

2.1. SUMMARY

Starry and cloudy milia-like cysts (MLCs) have certain unique characteristics which can be used to separate them from competing structures such as small areas of keratin scale, bubbles from the contact dermoscopy fluid interface and tiny pigment pores. For every hair on the body there is a corresponding skin pore. Scales are nothing but excess dead skin cells that are produced due to some abnormal skin conditions. Many stars were studied by our team comprising Dr. Stoecker, Sneha Mahajan, Sherea Stricklin and me, and we determined many characteristics. A study to determine the statistical occurrence of MLCs in benign and malignant lesions was presented in [18].

All the images are taken from The Dermatology Center, Rolla, MO. All the images are of resolution 1024 by 768 and are in JPEG format. Now since starry MLCs are mostly white or yellowish in color, the most important factor to be considered is the intensity of the image. Skin lesions can be of different colors, varying in intensities. Figure 2.1 shows four different seborrheic keratosis lesions. As can be seen from Figure 2.1, lesion areas vary in color and can be pinkish, brown or black in color. Irrespective of the color of the seborrheic keratosis lesion, the color, size and shape of starry milia-like cysts does not change. Due to the color of the lesion, starry milia-like cysts are clearly seen in dark backgrounds, i.e. black or brown lesions, whereas it is difficult to identify starry MLCs in lesions having a light background.
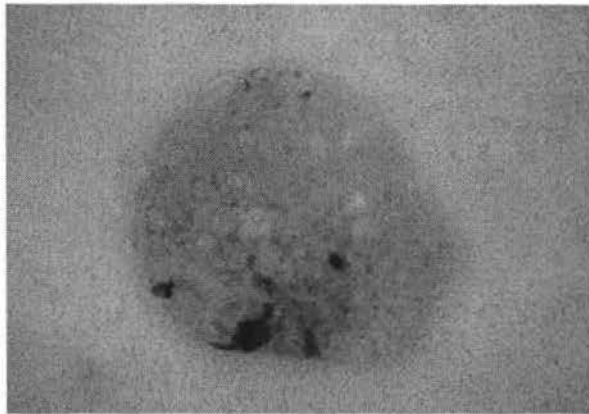
(a) jf101708rab841_CNP
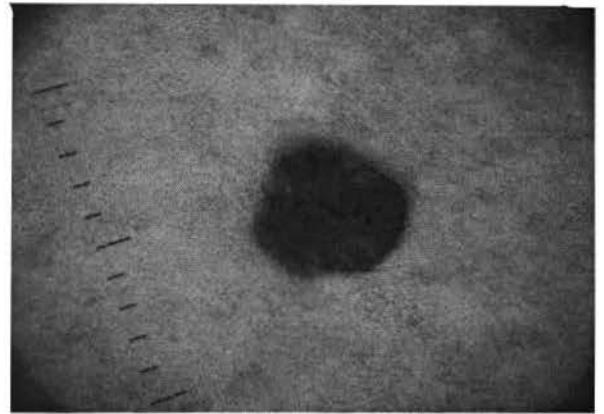
(c) oj070609rab1137_CNP
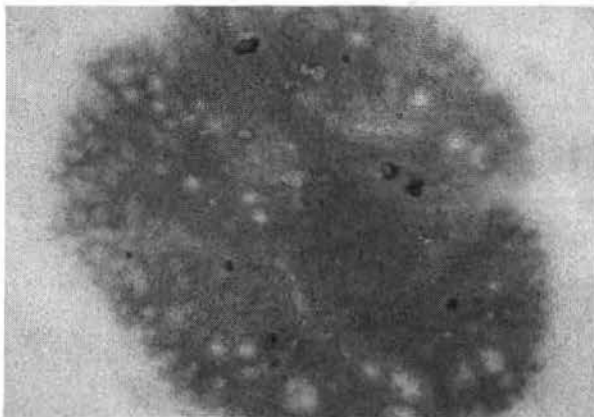
(b) EH050208per75_CNP
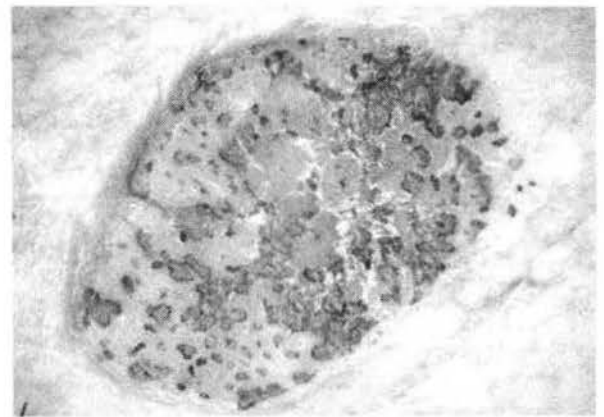
(d) ko041307rab104_CNP

Figure 2.1. Lesion areas of different seborrheic keratoses

2.1.1. Use of Blue Plane. Color images are stored in Matlab as an m-by-n-by-3

data array. It defines red, green and blue color components for each individual pixel in

the image. The actual color of each pixel as seen in the image is determined by the

combination of all the red, green and blue intensities that are stored in the different color

planes at that particular pixel location. In this research project, for the detection of starry

MLCs, the processing of the image is done only considering the blue plane of the image.

The blue plane gives the best contrast in intensities as compared to the red and blue

planes of the original true-color image as shown in Figure 2.2 for image -

BF092908per131 CNP from the seborrheic keratoses set.
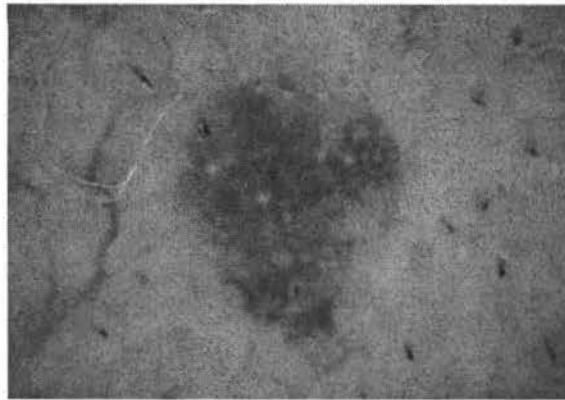
(a) Original Image
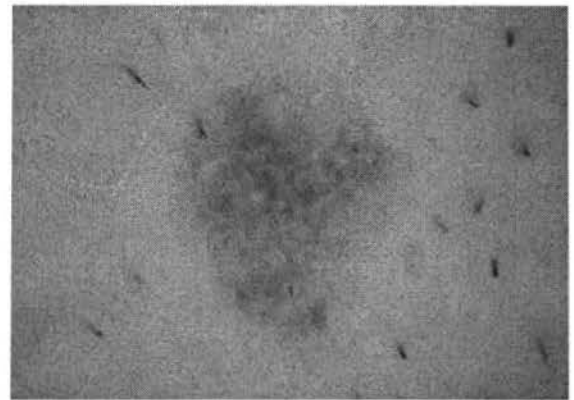
(b) Red Plane
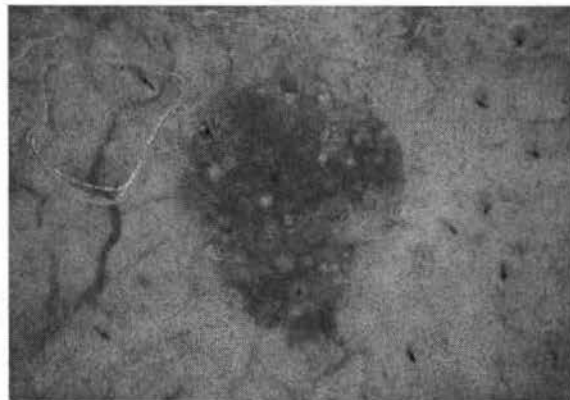
(c) Green plane
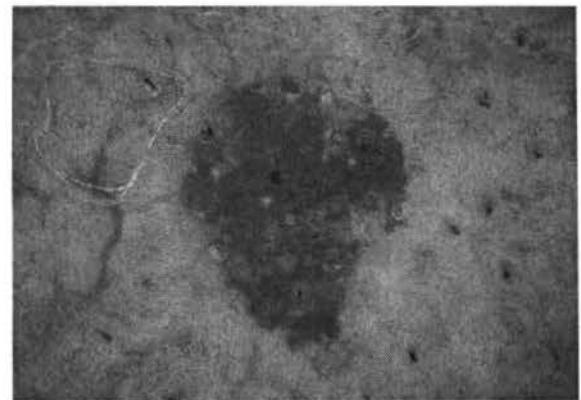
(d) Blue plane

Figure 2.2. Original image of BF092908per131_CNP along with the red, green and blue plane. The original image is shown in (a) and the red, green and blue planes are shown in (b), (c) and (d) respectively

2.1.2. Segmenting the Actual Lesion Area. Since only the lesion area is

required for processing instead of the entire image, a mask is applied on the original

image and only the lesion area is obtained. This lesion mask is manually drawn using

software named 'Winshow' (which creates border masks) and all the borders were checked by a seasoned dermatologist.

Also, the program can consider a bubble to be a starry MLC. In order to avoid this problem, a bubble mask is applied in order to get rid of bubbles which are formed from the contact dermoscopy fluid interface. The bubble masks were created by a program written by Thomas Mark Daniel Szalapski who is one of our research group students at Missouri University of Science and Technology. Figure 2.3 shows the original BF092908per131_CNP image along with the bubble masks, lesion mask and finally the actual lesion area which is obtained by applying the masks on the blue plane of the original image.

Two different bubble masks – a bright bubble mask and a dim bubble mask-were created for the same image. The two bubble masks are created using the same program but with different thresholds. The bright mask finds all the easier more obvious bubbles, whereas the dim mask finds more bubbles (and more false positives) and explicitly excludes the bubbles found with the bright mask. So in order to find all the bubbles, both the masks are used together. When the bubble masks are applied over the original image, the bubble areas are covered and hence one of the main mimics of starry MLCs i.e. bubbles, are removed with the help of these bubble masks. These bubble masks remove almost all the prominent bubbles from the image, but still, a few less noticeable bubbles are left untouched. These leftover bubbles do not have a certain well-defined characteristic which can be used to differentiate them from true starry MLCs.
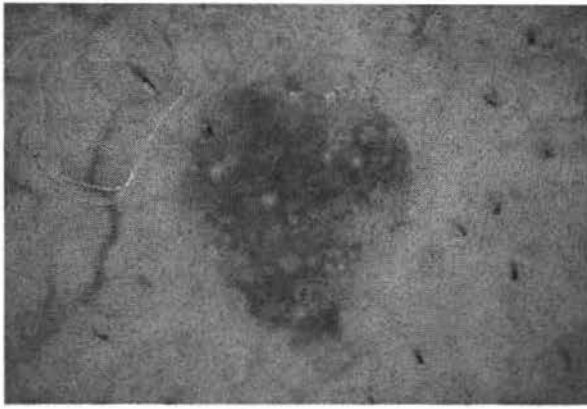
(a) Original Image
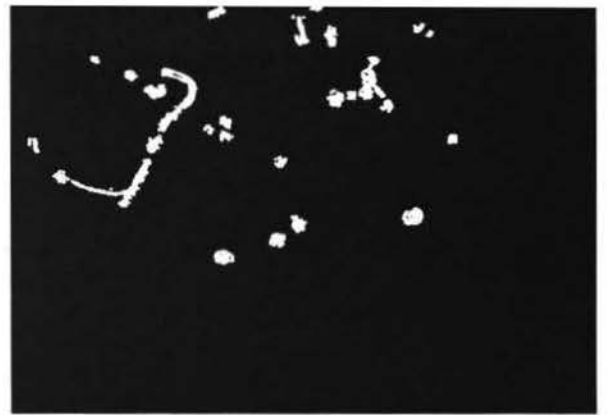
(c) Bright bubble mask

(b) Lesion Mask
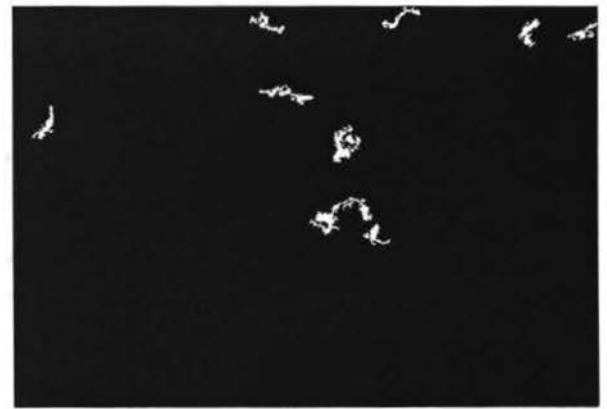
(d) Dim bubble mask
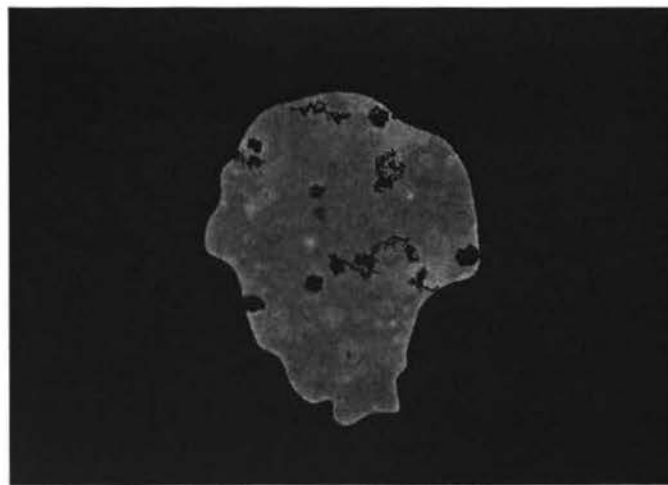
(e) Lesion Area

Figure 2.3. Obtaining lesion area from bubble masks and lesion mask. The original image is shown in (a), the lesion mask is shown in (b), the bubble masks are displayed in (c) and (d) and the final lesion area of the image is shown in (e)

3. SELECTION OF CANDIDATE PIXELS

The total number of pixels in an image is equal to 1024 times 768. If we consider each pixel to be a candidate star, then to run the actual code, which includes a lot of heavy calculations on each candidate star, would take far too much time. Hence, in order to reduce the computational load, certain codes were introduced which would reject the candidate pixels which are not starry MLCs. Certain methods used for this purpose are discussed in the following sections.

3.1. SELECTION OF CANDIDATE PIXELS BY APPLYING BLOCK BY BLOCK SELECTION OF MAXIMUM INTENSITY PIXEL VALUE

In order to close in on the candidate stars, the brightest pixel in the neighborhood of 13 by 13 pixels is selected for further processing. Since we know that stars have a bright center and the intensity goes on decreasing radially outwards, we use this brightest pixel in a 13 by 13 box to select the candidate stars. Also, the box size was selected as 13 by 13 since the smallest of stars were found to have a diameter in the range of 11 to 13 pixels.

With the help of this 13 by 13 box, the total number of pixels which are to be further processed reduces from 1024*768 = 786432 pixels to (1024*768) / (13*13) = 4654 pixels. The image in Figure 3.1 (a) is the original image and the image in Figure 3.1 (b) is the image which has the maximum intensity pixel in every sliding 13 by 13 neighborhood marked on the lesion area of the original image.
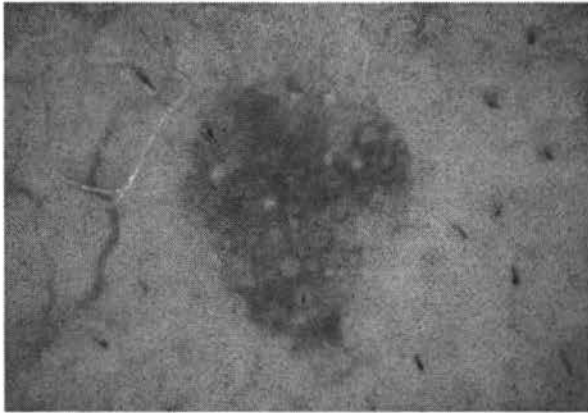
(a) Original image
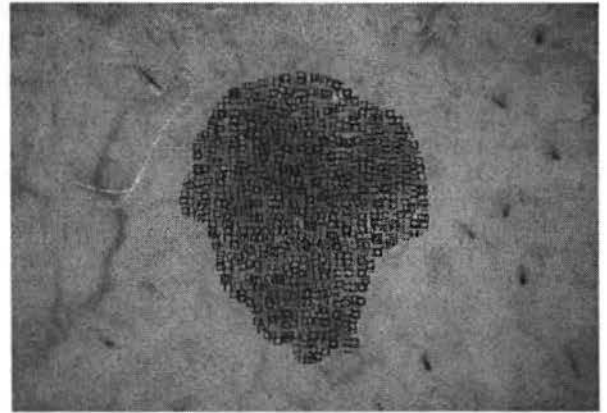
(b) Output of section 3.1

Figure 3.1. Application of block by block selection of maximum intensity pixel value. Original image is shown in (a) and the output of the selection of maximum intensity pixel value using sliding window is shown in (b)

3.2. APPLYING MEAN OF THE LESION AS THRESHOLD

Since the starry MLCs are much brighter than their surroundings, this feature can be used to reduce the candidate starry MLCs with the help of a mean threshold. First the mean intensity value of the entire lesion area is calculated. The original idea was that each pixel value selected as a candidate star from the previous 13 by 13 box code would be compared with the mean intensity value of the entire lesion area, and if the value of the individual pixel is greater than the mean intensity value, then it would be considered for further processing as a candidate star; else it would be discarded. In this method, it was found out that due to the presence of noise; a lot of noisy pixels get selected as candidate stars for further processing. In order to avoid this problem, an averaging filter with equal weights was applied using a 3 by 3 neighborhood function centered on each candidate pixel before comparing the mean value obtained with the mean of the lesion area.

The mean threshold was decreased from just the mean intensity value, to 0.85 times the mean intensity value. This did give better results in most of the images but at the expense of excluding a few true stars in a small number of images when compared with a threshold which is 0.5 times the mean intensity value.

It was decided that the mean threshold be 0.85 times the mean intensity value, because in certain cases like a lesion that contains both dark and bright areas with stars present in both areas, if the threshold is the mean intensity value or above, the stars in the dark area will not be identified. The image in Figure 3.2 (a) is the original image and the image in Figure 3.2 (b) is the image after applying mean of the lesion as the threshold.
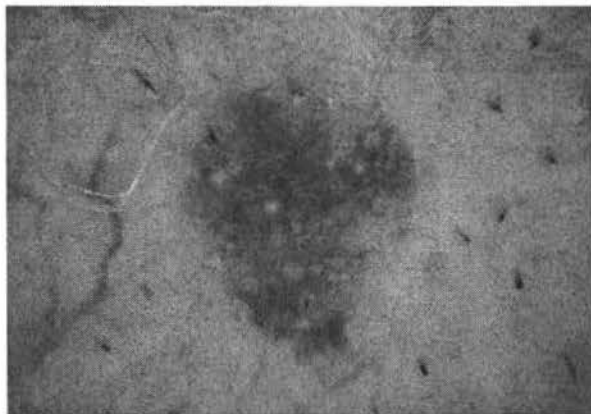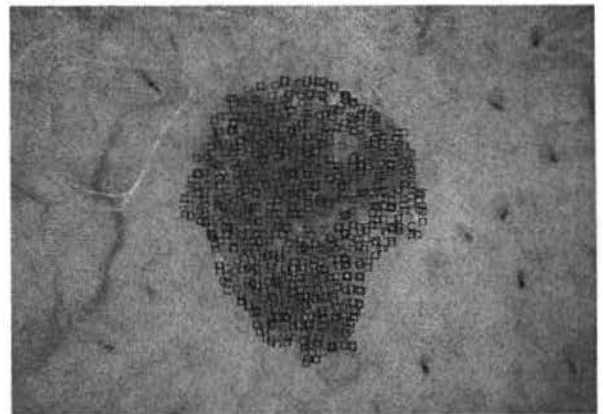
(a) Original image   (b) Output of section 3.2

Figure 3.2. The output obtained by application of 0.85 times the mean of the lesion as the threshold. The original figure is displayed in (a) along with the output of section 3.2 displayed in (b)

3.3. ELIMINATE CLOSE-BY STARS AND DUPLICATE STARS NEAR MASK EDGES

In Section 3.1, the candidate pixels were selected on the basis of having the highest intensity in a particular 13 by 13 pixel area. Now consider the situation of a starry

MLC whose size is much more than 13 by 13. In such a case, by applying the logic of Section 3.1 which is selection of candidate pixels by applying block by block assortment of maximum intensity pixel value, two pixels of the same starry MLC are selected. In such a case out of the two pixels selected, the pixel which has the highest intensity among them will definitely be the center pixel i.e. the brightest pixel is selected and the other candidate star is rejected. This concept is termed as rejection of close-by candidate stars.

Also a lot of candidate stars situated at the edges of the lesion mask are not actual starry MLCs, but they get selected mainly because they are brighter than their surroundings. These duplicate candidate stars are rejected in this particular step by checking the proximity of the candidate star to the lesion border. If the candidate stars are less than 13 pixels away from the lesion mask edges, then those candidate pixels are rejected. Figure 3.3 shows the output after rejecting close-by candidate starry MLCs and duplicate starry MLCs near lesion mask edges.
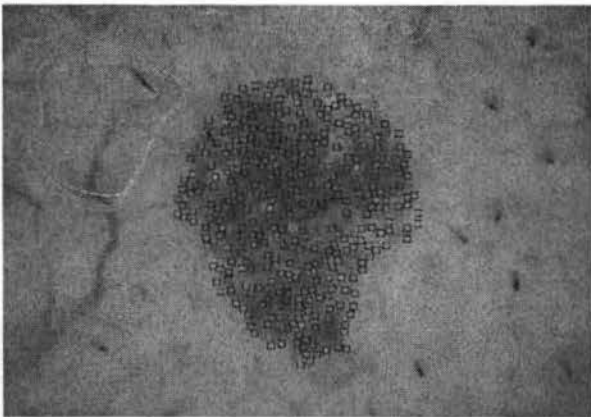
(a) Output after eliminating close-by candidate starry MLCs
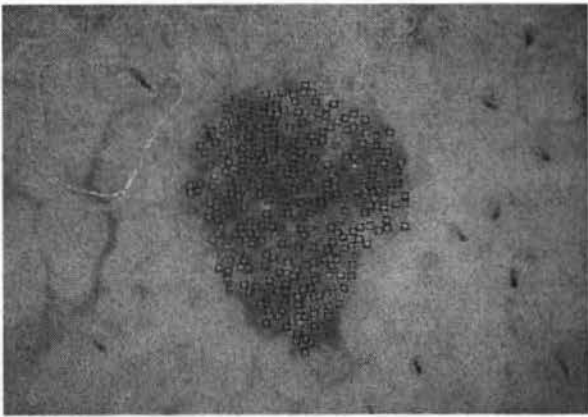
(b) Output after eliminating duplicate candidate starry MLCs

Figure 3.3. Output after rejecting close-by candidate starry MLCs and duplicate starry MLCs near lesion mask edges

3.4. STUDY OF STARRY MLCs

Since the starry MLCs have a bright center and then the intensity reduces as one goes radially outwards, this phenomenon can be used to describe a star completely. Initially stars were considered to be circular and hence various techniques were applied to check the consistency of the drop in intensity of the stars.

3.4.1. Intensity Values at Equidistance from the Center Pixel. One of the methods used to check the assumption that stars are circular was to check the mean intensity in a 3-by-3 neighborhood at an arbitrary distance from the center in all four directions. This can be explained with the help of an image as shown in Figure 3.4. Figure 3.4 shows the enlarged image of one of the candidate starry MLCs. The pixel whose value is 185 is the center pixel of the candidate starry MLC. The mean intensity is calculated along the horizontal and vertical axes through the center pixel at a distance of five pixels away from the center.

This technique to differentiate true starry MLCs from mimics like pores, bubbles and scales was not a big success because of the following reasons:

1. It was found that stars are not exactly circular in shape and the drop in intensity was not constant in all the directions radially outwards.

2. The arbitrary distance to find the mean was a difficult threshold to fix since the stars are of different size. Also since the shape of all the stars is not circular, the mean in different directions was not found to be in a range which can be fixed.

3. This technique was tried basically to differentiate between true stars and scale. But since the location of the drastic fall in the intensity in the case of scale cannot be pin-pointed exactly, this technique failed to give any significant results.
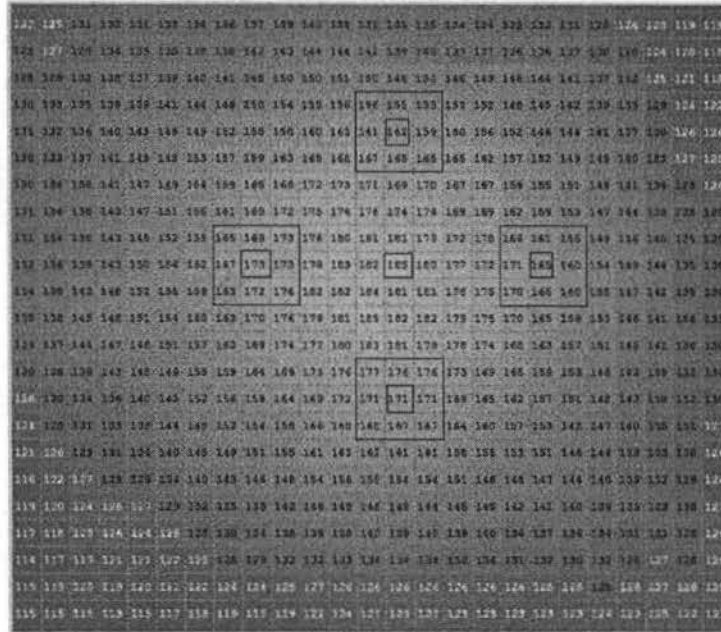
Figure 3.4. Enlarged image of a candidate starry MLC

3.4.2. Application of Row Vectors to Calculate Characteristics of a Star. Now

consider Figure 3.4. The intensity of the pixel on the center row of a starry MLC goes on

increasing from the first column till the center column and then starts decreasing. When

this center row is plotted, a peak-like structure at the center is obtained as shown in

Figure 3.5 and the shape of the graph, i.e. the rise time, fall time, the difference between

the peak and the surrounding, etc., can be used to statistically define a star.

The rise time and the fall time are calculated from the outermost pixel to the

center pixel along the center. So the number of pixels over which it is calculated varies

for different stars since the size of the star varies. Also the standard deviation and

variance is calculated over the entire starry MLC. The size of the star is not constant and

hence the parameters calculated till now may or may not be used as a useful parameter

for differentiating.
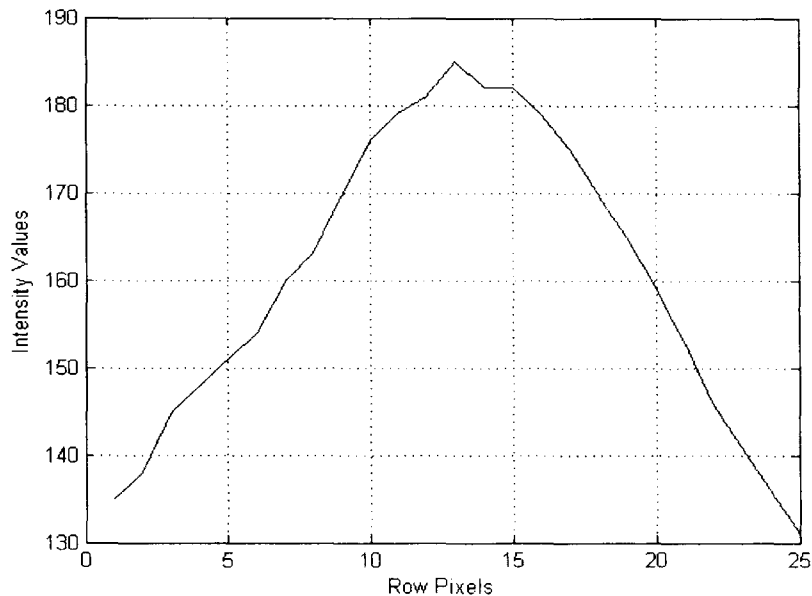
Figure 3.5. A graph of intensity values plotted across a starry MLC through the center pixel

To solve this problem, only the central area of 11 by 11 pixels of all the starry

MLC candidates was selected for calculating the rise time, fall time, standard deviation

and the difference between the mean of the star and its surround. Here the mean is

calculated over the star whose area is 11 by 11. The diameter of the star is 11 and hence

the radius of the star is 5, since the center pixel won't be considered in the calculation of

the radius. Consider two concentric circles where the inner circle of radius 5 pixels

contains the star, and the ring from a 5-pixel radius to an 8-pixel radius is considered to

be the surround of the star. The mean of the surround is calculated over this ring. The

difference between the mean of the star and mean of its surround is also calculated

according to the size of the starry MLC. The size of the surround is fixed as the ring

whose width is 3 pixels wide. For example, for a star of size 17, the ring has an inner radius of 8 pixels and an outer radius of 11 pixels.

From each candidate star, we can collect the following statistics with the help of this row vector technique.

1. Rise Time: The rate of change of intensity from the surrounding intensity value to the maximum central intensity value

2. Fall Time: The rate of change of intensity from the peak central intensity value to the surrounding intensity value

3. Standard deviation

4. Variance

5. Difference between the mean intensity value of the star with the mean intensity value of the surrounding

The rise time and the fall time which are defined as the change in intensity values with respect to the distance in terms of pixel length, give us an idea of how the intensity varies. But in order to get more details, two more important factors were introduced. One was used to check whether the intensity values keep decreasing radially outwards from the center. The second one was the analysis of the way the values decrease. Variance is the average of the squared differences from the mean. Standard deviation is the square root of variance. In simple words, standard deviation is a measure of how spread out, the intensity values are from the mean. Standard deviation is calculated using in-built Matlab function 'std2'. Different parameters like the standard deviation and variance of the intensity values along with the difference between the mean of the star and its surrounding were calculated. In order to get more details about the star, the difference

between each consecutive concentric circle around the central pixel was calculated. This can be explained in detail with Figure 3.6.
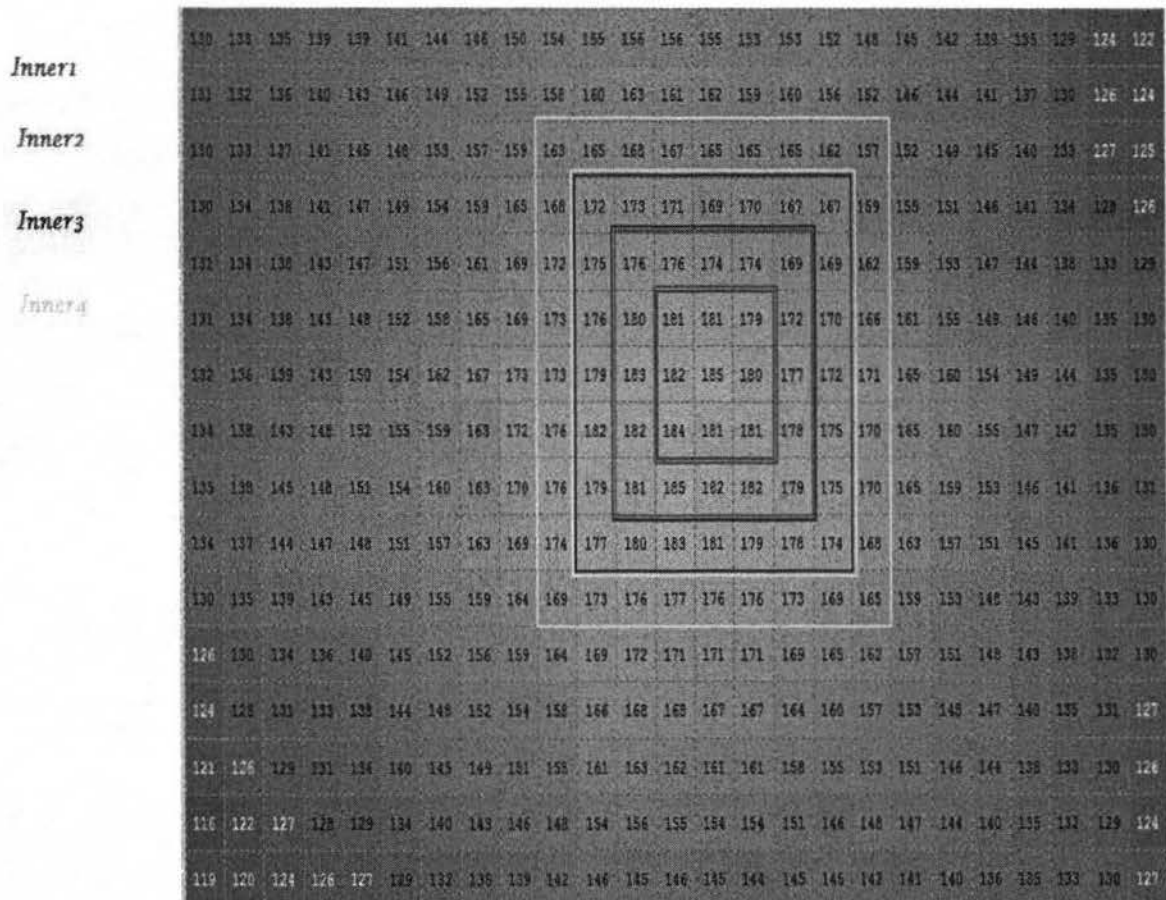
Figure 3.6. Difference between mean of concentric squares

The mean of each ring (here the ring is square in shape rather than circular) is calculated and the difference between the mean of consecutive rings are calculated. In Figure 3.6, each ring is given a name as inner1, inner2 and so on. Hence (inner1-inner2), which is the difference between the innermost ring and the next consecutive rings, is calculated. Similarly (inner2-inner3), and so on are calculated.

Since the appearance of a star is not constant, i.e. the color of the star can be either pink or black or brown, the shape of the star may either be circular or elliptical with different sizes ranging between 11 pixels wide to 30 pixels wide, the orientation of the elliptical star may differ as well and hence the graph of the row vector cannot be generalized and used as a parameter. But definitely the various other parameters listed above give us immense understanding of the starry MLCs and these parameters will be fed to statistical analysis software which will provide us with the details of how significant a particular parameter is to our decision of marking a candidate star as a starry MLC. The list of parameters that are fed to a neural network are given in Section 5.

3.5. CORRELATION COEFFICIENT OF THE CANDIDATE STAR WITH A STAR TEMPLATE

The row vector method gave us very good results of identifying the starry MLCs. But along with the true positive starry MLCs a lot of false negatives are also detected. In order to remove those starry MLC mimics, the candidate stars are correlated with a star template.

3.5.1. Correlation Coefficient and its Significance. Correlation is used extensively for matching. Let's consider $f(x, y)$ to be an image which consists of different objects or regions. If we want to check whether f contains a particular object which is of interest, and let $w(x, y)$ be that object, then if there is a match, i.e. if object w is present in f, the correlation will be maximum at the location where w finds a correspondence in f. Matching is performed using the correlation coefficient which is defined as follows:

$$\gamma(x,y) = \frac{\sum_s \sum_t [f(s,t) - \bar{f}(s,t)] \, [w(x+s,y+t) - \bar{w}]}{\{\sum_s \sum_t [f(s,t) - \bar{f}(s,t)]^2 \, \sum_s \sum_t [w(x+s,y+t) - \bar{w}]^2\}^{1/2}}$$

where, x = 0, 1, 2, 3......., M-1, y = 0, 1, 2..., N-1, M*N is the size of f, $\bar{w}$ is the mean value of pixels in w and $\bar{f}$ is the mean value of the region that is coincident with the current location of w. All the summations in the above equation are taken over the coordinates that are common to both f and w. The calculated correlation coefficient $\gamma(x,y)$ is scaled such that the maximum value is 1 and the minimum value is -1. The correlation coefficient is independent of the changes in the amplitude of f and w [20].

If the candidate star is very similar to the star template, then the correlation coefficient between the two will be close to 1, whereas if we correlate two very different objects then the correlation coefficient will be close to 0. There has to be a threshold to differentiate between true starry MLCs and the mimics. This correlation coefficient threshold is dynamic and is determined depending upon the number of candidate stars that are above a certain threshold. The lower bound for the correlation coefficient was set to 0.7 and the upper bound was 1. The correlation coefficient threshold set for each image is different and is dynamically calculated depending upon the total number of candidate stars that have a correlation coefficient value in a particular range. The entire range of coefficient values from 0.7 to 1 is divided into 15 sections and the best threshold is selected in such a way that a maximum of true positive starry MLCs and a minimum number of false positives are selected for further processing. The above idea can be explained with the help of an example: Consider that before correlation, there are 100

candidate stars in an image. Now all the candidate stars are correlated with three differently sized ideal stars and the maximum correlation coefficient for each candidate is noted. Consider the case where number of candidate pixels having a correlation coefficient above 0.7 is 86, number of candidate stars having correlation coefficient above 0.8 is 32 and the number of candidate stars having correlation coefficient above 0.9 is 3. In such a case, since we don't want to discard true positive starry MLCs but at the same time we wish to reject the maximum number of false positives, the correlation threshold for this particular image is set at 0.8. This correlation coefficient threshold is selected such that the number of candidate stars above this selected threshold is the maximum number found that is still less than 50. Thus the number of candidates selected for further processing is always less than 50, with the correlation coefficient being selected on the basis of the number of candidate stars above each threshold value ranging from 0.7 to 1 in steps of 0.025.

3.5.2. Correlation with Different Star Templates. Firstly, since the size of starry MLCs range from 11 pixels in diameter to more than 25 pixels in diameter, it was decided that each candidate star will be correlated with three star templates of sizes 11 by 11, 17 by 17 and 23 by 23. The size of the star is determined based on the correlation coefficient of the candidate star with the differently sized star templates.

The star template has to be chosen in such a way that it resembles the characteristics of an ideal star perfectly. Five different options were tried and the results of each of them were stored for further analysis. The various options to select a star template for each size are as follows:

1. Select the star template which has the most ideal characteristics from amongst the various starry MLCs in the medical images.

2. Study the various characteristics of starry MLCs and with the help of all the data that has been obtained; create a synthetic star template which encompasses the various characteristics of the true starry MLCs.

3. To create a star template, all the starry MLCs of the same size from various images were averaged and the resulting averaged image was considered as the star template.

4. Since the orientation of all the starry MLCs can be different, before creating the averaged star template, all the starry MLCs from the various images are rotated and summed so that the orientation does not create differences.

5. Since we know that the intensity of starry MLCs decreases radially outwards, contrast enhancement was applied on the averaged rotated star template which will be used for correlating with the candidate starry MLCs.

Out of the above five star templates in consideration, all the options were tried and it was found that the second option, i.e., the synthetic star template, gives good results for only certain starry MLCs. It was found out that, creating synthetic star templates from scratch with the help of the available data was a gargantuan task and can lead to errors if all the characteristics of the stars are not considered properly. So the synthetic star template was dropped.

The option to select the desired star template is given to the user at the start of the program. Figure 3.7 shows a few examples of star templates.
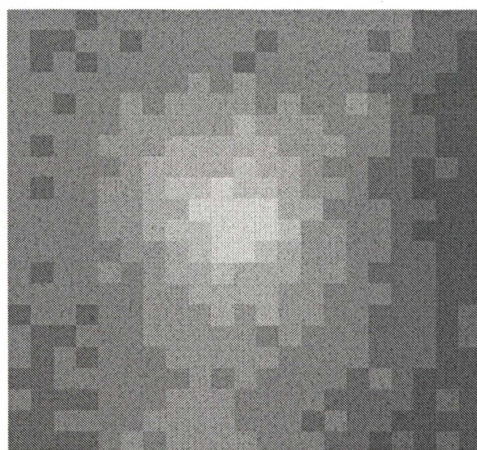
(a) Ideal star template of size 11 by 11
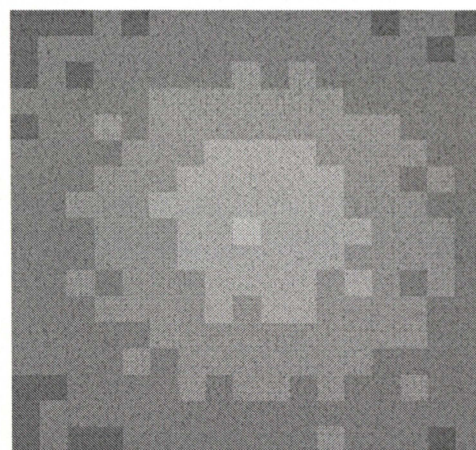
(c) Ideal Averaging star template of size 17 by 17
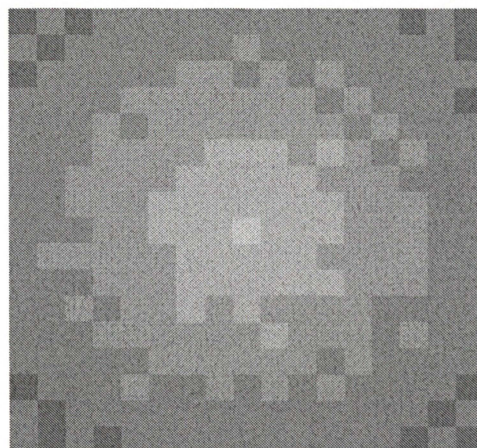
(b) Ideal averaging rotated star template of size 17 by 17
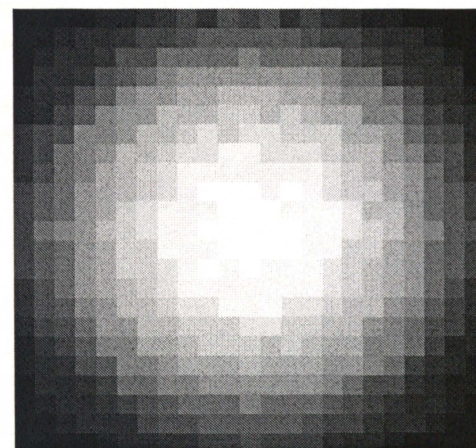
(d) Equalized averaging rotated star template of size 23 by 23

Figure 3.7. Illustration of different types and sizes of star templates

A set of images were run for different star templates and it was found that the ideal star template selected from amongst the various starry MLCs from the images, the averaged star template and the rotated averaged star template all gave very similar results as can be seen from the images shown below in Figure 3.8. Figure 3.8 shows the output after correlating the candidate stars of BF092908per131_CNP from Figure 3.3 (b) with four different types of star templates.
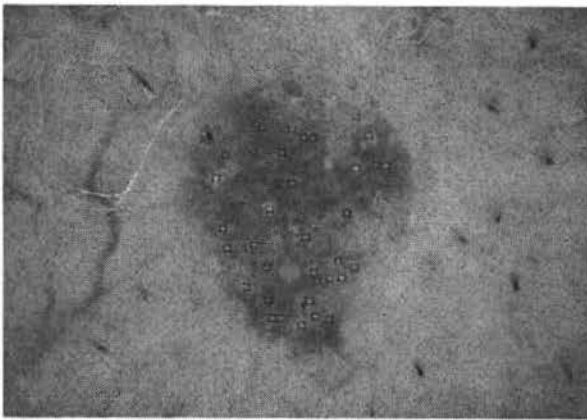
(a) Output after correlating with ideal star templates
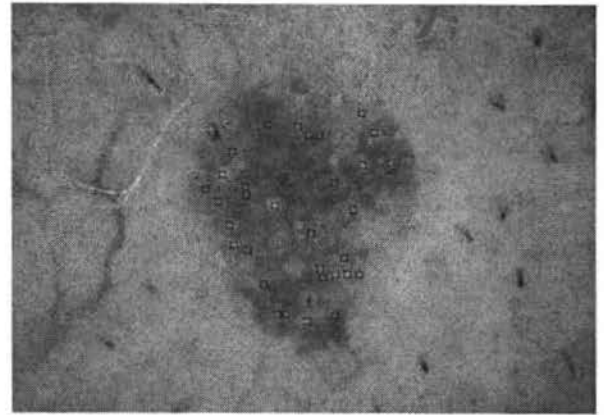
(c) Output after correlating with ideal averaging rotated star templates
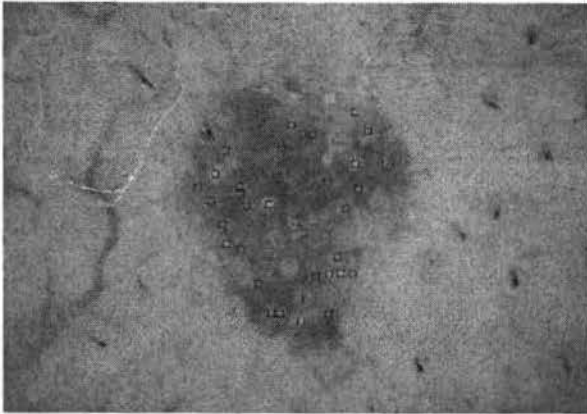
(b) Output after correlating with ideal averaging star templates
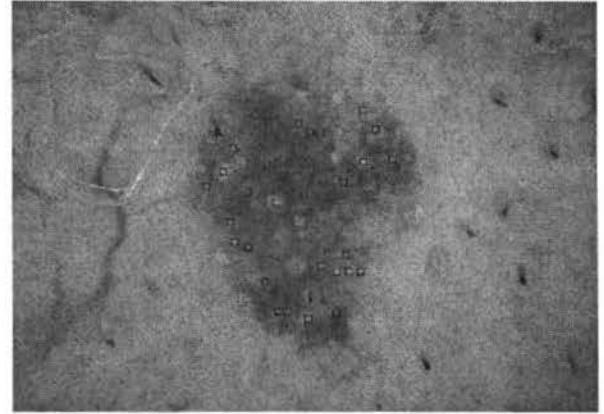
(d) Output after correlating with equalized averaging rotated star templates

Figure 3.8. Output showing the candidate starry MLCs after doing correlation with (a) Ideal star template, (b) Ideal Averaging star template, (c) Ideal averaging rotated star template and (d) Equalized ideal averaging rotated star template

The number of candidate stars selected in each of the above cases, were very similar. The output from the equalized star template was a little different from the other three star templates with respect to a few images. The difference was in the number of candidate starry MLCs selected for further analysis.

The code is written in such a way that it selects the size of the star template depending upon the candidate star. Each candidate star is correlated with the selected star template of different sizes and the size of the star is determined by the star template with which the correlation coefficient is highest.

3.6. APPLICATION OF DIFFERENCE BETWEEN MAXIMUM AND MINIMUM PIXEL VALUE IN AN 11 BY 11 STAR AS A THRESHOLD

All starry MLCs have a bright pixel at the center and then the intensity goes on decreasing. But the rate of decrease is not constant in different sized stars. So instead of taking the difference between the maximum and minimum pixel intensity in stars of different sizes, the difference between the maximum and the minimum pixel intensity considering only the 11 by 11 size of star irrespective of the actual size of the starry MLC will give us good threshold to differentiate between true starry MLCs and mimics. Since the pores have a similar characteristic as the starry MLC, but with the rate of change of intensity being low as compared to starry MLCs, this particular parameter can differentiate true starry MLC from pores.

From the 66 images of seborrheic keratoses that were considered as the training set, the data regarding the maximum and minimum pixel intensity considering only the 11 by 11 size of star irrespective of the actual size of star was analyzed and thresholds for different sized stars were calculated in such a way that a very few of the true starry MLCs are ignored for further processing. The thresholds are given in Table 3.1. The intensity values of the blue plane of the image are in the range of zero to one, where zero

corresponds to black and one corresponds to white. When the intensities are considered,

normally the range is from 0 to 255, but in this project the range is converted to 0 to 1.

Table 3.1. Difference between maximum and minimum pixel value threshold for
different size of starry MLC

Size of starry MLC	Difference between maximum and minimum pixel intensity in 11 by 11 size of star
11	0.14
17	0.075
23	0.03

For a particular test run, in which the image considered was

'BF092908per131_CNP' and the correlation method selected was 'Ideal star templates',

the output after applying difference between maximum and minimum pixel intensity

value threshold is as displayed in Figure 3.9.

As can be seen from Figure 3.9, the image that was obtained after the first

correlation had 34 candidate stars whereas the image that was obtained after applying the

threshold for difference between the maximum and minimum pixel intensity for the 11 by

11 size of star, has only 10 candidate pixels which includes all four true starry MLCs in

the image. The difference between max and min pixel intensity value as a threshold was

applied to all the 100 images (66 images of seborrheic keratoses and 34 images of

malignant melanoma) in the training set and it was found that almost none of the true

starry MLCs are rejected.
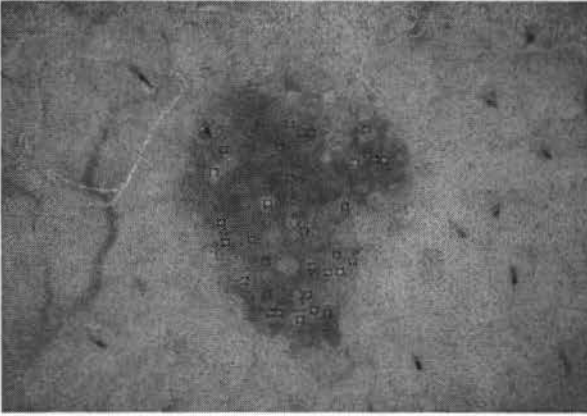
(a) Output of correlation coefficient method
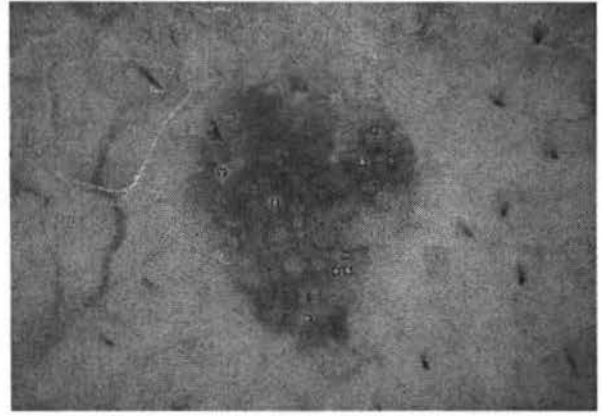
(b) Output after applying difference between max and min pixel value

Figure 3.9. Application of difference between maximum and minimum pixel value threshold. The input figure given to this method is the output of the correlation coefficient as shown in (a), the figure shown in (b) is the output after applying difference between max and min pixel value

4. SHAPE ANALYSIS OF CANDIDATE PIXELS

4.1. SHAPE OF STARRY MLC

It was determined that the drop in the intensity values as we go radially outwards was not constant for all starry MLCs. Also another important factor was that, in the same starry MLC, the rate at which the intensity values decrease is not constant in all directions from the center of the starry MLC. This can be used to find the shape of the star which will give us more insight into the understanding of the starry MLCs.

Various methods were used to find the shape of the star. The scales are extremely bright and hence the difference between the intensity values of the pixels in the scale and pixels outside the scale can be used to differentiate scales from true starry MLCs. But there was no consistency with this idea. Another idea was to see the intensity value of a 3 by 3 neighborhood at equidistant areas from the center pixel. But again this idea was not good since the shape of the star and the orientation of a star were unknown.

In order to find the shape of the star, we need a smooth image and hence the candidate star is filtered using a 3 by 3 mean filter by a sliding neighborhood operation. The smooth image has a constant rate of drop in intensity as compared to the actual star template. Figure 4.1 (a) shows an enlarged version of starry MLC and Figure 4.1(b) shows the starry MLC after passing it through a 3 by 3 mean filter using a sliding neighborhood operation. It can be clearly seen that the change in intensity in Figure 4.1 (b) is very smooth as compared to Figure 4.1 (a).
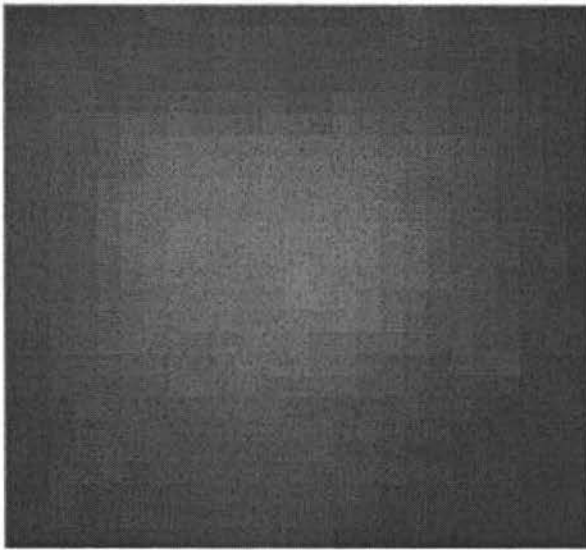
(a) Original starry MLC   (b) Smooth version of starry MLC using
3 by 3 averaging filter
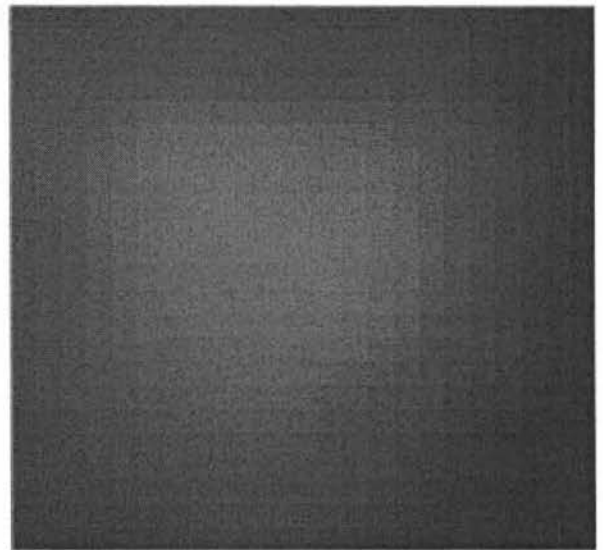
Figure 4.1. The enlarged and smoothed version of a starry MLC. (a) The enlarged version
of a starry MLC and (b) the smooth version of the same starry MLC after smoothing it
with a 3 by 3 averaging filter

While trying different ways to find the shape of the star, it was found out that the

histogram equalized high contrast image gives a very good idea of the shape of the star.

Figure 4.2 shows the histogram equalized image of the starry MLC which is shown in

Figure 4.1 (a). The histogram equalized image is a high contrast image. If the range of

intensity values of the shape of the star is between 0.4 and 0.7, then the histogram

equalized image will have intensity values in the complete range between zero and one.

The histogram equalization was used to just get an idea of the shape of the star. All the

further processing to find the shape of the star is done with the help of the averaged shape

of the star template and not the histogram equalized image. In this research, the shape of

the star is calculated with the averaged shape of the star template, but the histogram

equalized image can also be used to find the shape of the star.
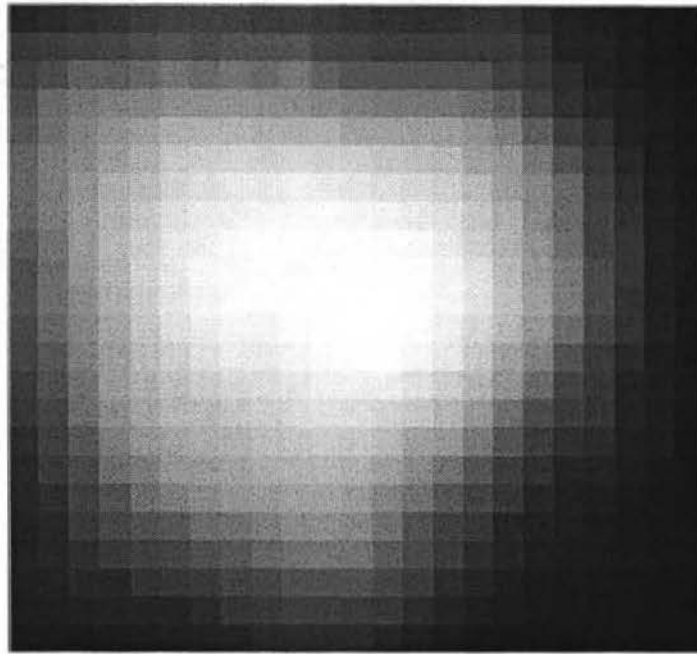
Figure 4.2. Histogram equalized image of a starry MLC

Once we obtain a smooth star template, depending upon the maximum and minimum intensity of pixel values in the star template, a closed figure is obtained which loosely represents the shape of the star. The shape of a star is calculated using the intensity values of the star candidate. Let's assume an ideal star in which the starry MLC has a bright center and then the intensity keeps on decreasing as we move away from the center at a constant rate in all directions. Let's consider that the brightest pixel in the star has an intensity value of 0.6 and the darkest pixel has a value of 0.4. Let's consider the size of the star to be 21 by 21 and the drop in intensity is 0.01 for every pixel width moving away from the center. If the outline of all the pixels that have a value in the range of 0.48 to 0.52 is drawn, a ring of width 5 pixels wide will be obtained. This is because we have assumed an ideal star which is circular in nature. After this, the closed figure is spurred and thinned in order to get a single pixel wide shape of the star. Spurring and

thinning is done with the help of the inbuilt function 'bwmorph' in the Image Processing Toolbox of Matlab. In the actual program, the mean of the maximum and minimum pixel value is calculated and then the pixel values in the range of the mean ± 0.03 are taken into consideration to obtain the outline of the shape of the star. Figure 4.3 (a) shows the closed figure which loosely represents the shape of the star. This is obtained by using the maximum and minimum intensity of pixel values in the star template as discussed above. Figure 4.3 (b) shows the thinned and spurred version of the shape of star. The single pixel width shape of the star is filled using the inbuilt Matlab function 'imfill'. Figure 4.3 (c) is the final shape of star obtained after filling the thinned and spurred version of shape of star.

While using this method, sometimes the ring that is formed is not closed. Consider that a star candidate is actually of size 13 by 13. It is correlated with three star templates of size 11, 17 and 23. That particular star candidate will have highest correlation coefficient with the star template whose size is 11. Now since the actual size of the candidate star is greater than the size that we assume from the correlation coefficient values, the ring that is formed may be incomplete and hence can't be filled fully. In order to solve this problem, a bigger size of the star candidate was considered for further calculations. The decision to consider a bigger size of the star candidate can be based on whether the final filled shape of star is complete or incomplete. Mathematically it can be decided by calculating the roundness i.e. the ratio of the square of the perimeter of the shape of the star with the value of 4*Pi*Area of the shape of the star.

$$Roundness = \frac{Perimeter^2}{4 * Pi * Area}$$

Roundness $\cong$ 1 *for completely filled shape of the star*

Roundness $\gg$ 1 *for incompletely filled shape of the star*
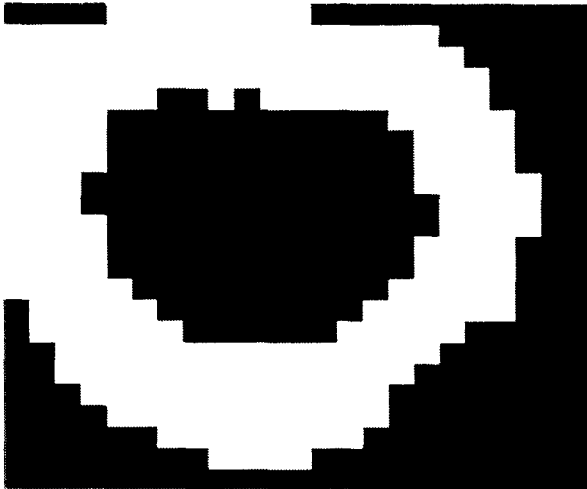
(a) Loose representation of the shape of star
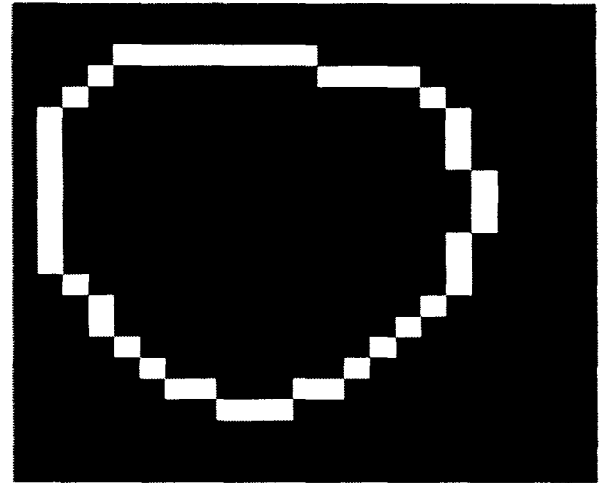
(b) Thinned and spurred version of the shape of star
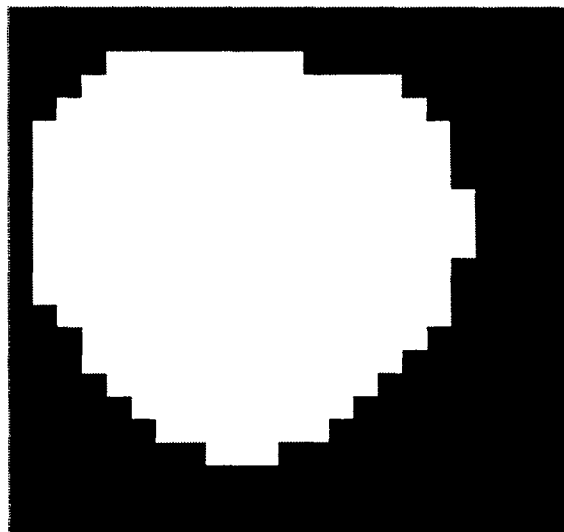
(c) Final shape of star

Figure 4.3. Various stages in the development of shape of the star. (a) Shows the loose representation of the shape of star, (b) shows the thinned and spurred shape of star and (c) shows the final shape of star after filling the thinned shape of star

In either case whether the shape of the star is filled or not, the perimeter of the object will remain the same whereas the area of the object will differ drastically because the area of an object is calculated depending upon the number of pixels in the object. If the shape of the star is not filled completely, then the ratio of (perimeter^2) / (4*pi*area) will be significantly greater than one, whereas if the shape of the star is filled, then the ratio of (perimeter^2) / (4*pi*area) will be close to one. Due to this, it is easy to differentiate whether the shape of the star is filled completely or not. Solidity is another parameter which is used to check whether the shape of the star is filled completely or not. The shape of the star is filled completely if the candidate star is a true starry MLC and the shape of the star is either incomplete or will have a tail-like structure if the candidate star is a mimic of a starry MLC. Solidity is the ratio of the area of the binary shape of the star to the convex hull area. Hence for a true starry MLC, solidity will be close to one, whereas solidity for a mimic of a starry MLC will be much less than one.

The shape of the star is developed for all candidate stars. The true starry MLCs and mimics of starry MLCs can be differentiated using the shape of the star. For true starry MLCs, the shape of the star will be complete, whereas the shape of the star for mimics such as keratin scales, skin pores and bubbles will be either complete in case the mimic very closely represents the true starry MLC or else the shape of the star will be incomplete or will have tail-like structures. The shape of the star of a mimic of a starry MLC will be incomplete or will have a tail-like structure as shown in Figure 4.4, since the drop in intensity is uneven in a certain direction. The complete process of developing the shape of the star for a true starry MLC is explained with diagrams in Figure 4.5. The same process for a mimic is explained with diagrams in Figure 4.6.
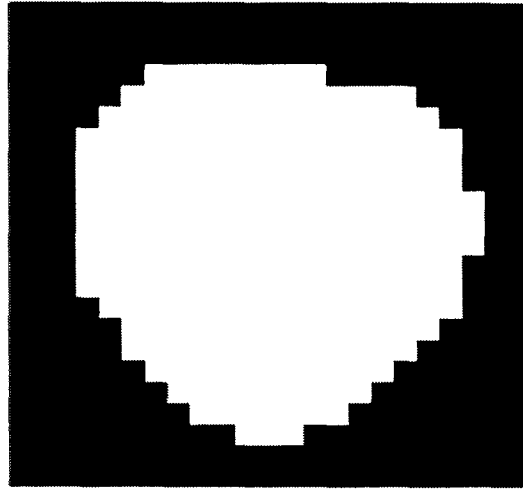
(a) Final shape of the star for a true starry MLC
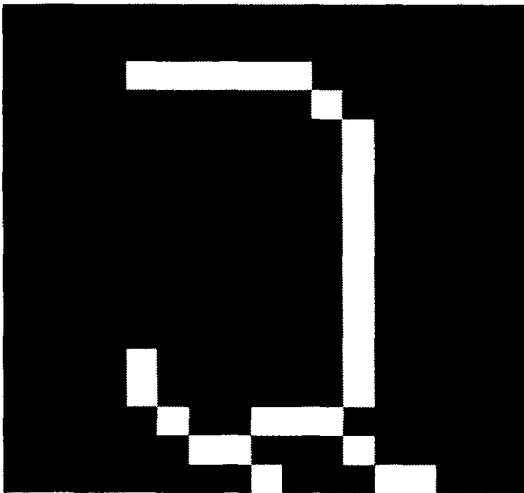
(b) Incomplete final shape of the star for a mimic of true starry MLC
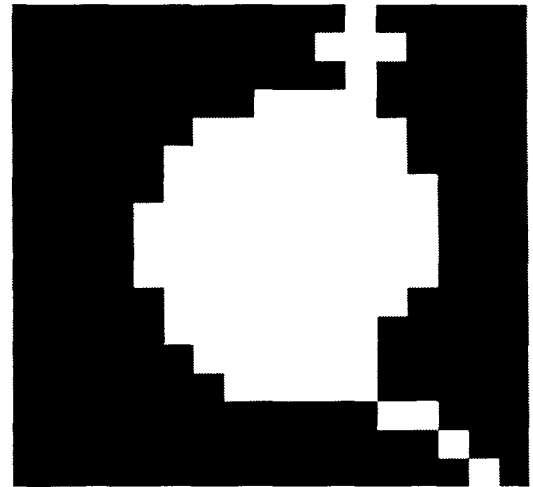
(c) Final shape of the star with a tail like structure for a mimic of true starry MLC

Figure 4.4. (a) Shows the shape of the star for a true starry MLC, (b) and (c) show the shape of the star for a mimic of starry MLC

The visible difference in the shape of the star of a true starry MLC and a mimic is captured with the help of the roundness and solidity.
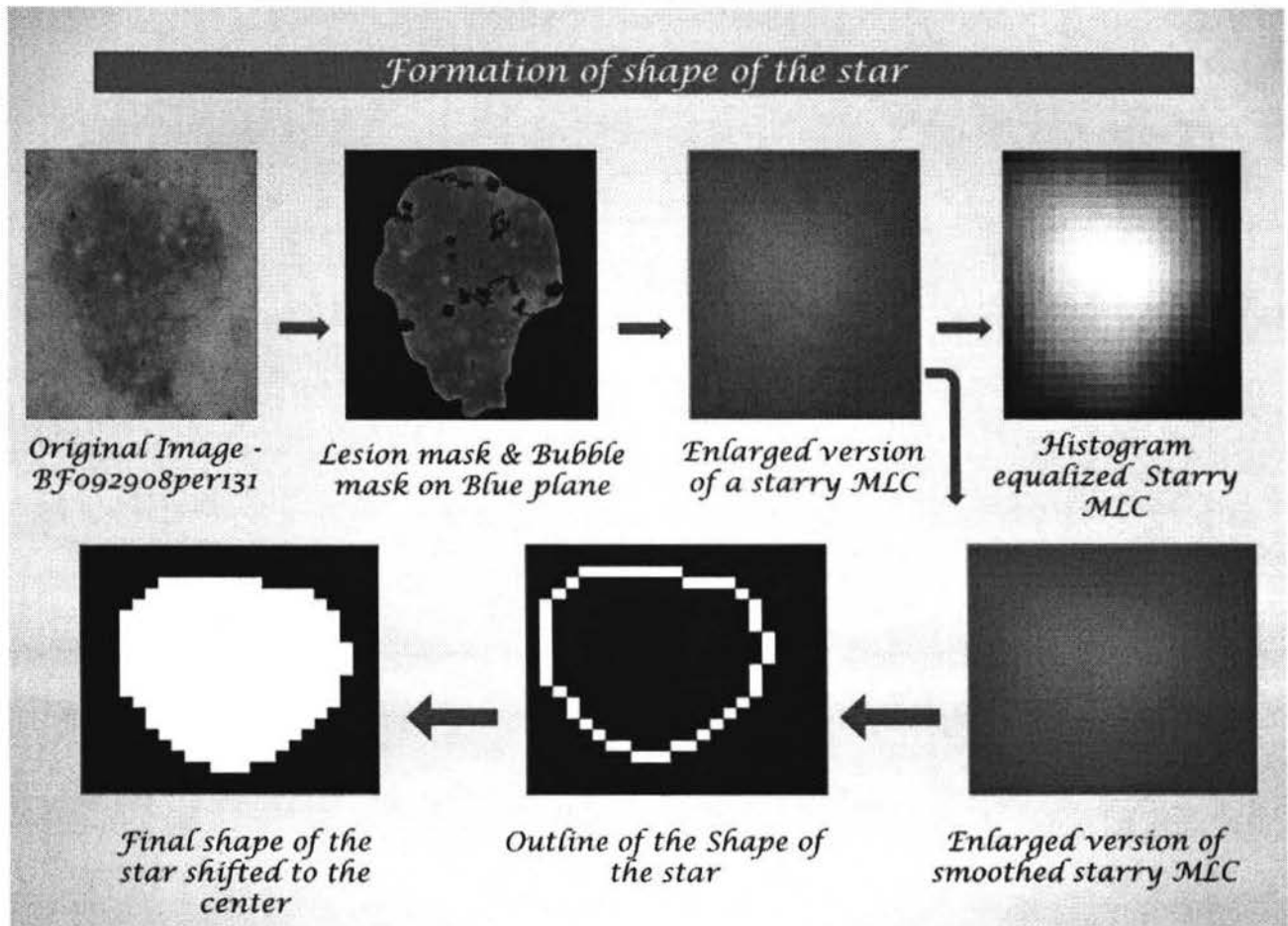
Figure 4.5. Formation of shape of the star for a true starry MLC taken from image BF092909per131_CNP from the seborrheic keratoses set

Once we have filled the shape of star, we can calculate the various properties such as the area, perimeter, orientation, diameter, major axis and minor axis. With the help of these calculations we can find out the actual shape of the star. If the major axis is equal to the equivalent diameter, then we can be sure that the candidate star closely resembles a circular ideal star, whereas if the difference between the major axis and the minor axis is greater than two pixels, then the star candidate will be considered as an elliptical starry MLC. All these calculations of the shape of the star are considered for the formation of star template used to correlate with the shape of the star as discussed in Section 4.2.
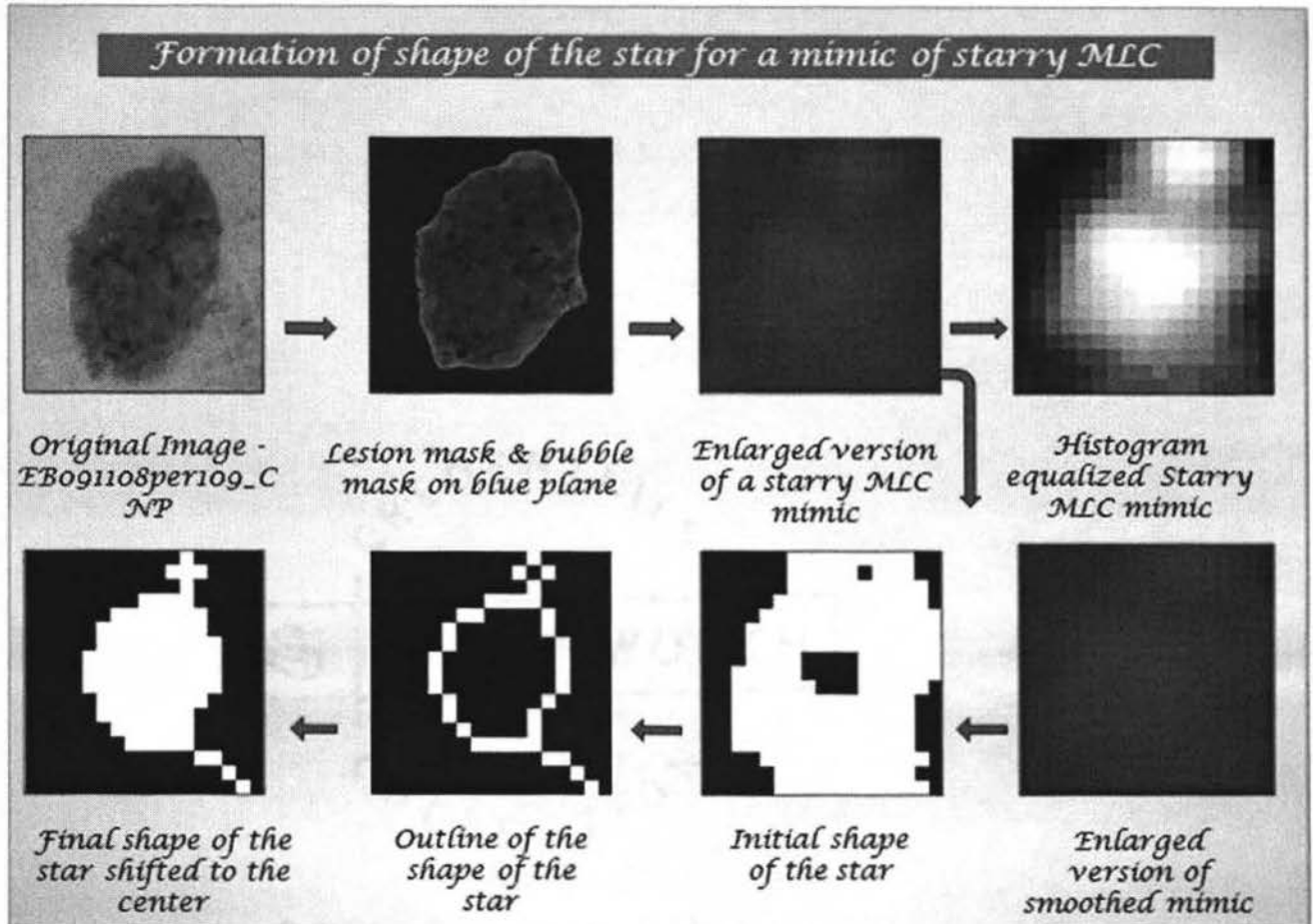
Figure 4.6. Formation of shape of the star for a true starry MLC taken from image
EB091108per109_CNP from the seborrheic keratoses set

4.2. SECOND CORRELATION WITH SHAPE OF THE STAR

Once the calculations are done and the shape of the star candidate is certain, then

with the help of the measurements made a new binary star template is created which is

the ideal star template for those measurements. Let's assume that the shape of the star is

elliptical in nature. The major axis and the minor axis of the shape of the star have been

calculated. With the help of this data, an elliptical binary star template is created whose

major and minor axes are the same as those of the shape of the star. Since the star

template created has the same axis length, area, perimeter, eccentricity and other

parameters, the appearance of the created shape of the star template will also be almost equal to that of the shape of the star. Similarly, if the shape of the star is circular in nature, with the help of the equivalent diameter a binary star template is created whose parameters diameter, area, perimeter, eccentricity, etc. will be almost equal to that of the shape of the star.

The original star candidate is correlated with the ideal star template that is created to see if the filled shape of the star is a true starry MLC or if it is a mimic of a starry MLC. Since the size of star is small, i.e. in the range of 11 by 11 pixels, the correlation coefficient may vary immensely even for small changes in the shape of the star. To overcome this problem, the filled shape of the star is scaled and similarly the star template formed is also scaled equivalently and then correlated. The scaling factor is different for different sizes of the candidate star. If the size of the star candidate is 11 by 11, then the scaling factor used is six. If the size of the star candidate is 17 by 17, then the scaling factor is four, and if the size of the star candidate is 23 by 23, then the scaling factor is three. The scaling factors are decided in such a way that the scaled template is almost equal irrespective of the actual size of the star candidate. Figure 4.7 shows the shape of the star which is the actual shape of a starry MLC from an image. The ideal star template along with the scaled and oriented version of the same is displayed in Figure 4.7.
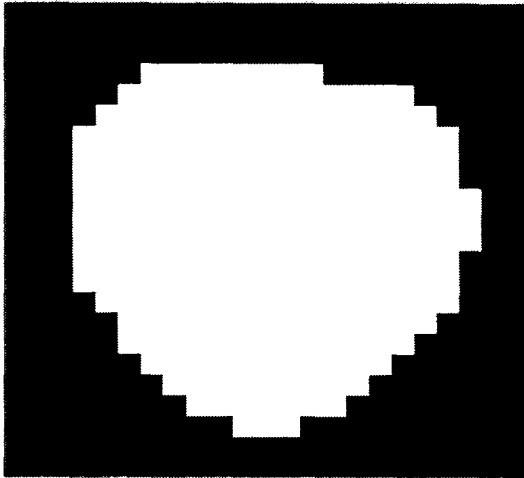
(a) Final shape of star shifted to the center
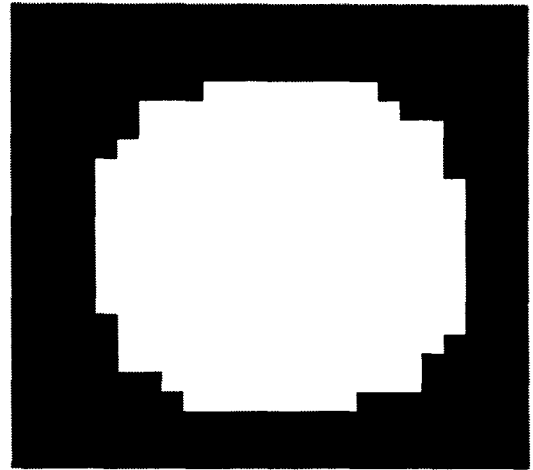
(c) Oriented ideal star template
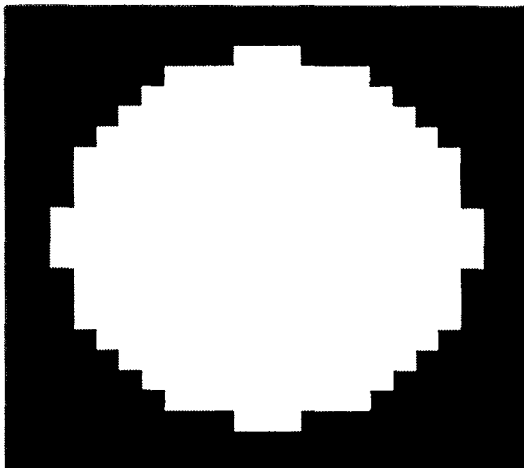
(b) Ideal star template of the same size and shape as the final shape of star

(d) Oriented scaled ideal star template

Figure 4.7. Second correlation with shape of the star. (a) is the filled shape of starry MLC shifted to the center, (b) is the ideal star template formed from the data obtained from the shape of starry MLC, (c) is the ideal star template having an orientation angle equal to that of shape of starry MLC and (d) is the scaled version of the oriented ideal star template

The ideal shape of the star template is created using various calculations done on the shape of the star. It was found out that the correlation coefficients in both scaled and un-scaled cases are almost the same. If the filled shape of the star is incomplete or if it has a tail, then that can be used to differentiate true MLCs from the mimics. Mathematically this differentiation can be done with the help of two parameters namely solidity and the circular condition. A convex hull is the smallest convex polygon that can contain the entire region of the filled shape of the star. Solidity specifies the fraction of pixels in the convex hull that are also in the filled shape of star [21].

$$S = \frac{A}{A_c}$$

where,

S = Solidity

A = Area of the filled shape of star

A_c = Area of Convex Hull

The circular condition is given as follows:

$$\frac{P^2}{4 * Pi * A} = 1$$

where,

P = Perimeter of the filled shape of star

Pi = 3.142

A = Area of the filled shape of star

If the shape of the star has a tail-like structure, which is the case for scales, solidity is much lesser than one, whereas for a true starry MLC, the shape of the star does not have any tail-like structure and the solidity is very close to one. The circular condition is used to check whether the shape of the star is completely filled or not. If the value of the circular condition is significantly greater than one, then it can be concluded that the shape of the star is not filled completely.

4.3. CORRELATION WITH A HISTOGRAM EQUALIZED STAR TEMPLATE

Out of the different mimics like bubbles from the dermoscopic fluid, keratin scales and skin pores, we were able to remove a lot of bubbles by using a bubble mask. We were also able to reduce the number of mimics wrongly identified as starry MLCs by using various methods as described in the previous sections. The skin pores are the toughest mimics of starry MLCs. Many of the pores are very similar to the starry MLCs in nature, i.e. they too have a bright center and the intensity goes on decreasing radially outwards, the only difference being that the rate of drop of intensity is slightly less than for starry MLCs. This characteristic is not trapped when the various parameters like fall time, rise time etc. are calculated because the difference is very small. In order to make this difference visible, the candidate star can be histogram equalized in order to enhance its contrast and then the equalized star candidate can be correlated with an ideal star template which is also equalized. The ideal star template that is histogram equalized is the same ideal star template that is used for the first correlation.

5. DATA COLLECTION

At every method that was discussed above many properties of the starry MLC are calculated and all these parameters are passed on to a neural network to get the best combination of the various parameters. The various parameters for each candidate starry MLC are as follows:

From the original image, the following data is stored:
- Image name.
- X and Y coordinate of the center of the candidate starry MLC.
- The pixel intensity at the center of the candidate star.

From correlating the candidate star with a star template:
- Correlation coefficient which is the maximum coefficient from correlating the candidate star with three differently sized ideal star templates.

- Size of star.

From the original star candidate:

The different parameters that were calculated from the original star template were calculated using the inbuilt Matlab functions. The size of the star was determined by correlating the candidate star with three differently sized ideal star templates. After determining the size of the star which can either be 11 or 17 or 23, the surround of the star is just a ring having a thickness of three pixels wide whose inner radius is the same as the radius of the star. The mean of the star and the mean of the surround are calculated by taking the average of all the pixel values in the star and in the surround respectively. The standard deviation is calculated using the two-dimensional standard deviation inbuilt

function 'std2' in Matlab. Variance is calculated by taking the square of the standard

deviation.

- Difference between the intensity values of rings around the center pixel of the candidate star of different radii ranging from 1 to 5.

- Rise time and fall time which gives the rate of change of intensity values.

- Mean of the actual size of star (actual star size can be 11, 17 or 23).

- Mean of the actual size of surround.

- Difference between the mean of the star and mean of the surround.

- Standard deviation of the star (actual star size can be 11, 17 or 23).

- Variance of the star (actual star size can be 11, 17 or 23).

- Rise time and the fall time for only the 11 by 11 size of the star (actual star size can be 11, 17 or 23).

- Mean of only the 11 by 11 size of the star (actual star size can be 11, 17 or 23).

- Mean of surround of 11 by 11 size of the star (actual star size can be 11, 17 or 23).

- Difference between the mean of the star and mean of the surround considering the size of the star to be 11 by 11 (actual star size can be 11, 17 or 23).

- Standard deviation of only the 11 by 11 size of the star (actual star size can be 11, 17 or 23).

- Variance of only the 11 by 11 size of the star (actual star size can be 11, 17 or 23).

- Difference between highest and lowest point in the star of size 11 by 11 (actual star size can be 11, 17 or 23).

Parameters calculated from the shape of the star:

All the parameters from the shape of the star are calculated using the

'regionprops' function in Matlab. The function 'regionprops' operates on a black and

white (binary) image in which the white area is considered as the object and the black

area is considered as the background.

- Area: The actual number of pixels in the region.

- Perimeter: The distance around the boundary of the region.

- Centroid: Specifies the x-coordinate and the y-coordinate of the center of mass of the region and is a two element vector.

- Major axis is calculated using the property 'MajorAxisLength' in 'regionprops'. It is gives the value in pixels of the major axis of the ellipse whose second central moments are the same as that of the region.

- Minor axis is calculated in the same way as the major axis is calculated.

- Equivalent Diameter: The diameter of a circle whose area is the same as the area of the region.

- Orientation: Angle between the x-axis and the major axis.

- Eccentricity: Ratio of the distance between the foci of the ellipse and the major axis length of the ellipse. Value of eccentricity always lies between zero and one.

- Solidity: Percentage of pixels in the convex hull that is also present in the region. Convex hull specifies the smallest convex polygon that can contain the complete region.

- Roundness (Circular condition): Perimeter^2 / (4*Pi*Area).

- Major axis / Equivalent diameter.

- Area of shape of the star is filled completely or not

- Shape of the star is circular or elliptical

Parameters from correlating the shape of the star with synthetic star template:

- Second correlation coefficient for scaled shape of star.

- Second correlation coefficient for un-scaled shape of star.

- Correlation coefficient of the candidate starry MLC with an equalized star template.

6. SAS AND LOGISTIC REGRESSION

SAS (Statistical Analysis System) is an assimilated system of many software products provided by SAS Institute Inc. that enables programmers to perform a range of analysis such as data mining, statistical analysis, business planning, forecasting and decision support, project management, data warehousing, remote computing, quality management etc. [22].

This statistical system is one of the most popular statistical software packages available in the market for all kinds of data analysis. The SAS system has in-built modules that are full-fledged applications built around the four basic data driven tasks that are commonly used: data access, data management, data analysis and data presentation. One of the modules is the STAT module which provides wide-ranging statistical capabilities for specialized as well as enterprise-wise methodical needs. The STAT module includes a wide range of statistical tools such as variance analysis, linear and logistic regression analysis, categorical data analysis, survival analysis, psychometric analysis, cluster analysis, matrix manipulation, graphics and graphical interface, multi-dimensional scaling, forecasting etc. In this research project, logistic regression is used for data analysis.

Logistic regression is used in cases where the dependent variable is anticipated to be non-linear with respect to one or more independent variables. There are different types of logistic regression that can be used for analysis. The different types are direct, sequential and stepwise logistic regression. In direct logistic regression, all the variables are entered at the same time into the equation. In this method, there is no change in order

or importance for each predictor variable. Sequential logistic regression is used in cases where there exists a certain order or importance of the predictor values. The stepwise logistic regression is a data screening tool and the decision of whether to include or exclude a particular variable is not that stringent as compared with other statistical methods [23]. The logistic regression model is explained in detail in Section 7.

7. RESULTS

7.1. MAXIMUM LIKELIHOOD ESTIMATES

All the parameters as discussed in Section 5 were passed on to the Statistical

Analysis System. The data was collected from the test set which included 66 images from

the seborrheic keratosis set and 34 images from the melanoma set. The total number of

candidate pixels that were fed to SAS was 1005 candidates from the seborrheic keratoses

set and 473 candidates from melanoma set. The output of SAS is given in Tables 7.1 and

7.2. Table 7.1 shows the analysis of maximum likelihood estimates. Out of the forty

parameters that were passed to SAS, Table 7.1 shows that only columns 4, 5, 12, 15, 17,

19, 37 and 39 are important.

Table 7.1. Analysis of maximum likelihood estimates

Parameter	Parameter name	DF	Estimate	Error	Chi-Square	Pr > ChiSq
Intercept	Intercept	1	-21.6256	1.5429	196.456	< .0001
C4	Size of the star	1	0.1004	0.0202	24.6856	< .0001
C5	First correlation coefficient	1	10.6018	1.9629	29.1704	< .0001
C12	Difference between star and surround	1	21.4616	4.2420	25.5965	< .0001
C15	Mean of 11 by 11 star	1	1.7249	0.5853	8.6837	0.0032
C17	Variance of 11 by 11 star	1	-277.5	87.4929	10.0570	0.0015
C19	Rise time of 11 by 11 star	1	89.7234	16.1604	30.8251	< .0001
C37	Second correlation coefficient scaled	1	2.0984	0.6273	11.1896	< .0008
C39	Equalized correlation coefficient	1	6.7850	1.2547	29.2425	< .0001

The important parameters as indicated by SAS are size of the star, correlation coefficient with an ideal star template, difference between intensity of a star and its surround, mean of the star considering the size of the star to be 11 by 11, variance of the star considering the size of the star to be 11 by 11, rise time of the 11 by 11 size of star, correlation coefficient of the scaled version of the shape of the star with the scaled version of the ideal shape of star and the correlation coefficient of the equalized star with the equalized star template.

7.2. ANALYSIS OF MAXIMUM LIKELIHOOD ESTIMATES

The importance of Table 7.1 can be explained with a little explanation of the Logistic Regression model that has been used in SAS. Logistic regression is nothing but a logistic function whose value is always bound between zero and one like probabilities. The logistic function is given by the following equation: [24]

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

The graph of the function is shown in Figure 7.1. The input is z and the output is $f(z)$. The important characteristic of the logistic function is that the input can take on values anywhere between negative infinity to positive infinity, whereas the output is always between 0 and 1. The variable z represents a set of independent variables and $f(z)$ represents the probability of a particular outcome for those set of variables.

The variable z is defined as follows: [24]

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots\ldots + \beta_k x_k$$
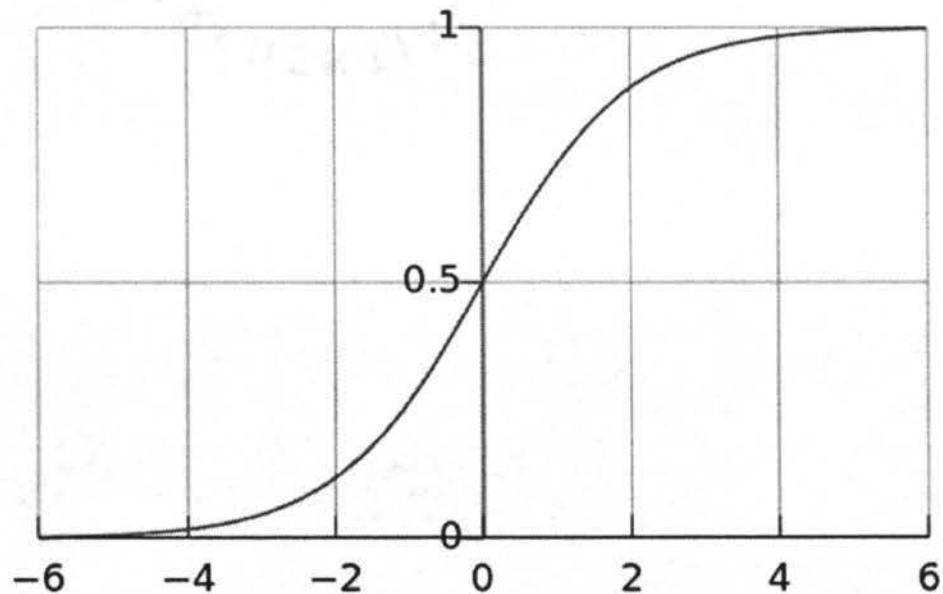
Figure 7.1. Graph of logistic function [25]. z is along the horizontal axis and $f(z)$ is along the vertical axis

β_0 is called the intercept and β_1, β_2, β_3, and so on are the regression coefficients of x_1, x_2, and x_3 respectively. x_1, x_2, x_3 and so on are the parameters displayed in Table 7.1. β_0 is the intercept when all the variables are zero. Each of the regression coefficients defines the contribution of that particular variable. If the value of a regression coefficient is greater than one, then, that particular variable increases the probability of the outcome, whereas if the value of a regression coefficient is negative then that particular variable decreases the probability of the outcome. Also if the magnitude of the regression coefficient is large, then the impact of that variable is very high on the probability of the outcome. Similarly a low magnitude regression coefficient means that the variable has very little influence on the probability of the outcome.

The logistic equation is used to check whether a candidate pixel is a starry MLC or not. This decision is taken after comparing the value of $f(z)$ with a probability level which is selected from the classification Table 7.2 for a specific purpose. For example, one might need the specificity or the sensitivity to be very high. In this research, the correctness of identifying a candidate star as a starry MLC is most important and hence the probability level for maximum correctness is used for comparing the corresponding probability level with the value of $f(z)$. For each candidate pixel, the columns 4, 5, 12, 15, 17, 19, 37 and 39 are calculated. These columns represent the following data:

- Size of the star
- First correlation coefficient with an ideal star template
- Difference between intensity of the star and its surround
- Variance of the star considering only the 11 by 11 size of the star
- Rise time of 11 by 11 size of star
- Correlation coefficient of the shape of the star with an ideal star template
- Equalized correlation coefficient

The value of the intercept is the value of β_0 in the equation for calculating z. The estimate values for each column from Table 7.1 are the values of β_1, β_2, β_3 and so on for the respective column values c4, c5, ..., , c39 which are nothing but x_1, x_2, x_3 and so on. Once 'z' is calculated, the value of z is used to find $f(z)$. This value of $f(z)$ for each candidate star is compared with the probability level corresponding to the maximum correctness value from the classification table. The selection of probability level is highly flexible. Depending upon what is necessary, we can select the probability level. For example the probability level corresponding to maximum sensitivity or maximum specificity can be selected.

7.3. SAS CLASSIFICATION TABLE

Table 7.2 shows the classification table received from SAS. Each row of the table corresponds to a cut-point that is applied to the predicted probability.

Table 7.2. SAS Classification table

Prob level	Event	Non-event	Event	Non-event	Corr ect	Sensi tivity	Speci ficity	FALSE POS	FALSE NEG
0.06	288	549	617	20	56.8	93.5	47.1	68.2	3.5
0.12	265	746	420	43	68.6	86	64	61.3	5.4
0.18	246	856	310	62	74.8	79.9	73.4	55.8	6.8
0.24	226	951	215	82	79.9	73.4	81.6	48.8	7.9
0.3	211	1001	165	97	82.2	68.5	85.8	43.9	8.8
0.36	194	1054	112	114	84.7	63	90.4	36.6	9.8
0.38	189	1062	104	119	84.9	61.4	91.1	35.5	10.1
0.4	185	1074	92	123	85.4	60.1	92.1	33.2	10.3
0.42	*181*	*1087*	*79*	*127*	*86*	*58.8*	*93.2*	*30.4*	*10.5*
0.44	172	1090	76	136	85.6	55.8	93.5	30.6	11.1
0.46	166	1093	73	142	85.4	53.9	93.7	30.5	11.5
0.48	158	1097	69	150	85.1	51.3	94.1	30.4	12
0.54	144	1119	47	164	85.7	46.8	96	24.6	12.8
0.6	127	1133	33	181	85.5	41.2	97.2	20.6	13.8
0.66	106	1146	20	202	84.9	34.4	98.3	15.9	15
0.72	87	1156	10	221	84.3	28.2	99.1	10.3	16
0.78	61	1161	5	247	82.9	19.8	99.6	7.6	17.5
0.84	43	1164	2	265	81.9	14	99.8	4.4	18.5
0.9	23	1164	2	285	80.5	7.5	99.8	8	19.7
0.96	5	1165	1	303	79.4	1.6	99.9	16.7	20.6
0.98	1	1166	0	307	79.2	0.3	100	0	20.8
1	0	1166	0	308	79.1	0	100	.	20.9

The probability values are given in the "Prob level" column, i.e. the first column. The next four columns can be explained with an example. Consider the row having probability 0.42. For this particular cut point, 181 events and 1087 nonevents were correctly classified. The candidate star being considered as a starry MLC is an event and

a candidate star not being considered as a starry MLC is a non-event. On the other hand, 79 events and 127 non-events were incorrectly classified. For this particular probability level, the correct classification rate is (181+1087) / (181+1087+79+127) = 86% which is given in the sixth column. The last four columns sensitivity, specificity, false positives and false negatives give the accuracy of the classification [26].

The correct percentage given by SAS is defined as P (correct classification | total number of predictions), where 'P' represents the probability. Sensitivity is the percentage of occurrences correctly predicted i.e. P (correct | event did occur). Similarly percentage of non-occurrences correctly predicted P (correct | event did not occur) is known as the Specificity. The errors are defined using the false positive and false negative rate. False positive rate is the percentage of predicted occurrences that are incorrect, P (incorrect | occurrence was predicted). False negative rate is the percentage of predicted occurrences which are incorrect, P (incorrect | non-occurrence was predicted) [27].

7.4. OUTPUT OF SAS ON THE TRAINING AND THE TEST SET

The output of SAS was fed into the Matlab code. The selected parameters are multiplied with the corresponding estimates from the classification table and then added. The resultant sum is compared with the probability level corresponding to maximum correctness which is 0.42 in this case. If the resultant sum is greater than 0.42, then the candidate star is considered to be a true starry MLC, whereas if the resultant sum is less than 0.42, then the candidate star is considered as a mimic and is rejected. Then the resulting code was again run on the training set which includes 66 seborrheic keratoses and 34 melanomas. The output is given in Table 7.3.

Table 7.3. Output of SAS on the seborrheic keratoses training set

Image name	TP	FP	FN	TN	Sensi tivity	Speci ficity	Odds Ratio
BF092908per131_CNP	3	0.00 1	3	4	0.5	0.999 75	4000
EB091108per109_CNP	0.0 01	0.00 1	1	8	0.000 999	0.999 875	8
EH050208per75_CNP	9	1	2	21	0.818 182	0.954 545	94.5
EH110508per172_CNP	0.0 01	0.00 1	0.00 1	0.0 01	0.5	0.5	1
FA093008per136_CNP	2	0.00 1	10.0 01	11	0.166 653	0.999 909	2199.78
JH072307per13_CNP	32	0.00 1	3	1	0.914 286	0.999 001	10666.6 7
JJ111207per43_CNP	1	1	1	7	0.5	0.875	7
JP111708per185_CNP	0.0 01	0.00 1	0.00 1	0.0 01	0.5	0.5	1
KS101408per155_CNP	1	0.00 1	2	4	0.333 333	0.999 75	2000
LS083007per18_CNP	0.0 01	2	0.00 1	21	0.5	0.913 043	10.5
LW102708per164_CNP	0.0 01	0.00 1	0.00 1	2	0.5	0.999 5	2000
RS111008per176_CNP	0.0 01	1	0.00 1	5	0.5	0.833 333	5
SW032008per71_CNP	2	0.00 1	1	11	0.666 667	0.999 909	22000
WT101408per151_CNP	0.0 01	0.00 1	0.00 1	5	0.5	0.999 8	5000
WZ082407per16_CNP	0.0 01	1	0.00 1	17	0.5	0.944 444	17
ac103108rol517_CNP	0.0 01	0.00 1	1	0.0 01	0.000 999	0.5	0.001
ad110807rab398_CNP	0.0 01	0.00 1	0.00 1	1	0.5	0.999 001	1000
ah122107rol256_CNP	2	0.00 1	13	4	0.133 333	0.999 75	615.384 6
bb101608rab835_CNP	0.0 01	0.00 1	4	0.0 01	0.000 25	0.5	0.00025
bg072307rab238_CNP	11	0.00 1	13	12	0.458 333	0.999 917	10153.8 5
bl110107rab381_CNP	1	0.00 1	3	9	0.25	0.999 889	3000

Table 7.3. (Continued)

br051107rol75_CNP	0.0001	0.0001	0.0001	14	0.5	0.999929	14000
cb031008rol338_CNPa	0.0001	3	0.0001	21	0.5	0.875	7
cg101507rol236_CNP	1	1	0.0001	13	0.999001	0.928571	13000
db050707rol70_CNP	3	1	0.0001	20.001	0.999667	0.952383	60003
dk072607rab241_CNP	3	0.0001	1	11	0.75	0.999909	33000
ec032408rol341_CNP	0.0001	0.0001	0.0001	18	0.5	0.999944	18000
eg030607tre12_CNP	8	0.0001	2	9	0.8	0.999889	36000
eh071607rolla158_CNP	1	1	0.0001	11	0.999001	0.916667	11000
ew011909rol558_CNP	0.0001	0.0001	3	10.001	0.000333	0.9999	3.333667
fr051208rab704_CNP	2	0.0001	0.0001	5	0.9995	0.9998	10000000
gm091807rol213_CNP	2	5	1	8	0.666667	0.615385	3.2
hf102908rab849_CNP	0.0001	0.0001	0.0001	2	0.5	0.9995	2000
hs061407rolla125_CNP	0.0001	4	0.0001	4	0.5	0.5	1
if030509rol588_CNP	0.0001	1	0.0001	15	0.5	0.9375	15
ip082807kolm11_CNP	3	1	1	18	0.75	0.947368	54
jb071607rol161_CNP	0.0001	1	0.0001	12	0.5	0.923077	12
jc051607rab139_CNP	1	0.0001	1	7	0.5	0.999857	7000
jf101708rab841_CNP	1	0.0001	3	10.001	0.25	0.9999	3333.667
jh112607rab442_CNP	2	0.0001	0.0001	1	0.9995	0.999001	2000000
jm021207rab22_CNP	13	4	3	11	0.8125	0.733333	11.91667
jn091707rol209_CNP	4	1	0.0001	12	0.99975	0.923077	48000
jn091707rol210_CNP	13	2	1	10.001	0.928571	0.833347	65.0065

Table 7.3. (Continued)

js111408rol520_CNP	1	0.001	1	11	0.5	0.999909	11000
kh052207rol90_CNP	1	0.001	2	10.001	0.333333	0.9999	5000.5
kk100708rol499_CNP	0.001	1	0.001	16	0.5	0.941176	16
kk121908rab905_CNP	0.001	0.001	1	8	0.000999	0.999875	8
km022508rol314_CNP	0.001	1	0.001	18	0.5	0.947368	18
ko041307rab104_CNP	0.001	0.001	0.001	12	0.5	0.999917	12000
ld102108rol510_CNP	2	0.001	0.001	29	0.9995	0.999966	58000000
md013009rol563_CNP	0.001	3	0.001	31	0.5	0.911765	10.33333
nm082007rab269_CNP	21	2	2	17	0.913043	0.894737	89.25
oj070609rab1137_CNP	0.001	0.001	0.001	3	0.5	0.999667	3000
rg111708rol521_CNP	3	0.001	3	17	0.5	0.999941	17000
rr090707rab298_CNP	0.001	0.001	2	19	0.0005	0.999947	9.5
rr090707rab299_CNP	1	0.001	1	15	0.5	0.999933	15000
rs062807dru39_CNP	1	0.001	0.001	4	0.999001	0.99975	4000000
sg110807rab399_CNP	0.001	0.001	1	6	0.000999	0.999833	6
td061109dru258_CNP	0.001	0.001	0.001	2	0.5	0.9995	2000
tt070607rab219_CNP	1	0.001	1	10.001	0.5	0.9999	10001
wf040408rol354_CNP	9	1	3	7	0.75	0.875	21
wm110308rab853_CNP	0.001	0.001	1	0.001	0.000999	0.5	0.001
wp022208rol307_CNP	6	0.001	1	12	0.857143	0.999917	72000
wp022208rol308_CNP	6	1	3	14	0.666667	0.933333	28
wp070808rol431_CNP	9	0.001	3	21	0.75	0.999952	63000

Table 7.3. (Continued)

zs052907rol99_CNP	0.0 01	2	0.00 1	16	0.5	0.888 889	8
TOTAL	182 .03	42.0 42	98.0 31	683 .01 1	0.649 966	0.942 015	30.1664 3

The first column in Table 7.3 is the name of the image. Columns 2, 3, 4 and 5 are true positives, false positives, true negatives and false negatives respectively. Columns 5 and 6 give the sensitivity and specificity and the last column is the odds ratio. The output of one of the images of the training set of seborrheic keratoses is as shown in Figure 7.2.
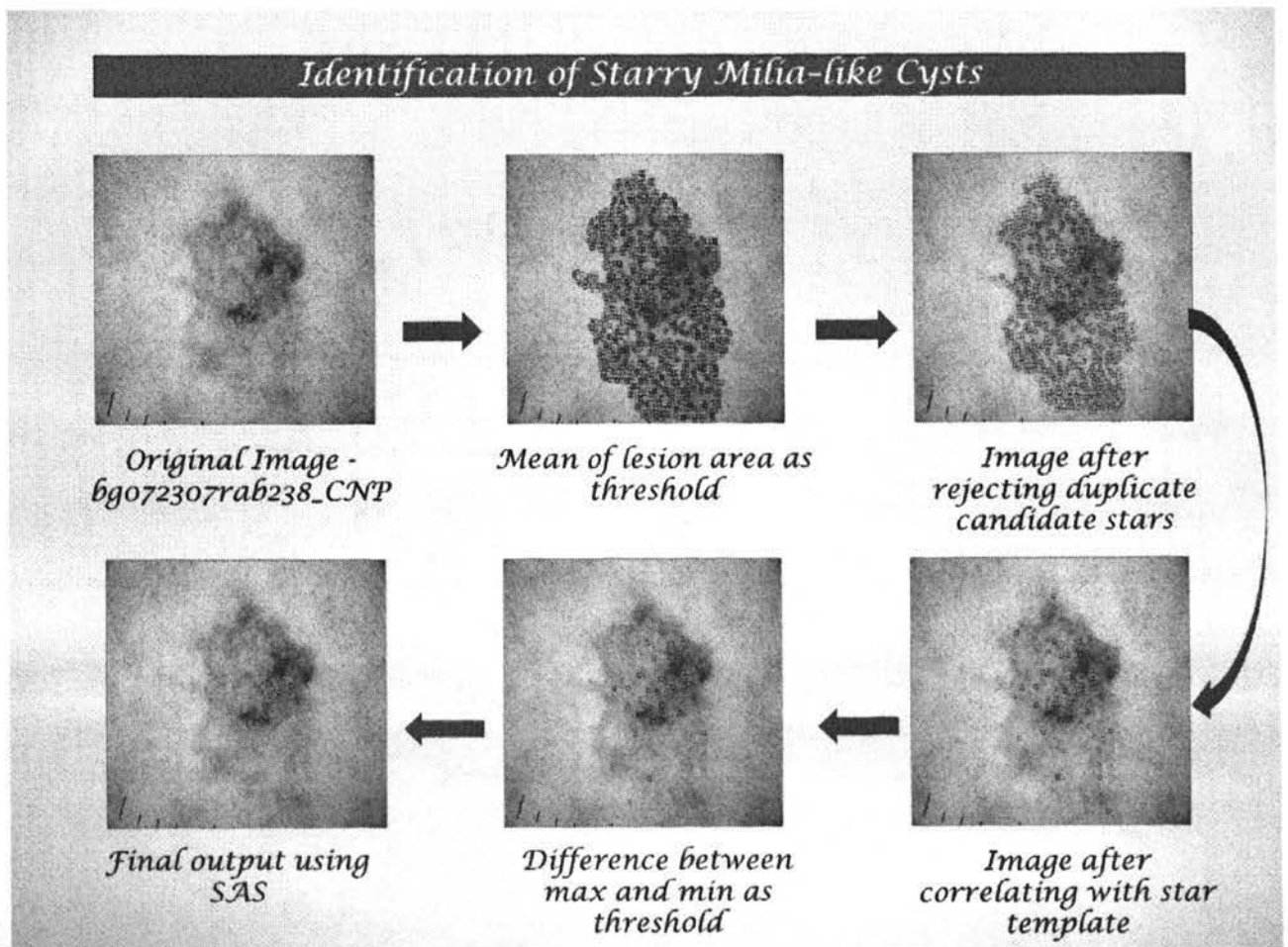

Figure 7.2. SAS output on image bg072307rab238_CNP of seborrheic keratoses set

The output of SAS on the training set of seborrheic keratoses of 66 images is

shown in Table 7.3. For each image, the following observations are made manually:

1. True Positives (TP – Starry MLC correctly identified as an MLC)
2. False Positives (FP – Star mimics identified as an MLC)
3. True Negatives (TN – Starry mimics not identified as an MLC)
4. False Negatives (FN – Starry MLC not identified as an MLC)

Sensitivity is the percentage of occurrences correctly predicted and percentage of

non-occurrences correctly predicted is defined as Specificity. Mathematically, sensitivity

and specificity are defined as follows: [28]

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

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

The last column of Table 7.3 is the odds ratio. In statistical terms, odds of an

event occurring is defined as the probability of the event divided by the probability of an

event not occurring. For example, let's assume a loaded coin that lands on heads 75% of

the times, then the odds of heads is 75/25 = 3. In simple words, odds can be thought of as

how likely will the event occur compared to how likely will the event not occur. For a

fair coin, the probability of a heads or a tail is 0.5 and hence the odds for either a head or

a tail is 50/50 = 1. This means that chances of getting a head is same as getting a tail.

Odds ratio is used when there is a need to compare the odds of an event occurring

to two different groups. It is defined as the ratio of the odds for the first group and the

odds for the second group. The formula for odds ratio is given by the following equation.

$$Odds\ ratio\ =\ \frac{p/(1-p)}{q/(1-q)}$$

where p is the probability for the first group and q is the probability for the second group. The odds ratio can be explained with an example. For women undergoing IVF fertility treatment, the chances of getting pregnant is somewhere around 50% for women whose age is less than 35. Hence the odds for this particular group is approximately one. The second group is a group of women over the age of 40. The likelihood of successful IVF treatment for the women in second group is 20%. Hence the odds for a woman over 40 years of age is 20/80 (20 being the chance of getting pregnant and 80 being the chance of not getting pregnant). Hence the odds ratio is given by (0.5)*(1-0.5) / (0.2)*(1-0.2) = 1 / (0.25) = 4. The odds ratio being four does not mean that women under 35 years of age are four times more likely to get pregnant as compared to women over the age of 40. Instead, women under 35 are 50/20 = 2.5 times likely to get pregnant as compared to women over 40. The odds ratio being four actually means that women under the age of 35 have 4 times the odds (not four times the probability) as compared to women over the age of 40 at getting successfully pregnant by IVF treatment [29]. The odds ratio can also be defined in terms of TP, FP, TN and FN. Also the odds ratio can be defined in terms sensitivity and specificity [30].

$$Odds\ ratio\ =\ \frac{TP/FN}{FP/TN}$$

$$Odds\ ratio\ =\ \frac{sensitivity * specificity}{(1 - sensitivity) * (1 - specificity)}$$

$$Odds\ ratio = \frac{The\ odds\ of\ a\ starry\ MLC\ correctly\ identified\ as\ a\ starry\ MLC}{The\ odds\ of\ a\ mimic\ incorrectly\ identified\ as\ a\ starry\ MLC}$$

In simple words, if the odds ratio is 1, then the odds of the first event occurring with respect to the second is the same. If the odds ratio is positive, then the odds of the first event occurring is higher than the odds of the second event occurring. In finding the odds ratio for the SAS output, all the entries with value '0' were converted to '0.001', so that there is no divide by zero error. Considering all the candidate stars from the 66 images of the training set of seborrheic keratoses, the odds ratio was determined to be 30.20068 which is very high compared to one. Similarly, the odds ratio was found for the training set of melanoma containing 34 images. The odds ratio is given in Table 7.4.

SAS was run on the data from the test set. The test set contained 4 folders, of which two contained melanomas and the remaining two contained seborrheic keratoses. There were 32 images from the melanoma set and 35 images from the seborrheic keratoses set. The odds ratio for each set is given in Table 7.4.

Table 7.4. Odds ratio for training and test set

		Number of images	Odds Ratio
Training set	Seborrheic Keratoses	66	30.20068
	Melanoma	34	11.605
Test set	Seborrheic Keratoses set 1	8	52.0625
	Seborrheic Keratoses set 2	27	35.69957
	Melanoma set 1	6	87
	Melanoma set 2	26	45.90484

Most of the mimics that were identified as a starry MLC were mainly because the mimics are very similar to the actual starry MLC. The other reason being that even after applying the bubble mask, there are still some bubbles left out of which few are considered as starry MLCs.

As can be seen from Table 7.4, the odds ratio for all the sets is greater than one which means that the ratio of odds of a starry MLC correctly identified as an MLC and odds of a mimics incorrectly identified as an MLC is higher than one. From the odds ratio, it is clear that the accuracy of the code in correctly identifying starry milia-like cysts is very high and at the same time the code also differentiates the starry milia-like cysts from competing structures such as small areas of keratin scale, bubbles from the contact dermoscopy fluid interface and tiny pigment pores with high probability.

8. CONCLUSION

It is very important to differentiate benign lesions from malignant lesions in dermoscopy. Benign lesions like seborrheic keratoses are not dangerous whereas malignant lesions are cancerous and hence dangerous. Malignant lesions if not treated in the early stages can become fatal. This research was conducted in order to classify any lesion area into seborrheic keratoses or not, depending upon the existence of a prominent feature, i.e. milia-like cysts. Seborrheic keratoses are characterized by a few features like milia-like cysts, comedo-like openings and network-like structures. This research concentrates on one of the prominent features of seborrheic keratoses, i.e. milia-like cysts. Further our research group differentiated milia-like cysts into starry and cloudy milia-like cysts depending upon the size and the appearance of milia-like cysts. Multiple starry milia-like cysts characterize seborrheic keratoses and hence it is important to identify true starry milia-like cysts from other competing structures like keratin scales, skin pores and bubbles from dermoscopy fluid interface.

In order to identify a certain object in an image, template matching is a very useful tool. In this research, template matching was used to (a) Correlate the candidate star with an ideal star template, (b) Correlate the binary shape of the star with the generated ideal shape of the star and (c) To correlate the histogram equalized candidate star with histogram equalized ideal star template. Various characteristics of a starry MLC like the drop in intensity, size and shape of a starry MLC, etc., were studied in detail and with the help of the data collected, star templates were created for template matching. Also all the data was fed to a statistical analysis software SAS to decide which

parameters are important in deciding whether a candidate star is a starry MLC or not. Out of all the parameters discussed in Section 5, with the help of the logistic regression model which is a statistical tool in SAS software, the size of the star, correlation coefficient of correlating the candidate star with an ideal star template, difference between star and surround, mean of 11 by 11 size of star, variance of 11 by 11 size of star, rise time of 11 by 11 size of star, correlation coefficient of correlating the scaled version of the shape of the star with the generated binary shape of the star template and correlation coefficient of the correlation between histogram equalized candidate star with the histogram equalized star template were found to be the most important parameters. The sensitivity, specificity and the odds ratio were calculated for both the training and the test set. It was found that the odds ratio which is the ratio of odds of a starry MLC correctly identified as a starry MLC to the odds of a mimic of starry MLC incorrectly identified as a starry MLC was significantly greater than one. The higher odds ratio indicates that the program correctly identifies the starry MLCs as starry MLCs and at the same time the degree of differentiation of starry MLCs and the mimics of stars is very high.

9. FUTURE WORK

With the help of SAS, 86% accuracy is obtained in identifying a candidate star as a starry MLC. This accuracy can be increased further if the following steps are implemented:

1. Milia-like cysts are classified as starry MLCs or cloudy MLCs depending upon the size of the MLC. It was decided that the MLCs whose diameter is greater than 30 pixels wide will be classified as cloudy MLCs and all the MLCs whose diameter is less than 30 pixels wide are classified as starry MLCs. In this research project, the size of the star was varied from 11 to 23 in steps of six. Hence, three different sizes of star having diameter 11, 17 and 23 were considered. It is possible that the accuracy will increase if we consider more steps in the range 5 to 31 pixels in diameter, like having sizes 5, 9, 13, 17, 21, 25, 29 and 33.

2. Just like the characteristics of the stars were designed, similarly the characteristics of the mimics of star like pores, bubbles and scales can also be calculated and used to differentiate between true starry MLC and the mimics of stars.

3. It was found that many of the false positives that were identified as starry MLC were bubbles. The bubble mask that was used gave very good results in the sense that most of the bubbles were removed using the bubble mask. But still the bubble mask created is not 100% correct and there is still a possibility to improve the bubble masks.

4. SAS was trained with the help of 66 images from the seborrheic keratoses set and 34 images from the melanoma set. The accuracy of SAS increases if it is trained

using more images. Hence, the accuracy of the code can be increased by training SAS with more images.

5. The FFT can be used just like correlation coefficient for template matching. The FFT can be used to find the frequency of the true starry MLCs. In the complete original image, find for similar frequencies using sliding neighborhood operation. If there is a high correlation between the frequencies of the true starry MLC and the frequency of the small portion of the complete original image, then we have a match and the starry MLC in the image can be identified. The advantage of using FFT is that the process of template matching can be speeded up.

6. The complete research work is concentrated only on the blue plane of the image. Even though just the blue plane in itself is giving an overall accuracy equal to 86%, but still the use of colors can improve the results of the code.

7. Also, for correlation, a square template was considered. Instead of a square template, a circular template can be used. Since starry MLCs are mostly circular in shape, a circular template might give better correlation coefficient as compared to a square template.

BIBLIOGRAPHY

1. Gill D, Dorevitch A, Marks R. "The Prevalence of Seborrheic Keratoses in People Aged 15 to 30 Years," *Arch Dermatol*; 136:759-762, 2000.

2. Tindall JP, Smith JG Jr. "Skin lesions of the aged and their association with internal changes," *JAMA*; 186:1039-1042, 1963.

3. Young AW Jr. "Dermatogeriatric problems in the chronic disease hospital," *N Y State J Med*; 65:1748-1752, 1965.

4. Memon AA, Tomenson JA, Bothwell J, Friedmann PS. "Prevalence of solar damage and actinic keratosis in a Merseyside population," *Br J Dermatol*; 142(6):1154-9, 2000.

5. Yeatman JM, Kilkenny M, Marks R. "The prevalence of seborrhoeic keratoses in an Australian population: does exposure to sunlight play a part in their frequency?" *Br J Dermatol*; 137(3):411-4, 1997.

6. Butler DF, Heymann WR, Quirk CM. "Seborrheic Keratoses," Elston DM (ed), http://emedicine.medscape.com/article/1059477-overview#a0199, last viewed Nov 2011.

7. Herzog C, Pappo A, Bondy M, Bleyer A, Kirkwood J. "Cancer Epidemiology in Older Adolescents and Young Adults 15 to 29 Years of Age, Including SEER Incidence and Survival: 1975-2000 Ch 5," Bleyer A , M O'Leary, Barr R, Ries LAG (eds), *Malignant Melanoma, National Cancer Institute*, NIH Pub. No. 06-5767.Bethesda, MD, 2006. pp 53-63.

8. Jemal A, Murray T, Ward E, Samuels A, et al. Cancer statistics, Q6 2005. *CA Cancer J Clin* 55:10-30, 2005.

9. R. Siegel, E. Ward, O. Brawley, A. Jemal, "Cancer statistics, 2011. The impact of eliminating socioeconomic and racial disparities on premature cancer deaths," *CA A Cancer Journal for Clinicians*, 61(4):212-236, 2011.

10. SEER Stat Fact Sheets: Melanoma of the Skin, Surveillance Epidemiology and End Results, National Cancer Institute, http://seer.cancer.gov/statfacts/html/melan.html, last viewed Nov 2011.

11. Lotze MT, Dallal RM, Kirkwood JM, Flickinger JC. "Cutaneous melanoma," *Principles and Practice of Oncology*, 6e. Philadelphia: Lippincott, 2001.

12. Tan WW. "Malignant Melanoma," Schulman P, Talavera F, Harris JE (eds), http://emedicine.medscape.com/article/280245-overview#aw2aab6b2b2aa, last viewed Nov 2011.

13. Sasieni PD, Shelton J, Ormiston-Smith N, Thomson CS, Silcocks PB. "What is the lifetime risk of developing cancer?: the effect of adjusting for multiple primaries," *Br J Cancer*; 105(3): 460-5, 2011.

14. Statistical Information Team at Cancer Research UK, 2011, http://publications.cancerresearchuk.org/downloads/Product/cs_pdf_melanoma_ja n_2006.pdf, last viewed Dec 2011.

15. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. GLOBOCAN 2008 v1.2, Cancer Incidence and Mortality Worldwide: IARC Cancerbase No.10. Lyon, France: International Agency for Research on Cancer, 2010. Available from http://globocan.iarc.fr.

16. Izikson K, Sober AJ, Mihm MC, Jr, Zembowicz A. "Prevalence of Melanoma Clinically Resembling Seborrheic Keratosis," *Arch Dermatol*; 138: 1562-1566, 2002.

17. Soyer HP, Argenziano G, Chimenti S, Menzies S, Pehamberger H, Rabinovitz H, et al. *Dermoscopy of pigmented skin lesions*, An atlas based on the consensus Net meeting on Dermoscopy 2000. Milan: Era Medical Publishing and New Media; 2001.

18. Braun RP, Rabinovitz HS, Krischer J, Kreusch J, Oliviero M, Naldi L, Kopf AW, Saurat JH. "Dermoscopy of Pigmented Seborrheic Keratosis: A Morphological Study," *Arch Dermatol*; 138: 1556-1560, 2002.

19. Stricklin SM, Stoecker WV, Oliviero MC, Rabinovitz HS, Mahajan SK. "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*, 25(10): 1222–1224, 2011.

20. Gonzalez RC., Woods RE. *Digital Image Processing*. Prentice Hall, Second Edition, 2002.

21. MATLAB version 7.8.0 (R2009a), Image Processing Toolbox. Natick, Massachusetts: The MathWorks Inc., 2009.

22. SAS/ETS 9.2 User's Guide. "Uses of SAS/ETS Software," http://support.sas.com/documentation/cdl/en/etsug/60372/PDF/default/etsug.pdf, last viewed Dec 2011.

23. Logistic Regression in Dissertation and Thesis Research.
http://www.researchconsultation.com/logistic-regression-statistics.asp, last
viewed Dec 2011.

24. Garson GD. "Logistic regression,"
http://faculty.chass.ncsu.edu/garson/PA765/logistic.htm#stepwise, last viewed
Dec 2011.

25. Baez JC. "This Week's Find,"
https://johncarlosbaez.wordpress.com/2011/02/17/this-weeks-finds-week-309/,
last viewed Dec 2011.

26. Step-wise Logistic Regression and Predicted values,
http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.
htm#statug_logistic_sect052.htm, last viewed Dec 2011.

27. Wuensch KL. "Logistic Regression with SAS,"
core.ecu.edu/psyc/wuenschk/mv/multreg/logistic-sas.doc, last viewed Dec, 2011.

28. Umbaugh SE. *Digital Image Processing and Analysis: Human and Computer
Vision Applications with CVIP tools.* CRC press, Second Edition, 2011.

29. Goldin R. "Odds Ratio,"
http://stats.org/stories/2008/odds_ratios_april4_2008.html, last viewed Dec 2011.

30. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. "The diagnostic odds
ratio: a single indicator of test performance," *Journal of Clinical Epidemiology*,
56 (2003), 1129-1135.

VITA

Viswanaath Subramanian was born in Coimbatore, India. He received his primary and secondary education from Chandrakant Patkar Vidyalaya, Dombivli, Maharashtra, India. He completed his secondary education in May 2003 and then joined the Birla college of Arts and Science, Kalyan in the science stream. He graduated with a Bachelor of Engineering degree in Electronics Engineering from Fr. Conceicao Rodrigues College of Engineering, Bandra, Mumbai in May 2009.

He enrolled as a graduate student in Missouri University of Science and Technology, Rolla, MO in August 2009, majoring in Electrical Engineering. He was a Graduate Research Assistant in the DERMVIS group of cancer research from January 2010 to December 2010 and received his MS degree in Electrical Engineering in May 2012. His area of interest is in wireless communication, networks, signal processing and image processing.