

---

Masters Theses

Student Theses and Dissertations

---

Spring 2013

## Automatic detection of polypoid skin lesions: The squash sign

Saurabh G. Karnik

Follow this and additional works at: [https://scholarsmine.mst.edu/masters\\_theses](https://scholarsmine.mst.edu/masters_theses)



Part of the [Bioimaging and Biomedical Optics Commons](#), and the [Computer Engineering Commons](#)

Department:

---

### Recommended Citation

Karnik, Saurabh G., "Automatic detection of polypoid skin lesions: The squash sign" (2013). *Masters Theses*. 8047.

[https://scholarsmine.mst.edu/masters\\_theses/8047](https://scholarsmine.mst.edu/masters_theses/8047)

This thesis is brought to you by Scholars' Mine, a service of the Missouri S&T Library and Learning Resources. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact [scholarsmine@mst.edu](mailto:scholarsmine@mst.edu).

AUTOMATIC DETECTION OF POLYPOID SKIN LESIONS: THE SQUASH SIGN

by

SAURABH G. KARNIK

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 COMPUTER ENGINEERING

2013

Approved by

Randy H. Moss, Advisor  
Ronald J. Stanley  
Bijaya Shrestha

© 2013

Saurabh G. Karnik

All Rights Reserve

**PUBLICATION THESIS OPTION**

This thesis consists of the following one article formatted in the accordance with the university specifications:

Pages 3-37 To be submitted to *Transactions in Medical Imaging*, IEEE 2013.

## ABSTRACT

Earlier detection of malignant melanoma can save lives. Automatic in-vivo methods of melanoma detection including smart-phone applications are now available, but specificity is low, i.e. too many benign lesions are detected as possible melanomas. Polypoids, elevated and rounded lesions are one class of benign lesions. These raised lesions with piecewise circular borders, predominantly intradermal nevi and cutaneous polyps (IDNs), are always benign. Variegated coloring present in polypoid lesions can lead physicians to perform an unneeded biopsy. Contact dermatoscope plates with gel squash these lesions, leaving a “squash sign” marker of benignancy which can be automatically detected. We present a method to automatically recognize the squash sign in raised benign skin lesions. This method is a two-stage process. First an algorithm is implemented to narrow down upon the region of interest which is the area where the squash sign may possibly be present in the lesion. In the second stage, image features are extracted from the lesion as a whole and the region of interest described above, to mathematically model these benign lesions. The data set for training and testing set consisted of a variety of images containing intradermal nevi and cutaneous polyps, malignant melanomas and other benign lesions. All the images had a spatial resolution of 1024 by 768. It was observed that the mathematical model was able to recognize over 64% of the intradermal nevi and cutaneous polyps on both the training and testing sets. This method has the potential to reduce the number of false diagnoses in automatic melanoma detection devices.

## ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my advisor Dr. Randy Moss for support of my M.S. study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I value his enormous patience in guiding me through the tough times. I would sincerely like to thank Dr. William V. Stoecker, who has supported me throughout this research and always made himself available to guide me through the research. He has helped me integrate the medical and image processing concepts which contributed greatly to success of my research. I also wish to thank Dr. R. Joe Stanley and Dr. Bijaya Shrestha for their thoughtful inputs throughout the way. Also, I would like to express my gratitude to Jason Hagerty and Nabin Mishra who guided me in the course of my research by giving me their valuable inputs.

I would also like to thank my family for the overwhelming support they have provided me through my life that motivated me to make me what I am today.

## TABLE OF CONTENTS

	Page
PUBLICATION THESIS OPTION .....	iii
ABSTRACT.....	iv
ACKNOWLEDGMENTS .....	v
LIST OF ILLUSTRATIONS .....	viii
LIST OF TABLES .....	ix
 SECTION	
1. INTRODUCTION .....	1
 PAPER	
I. AUTOMATIC DETECTION OF POLYPOID SKIN LESIONS: THE “SQUASH SIGN” .....	3
ABSTRACT.....	4
1. INTRODUCTION.....	5
2. PRE-PROCESSING OF IMAGES.....	10
2.1 DATA SETS AND IMAGES USED .....	10
2.2 IMAGE ANALYSIS ALGORITHM .....	11
2.3 CONTRAST ENHANCEMENT ALGORITHM.....	13
2.4 I - Q PLANE.....	13
3. PROCEDURE .....	15
3.1 THRESHOLDING AND AREA FILTERING .....	15
3.2 FFT BANDPASS FILTERING.....	16
3.3 GENERATING THE OBJECT MASK .....	18

3.4	RADIAL FILTERING .....	19
3.5	HISTOGRAM FILTERING.....	21
3.6	FEATURE EXTRACTION.....	23
3.7	SAS LOGISTIC REGRESSION FEATURES.....	25
4.	RESULTS .....	31
5.	CONCLUSION .....	36
6.	REFERENCES .....	37
SECTION		
VITA	.....	39

## LIST OF ILLUSTRATIONS

	Page
Figure 1.1 The Sqash Sign.....	6
Figure 1.2 Flowchart for Algorithm to Converge ROI.....	9
Figure 2.1 Lesion Image.....	10
Figure 2.2 Dilated Lesion Border.....	12
Figure 2.3 I-Q Result After I-Q Subtracted Contrast-Enhanced Image Operation on Figure 2.1.....	14
Figure 3.1 Image after Thresholding Figure 2.3 at 0.7 .....	16
Figure 3.2 Image after Applying the Area Filter to Figure 2.3.....	16
Figure 3.3 Image after Applying the Bandpass Filter to Figure 2.3.....	17
Figure 3.4 Result of Thresholding Figure 3.3 with Lesion Mask Applied.....	18
Figure 3.5 Result of Figure 3.4 with Holes Filled.....	19
Figure 3.6 Image after Object Mask is Applied to Figure 3.2.....	19
Figure 3.7 Image after Radial Filtering.....	20
Figure 3.8 Final Mask Obtained for the ROI Shown in Black after Histogram Filtering.....	23
Figure 3.9 Sobel Gradients.....	25
Figure 3.10 B-Spline Control Points Obtained.....	25
Figure 4.1 ROC Curve.....	34
Figure 4.2 Lesion Image with Detected Border.....	35

**LIST OF TABLES**

	Page
Table 2.1 Plane Statistics.....	13
Table 4.1 Logistic Regression Model.....	31

## 1. INTRODUCTION

In this research, the statistical modeling of dermoscopic structures called Intra-dermal nevi and cutaneous polyps together termed as IDN vs. malignant melanoma and other benign lesions is analyzed to develop models based on selected features via logistic regression to differentiate IDNs from other benign lesions. The intra-dermal nevi are raised lesions with piecewise circular borders giving a three-dimensional appearance due to change in focus. This raised lesion is a sign of the lesion being benign. Due to its variegated coloring often a false biopsy is performed thus resulting in trauma to the patient and wastage of money.

This research presents a paper "Automatic Detection of Polypoid Skin Lesions: The Squash Sign" to detect intra-dermal nevi and polypoid lesions in contact dermoscopy images without special lighting equipment. This paper proposes a two stage process. In the first stage an algorithm is proposed to converge on the region of interest (ROI) where the raised sharp lesions may possibly be present. This process is done so that we can extract features in this ROI. In the second stage the statistical features from the contrast-enhanced result of the contrast-enhanced Q plane subtracted from the contrast-enhanced I plane, the bandpass filtered result of the contrast enhanced I-Q plane and features from the lesion as whole are extracted. These features will further be used to build a model using a stepwise logistic regression procedure. The stepwise logistic regression procedure will select the most important features using the  $P > \text{ChiSq}$  statistics by comparing it with the user-defined  $s_{\text{entry}}$  and  $s_{\text{stay}}$  parameters.

The above mentioned procedure is applied to all the images in the training and testing sets. Images used for the training set consisted of 38 IDNs, 120 melanomas and 25 other benign images. Images on the testing set consisted of 47 IDNs, 75 melanomas and 604 other benign images. All the images had a spatial resolution of 1024 by 768. The mathematical model was able to recognize over 64% of the IDNs on both the sets with 100% specificity on the training set and 99% on the testing set. The area under the receiver operating characteristic curve for the model was calculated to be 98%.

**PAPER****I. Automatic Detection Of Polypoid Skin Lesions: The “Squash Sign”**

**Keywords:** Intradermal Nevi, Polypoid, Melanoma, Cutaneous Polyps

**Authors:** Saurabh Karnik, Randy H.Moss, William V. Stoecker, Nabin Mishra, Jason Hagerty

**Author Affiliations:**

Stoecker & Associates, Rolla, MO (Dr Stoecker); Missouri University of Science and Technology (Mr. Saurabh Karnik, Dr. Randy Moss, Mr. Nabin Mishra, Mr. Jason Hagerty)

**All correspondence should be sent to:**

William V. Stoecker, MS, MD

Stoecker & Associates

1702 East 10th Street

Rolla, MO 65401-4600

Tel: 573-364-0122

Fax: 573-364-0129

Email: [wvs@mst.edu](mailto:wvs@mst.edu)

**Funding/Support:** Stoecker & Associates, Rolla, MO and Number NIH SBIR R44 CA-101639-02A2 of the National Institutes of Health (NIH).

## ABSTRACT

Earlier detection of malignant melanoma can save lives. Automatic in-vivo methods of melanoma detection including smart-phone applications are now available, but specificity is low, i.e. too many benign lesions are detected as possible melanomas. Polypoids, elevated and rounded lesions are one class of benign lesions. These raised lesions with piecewise circular borders, predominantly intradermal nevi and cutaneous polyps (IDNs), are always benign. Variegated coloring present in polypoid lesions can lead physicians to perform an unneeded biopsy. Contact dermatoscope plates with gel squash these lesions, leaving a “squash sign” marker of benignancy which can be automatically detected. We present a method to automatically recognize the squash sign in raised benign skin lesions. This method is a two-stage process. First an algorithm is implemented to narrow down upon the region of interest which is the area where the squash sign may possibly be present in the lesion. In the second stage, image features are extracted from the lesion as a whole and the region of interest described above, to mathematically model these benign lesions. The data set for training and testing set consisted of a variety of images containing intradermal nevi and cutaneous polyps, malignant melanomas and other benign lesions. All the images had a spatial resolution of 1024 by 768. It was observed that the mathematical model was able to recognize over 64% of the intradermal nevi and cutaneous polyps on both the training and testing sets. This method has the potential to reduce the number of false diagnoses in automatic devices.

## 1. INTRODUCTION

Raised, smooth, benign skin lesions including intradermal nevi (IDN), neurofibromas, and cutaneous polyps, even those that vary in color, are always benign, which makes them important to reduce the number of false biopsies. The intradermal nevi are raised lesions with piecewise circular borders giving a three-dimensional appearance due to change in focus. This research presented here is our response to the clinical need to detect intradermal nevi and polypoid lesions in contact dermoscopy images without special lighting equipment. One in-vivo method has been proposed—the wobble sign [1]. The wobble sign is positive if the dermatoscope glass plate pressing against the lesion can be rocked or wobbled back and forth. Sometimes the polyp is too small to wobble. A general practitioner sent an image to a dermatology chat group, generating significant concern about the variegated coloring. Experienced specialists, noting the smooth border, reassured the general practitioner that the lesion was benign.

Before acquiring the image of the lesion, ultrasound gel is applied to the skin after which a contact dermatoscope plate is placed on the skin. Due to the elevation of the lesion, the lesion is pressed against the dermatoscopic glass plate which forms a sharp gel-like border around the raised lesion. In this paper we call the sharp gel-like border generated by the ultrasound gel due to a dermatoscopic plate being pressed against it the "squash sign". Our observation is that a smooth border observed at the squashed lesion edge, the "squash sign", is a strong indicator of a benign lesion. Figure 1.1 shows the

squash sign for such lesions. This research aims to provide an alternate to other methods of detecting elevation, shape from shading, and specialized lighting systems [2].

The IDNs are approximately circular in shape and have the sharp raised border which is close to the lesion border. We make three assumptions while finding the IDNs which are as follows: 1) the sharp borders are close to the lesion mask borders, 2) intradermal nevi and cutaneous polyps are approximately piecewise circular and the distance of the sharp edge from the center of lesion does not vary to a great degree, and 3) the image is taken such that the lesion is approximately at the center of the image.

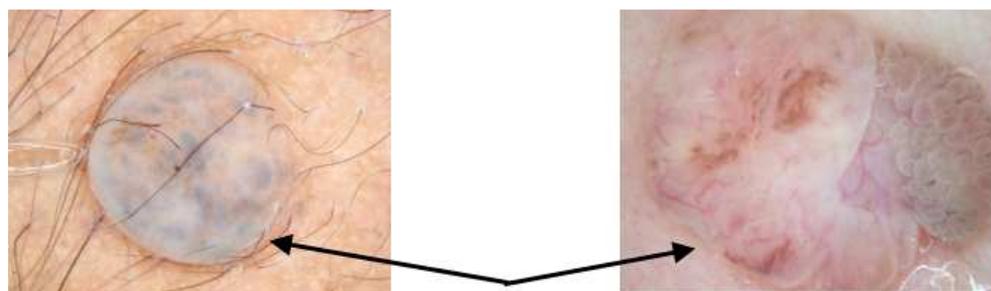


Figure 1.1 The Squash Sign

We present a two-stage method to detect the IDNs. In the first stage we implement an algorithm using image processing techniques to narrow down upon the region of interest which is the area where the squash sign may possibly be present in the lesion. In the second stage we extract features from the region of interest and the lesion itself as a whole and build a mathematical model to identify these benign lesions.

It was observed that a combination of planes in the YIQ color scheme, which is used for NTSC analog color television transmission, was ideal to extract this region of

interest (ROI), i.e. the area around the squash sign. The given algorithm was applied to the resultant plane obtained when the contrast-enhanced Q plane was subtracted from the contrast-enhanced I plane of the YIQ color scheme. This result was further contrast-enhanced again.

Step 1: Threshold and area filtering: The result of the contrast-enhanced Q plane subtracted from the contrast-enhanced I plane is thresholded. An area filter is applied after it is thresholded to remove the objects that have area less than 15% of the largest object in the image.

Step2: A bandpass filtering: A bandpass filter is applied to the (I-Q) contrast-enhanced image. This will filter out the smooth area inside the lesion leaving the ROI, i.e. the area around the squash sign.

Step 3: Generating the object mask: The result of image the obtained after step 2 is thresholded. The holes in the thresholded image are filled. This binary image depicts the area in the lesion where the squash sign is not present. When the object mask is inverted and ANDed with the image in step 1, a significant amount of noise is removed from the image, helping us converge on the ROI.

Step 4: Radial Filtering: The image obtained after applying the object mask in step 3 is divided into 66 sectors with the center of image as the central point of reference. The radius for an object is calculated as the distance of the centroid of the object to the center of image. One object per sector is selected. The object selected in each sector is the one which is furthest away from the center of image.

Step 5: Histogram Filtering: Radii of the objects, i.e. distance of the centroid of the objects from the center of image in each sector, are plotted in the form of a histogram. Histogram filtering is done on the image obtained after radial filtering at step 4, such that there is minimum fluctuation in the distance between the objects with respect to center of the image. This will further help to converge on the ROI.

In the second stage, features were extracted from the lesion as whole on various color planes. They were analyzed by the software package SAS using a stepwise logistic regression model. The probability given by the stepwise logistic regression model was calculated on the training and testing sets. A threshold was chosen based on the training set and the classifier was defined. The flowchart in Figure 1.2 summarizes the steps included for finding the ROI for the images.

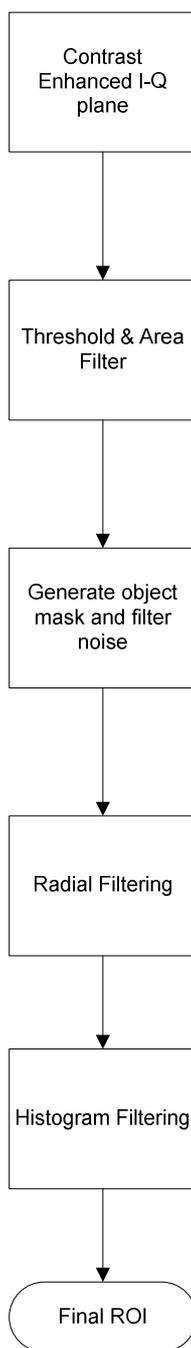


Figure 1.2 Flowchart for Algorithm to Converge ROI

## 2. PRE-PROCESSING OF IMAGES

### 2.1 DATA SETS AND IMAGES USED

The image set used for this work consisted of 120 melanoma images and 38 elevated intradermal nevi images along with 25 images that were neither IDNs nor melanomas. All images were reduced to 1024 by 768 resolution and retained full 24-bit color. Figure 2.1 is the original dermoscopic image of an IDN lesion. For processing, instead of the entire image, a mask is applied and only the lesion area is obtained. The lesion masks were drawn manually using software named 'Winshow' [3] which allowed the manual creation of border masks. All the borders were checked by an expert dermatologist. It should be noted that the classification done was binary, in which the positives were considered as the IDN lesions and the negatives were other benign lesion and malignant melanomas.

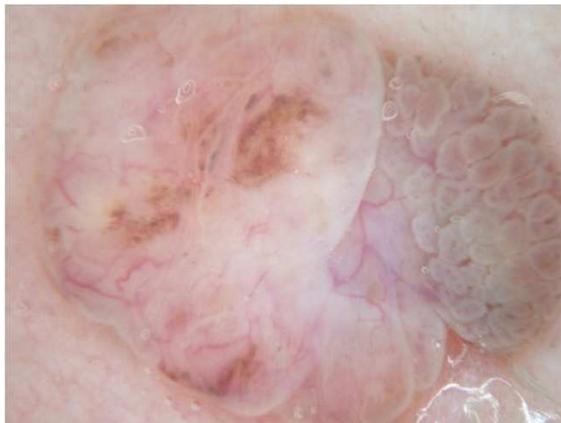


Figure 2.1 Lesion Image

## 2.2 IMAGE ANALYSIS ALGORITHM

One of the major features that needed to be examined was the elevation due to the lesion coming out of the skin. Edge detectors along with Gaussian filters of different sizes and sigma values were used on the RGB planes to obtain the borders of the elevated lesions or the squash sign. However, the edge detectors were unable to obtain the squash sign or the sharp elevated edges. Moreover the edge parameters like edge length and longest continuous edge due to the squash sign would not have been enough to classify the IDNs. This made it necessary to look into different color plane schemes and combinations that would help us to converge upon the ROI, i.e. the area in the lesion possibly containing the squash sign. It was observed that the squash sign was close to the lesion border. So the border of the lesion mask was dilated by a disk-shaped structuring element with a radius of 12 pixels. In the dilated lesion border the objective was to select a plane in which the squash sign is highlighted. The dilated lesion border is shown in Figure 2.2. The difference of the standard deviation of the area of the dilated lesion border and the standard deviation inside the dilated border of the lesion mask was calculated for various image planes and their combinations. The area in the image which could possibly contain the squash sign, i.e. the ROI, would be highlighted in an image plane where the difference of these standard deviations obtained would be highest. Table 2.1 shows standard deviation parameters on various image planes and their combinations, calculated for the image shown in Figure 2.1. It is observed that the I-Q color plane (the contrast-enhanced Q plane subtracted from the contrast-enhanced I plane) of the YIQ color scheme is the best scheme to extract this elevation as it has the maximum difference of standard deviation between the dilated lesion border (where the squash sign is

expected to appear) and the interior of the lesion. Thus the contrast-enhanced I-Q plane highlighted the squash sign. This result was found to be consistent with other IDN images. It should be noted that the standard deviations of the subtraction of two planes is not the same as the standard deviation when they are subtracted in other direction, because the contrast-enhanced function was also applied the result after they were subtracted. The Y plane in the YIQ color scheme is the Luminance plane, the I plane stands for in-phase and the Q plane stands for the quadrature, which refers to the in-phase and the quadrature phase components in quadrature modulation. The YIQ color plane can be defined in terms of RGB plane[4] as follows.

$$Y = (0.299)(R) + (0.587)(G) + (0.114)(B)$$

$$I = (0.596)(R) - (0.274)(G) - (0.322)(B)$$

$$Q = (0.211)(R) - (0.523)(G) + (0.312)(B)$$

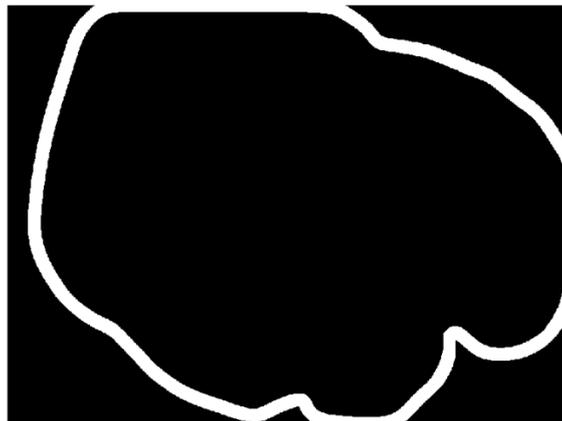


Figure 2.2 Dilated Lesion Border

Table 2.1 Plane Statistics

Plane	Std Dev Inside	Std Dev edge	Std Dev difference
Y	0.0691	0.0692	0.0001
I	0.0597	0.1748	0.1151
Q	0.0535	0.2772	0.2237
R	0.0642	0.1658	0.1016
G	0.0625	0.163	0.1005
B	0.0612	0.1516	0.0904
Y-I	0.0794	0.3665	0.2871
I-Y	0.1046	0.3645	0.2599
I-Q	0.0709	0.3642	0.2933
Q-I	0.1146	0.3443	0.2297
Y-Q	0.0715	0.3567	0.2852
Q-Y	0.0904	0.3493	0.2589
R-G	0.0724	0.235	0.1626
G-R	0.1255	0.1391	0.0136
R-B	0.0693	0.2269	0.1576
B-R	0.135	0.1044	-0.0306
G-B	0.0409	0.2917	0.2508
B-G	0.0848	0.0598	-0.025

### 2.3 CONTRAST ENHANCEMENT ALGORITHM

For preprocessing we enhance the contrast of the I and Q plane linearly, such that one percent of the pixels in the image are saturated at each of the 0 and 1 values, using the matlab function "imadjust". This contrast enhancement operation was performed upon the entire image.

### 2.4 I - Q PLANE

After the contrast enhancing is done, the contrast-enhanced Q plane is subtracted from the contrast-enhanced I plane. The result is also linearly contrast-enhanced again with one percent of the pixels saturated at 0 and 1 values each. Figure 2.3 shows the

result when the operation is performed on Figure 2.1. It was observed that that wherever there is a sudden change in elevation, around the elevated edges on the contrast-enhanced I-Q plane, there is a high contrast in the region containing those elevated edges, where the intensity changes from a lower value to a higher value. In Figure 2.3 we observe the intensity changes mentioned above around the sharp edges. We further try to converge on the areas around the sharp elevated edges or the squash sign and use them as a feature to recognize the elevated lesion images. The bright borders signify the sharp edges surrounded by a darker background.

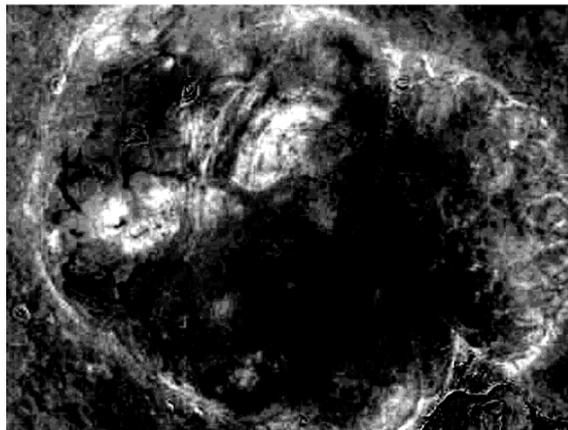


Figure 2.3 Result after I-Q Subtracted Contrast-Enhanced Image Operation on Figure 2.1

### **3. PROCEDURE**

#### **3.1 THRESHOLDING AND AREA FILTERING**

The contrast-enhanced I-Q image plane was used for converging upon the ROI as the results for the image analysis algorithm mentioned in section 2.2 was found to be consistent with all the IDN images. The following procedure described was performed upon all the images which were to be examined for potential IDNs, for converging upon the possible region of interest for the feature extraction. An example is illustrated with images, that were obtained after performing the following described procedure on the lesion image shown in Figure 2.1 .

To extract the sharp edges, which were highlighted by the bright area around the squash sign, thresholding was used on the contrast-enhanced I-Q image plane and 0.7 was found empirically to be the optimum threshold (where black is represented by zero and white as one). The result of this thresholding operation on Figure 2.3 is illustrated in Figure 3.1. After the thresholding was done many insignificant objects and noise were present. This noise was eliminated by applying an area filter such that all white objects that had an area less than 40 pixels inside the lesion, were eliminated. The number 40 was obtained empirically. The result of the area filtering operation on Figure 3.1 is illustrated by Figure 3.2 .

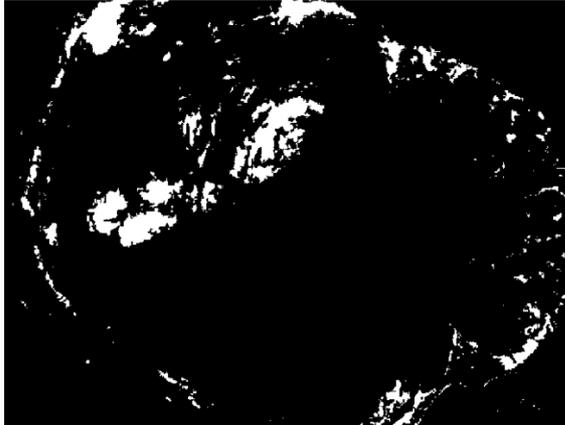


Figure 3.1 Image after Thresholding Figure 2.3 at 0.7

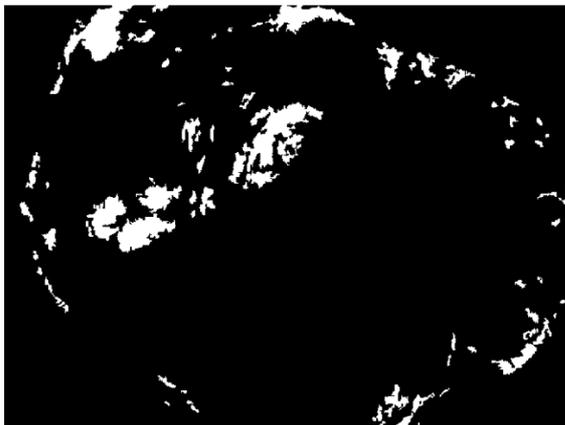


Figure 3.2 Image after Applying the Area Filter to Figure 3.1

### **3.2 FFT BANDPASS FILTERING**

After thresholding, possible areas of sharp edges or elevations were obtained. However, the region of interest which is the area around the elevated sharp edges or the squash sign still needs to be more precisely defined. It was observed that high-frequency granular noise and smooth low-frequency areas were present inside the lesion where there

was an absence of edges, which is illustrated by Figure 2.3. It was thus hoped that a bandpass filtering operation followed by a threshold operation could be used to create an object mask that would eliminate the unwanted area inside the lesion. Thus a bandpass filter was implemented. An ideal bandpass filter was defined in the frequency plane of value one where the distance from the origin was between 100 and 200 and of value zero otherwise. This bandpass filter was empirically determined and accepted as it preserved the area around the ROI after the filtering operation. After the application of this filter, the original image was subtracted from the bandpass filtered image. After the subtraction the absolute value of the most negative pixel was added to the resulting image making the smallest value of a pixel in the image zero. The resultant image was then normalized by dividing the image by the maximum value present. The result of this operation performed upon Figure 2.3 is shown in Figure 3.3 .

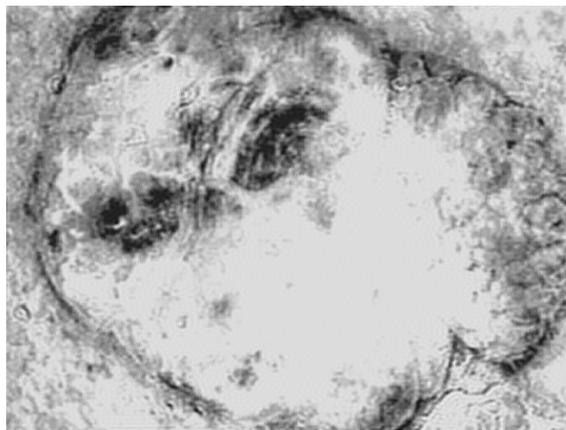


Figure 3.3 Image after Applying Bandpass Filter to Figure 2.3

### 3.3 GENERATING THE OBJECT MASK

The image obtained after the bandpass filtering operations was binary thresholded at a level of 0.9 which was determined empirically. The thresholded images often had holes present in them that had to be filled. The MATLAB function "imfill" was used to fill these potential holes. For the lesion image shown in Figure 2.1, Figure 3.4 shows the result after thresholding Figure 3.3. Figure 3.5 shows the result after the holes that occurred in Figure 3.4 were filled. The mask after holes were filled (Figure3.5) signifies the region where the region of interest is NOT present. The complement of this mask was ANDed with the lesion mask that was manually drawn. The resulting image was ANDed with the result obtained after area filtering in Section 3.1, illustrated by Figure 3.2, to reduce the noise present inside the lesion. This result after reducing the noise inside the lesion, thereby converging upon the region of interest, i.e. the area around the squash sign, is illustrated by Figure 3.6 .



Figure 3.4 Result of Thresholding Figure 3.3 with Lesion Mask Applied



Figure 3.5 Result of Figure 3.4 with Holes Filled



Figure 3.6 Image after Object Mask Applied to Figure 3.2

### 3.4 RADIAL FILTERING

After applying the object mask to erase the noise inside the lesion and the lesion mask, the majority of the noise is expected to be filtered out. It was observed that the sharp edges were close to the lesion mask and hence further away from the center. The image obtained after applying the object mask was divided into radial sectors, where the

central point of reference was the center of the image. Centroids of objects in a particular sector were calculated as well as the corresponding distance from each centroid to the center of the image. The object whose centroid had the maximum distance from the center in a given sector was retained and the others were erased. Thus, we obtained only one particular object in a given sector. It was observed from a number of IDN images that there were objects closely spaced adjacent to each other. Thus the sector for the radial filtering had to be large enough to reduce the computational complexity and small enough so that we do not lose any ROI in the image, as only one object per sector is chosen. Observing various images a five-and-a-half degree resolution was found to be appropriate. Hence the all the images, were divided into 66 sectors for the purpose of radial filtering. The result after radial filtering for lesion image in Figure 2.1 is shown by Figure 3.7 .



Figure 3.7 Image after Radial Filtering

### 3.5 HISTOGRAM FILTERING

Most IDN images are obtained such that the lesion was approximately centered in the image. Most lesions are approximately oval in shape. Thus there should not be much fluctuation in the radius where the radius is calculated as the distance between the centroid of selected object in each sector and the center of the image. The following algorithm eliminated the objects that constituted the ROI such that, there is minimum fluctuation of the distance between centroid of the objects of the ROI from the center of image. This algorithm was applied to all the images after the radial filtering was complete.

The following algorithm is applied to converge on the ROI i.e the area around the squash sign or the sharp elevated border:

Step 1: Divide the radii range from an image into ten equal subsets (divisions), for histogramming. The divisions were made such that each division had the range of  $(\text{Maximum radius obtained in an image})/10$ . Exclude the values from first four ranges of radii which will be from 0 to  $0.4 * (\text{maximum radius in an image})$ . We exclude the given range due to the observation that the ROI is closer to the lesion borders and hence the radii of interest will not range over the first four minimum value subsets as they will be well inside the lesion. After the described operations the range of radii present will be greater than  $0.4 * (\text{maximum radius present in an image})$  distributed over six subsets. Select the subset that has the maximum frequency of occurrence for the radii.

Step 2: If the maximum frequency of occurrence subset is subset 10 (the outermost subset), check subset 9 to see if the frequency of occurrence of subset 9 is 60% or more of that of subset 10. If not, include the radii in subset 10 and erase others as

outliers. If subset 9 is at least 60% of that of subset 10, check if subset 8 is 60% or more of that of subset 10. If both subset 8 and subset 9 are at least 60% of that of subset 10, include the radii in subsets 8, 9 and 10, and erase all others as outliers. If subset 8 is not 60% of that of subset 10 include radii in subsets 9 and 10 and erase all others as outliers. If the maximum (frequency of occurrence) subset is not subset 10 proceed to step 3.

Step 3: Assuming the maximum frequency of occurrence subset is subset  $N$ , where  $N \neq 10$  observe the frequency of occurrence of subset  $N-1$ . Check to see if its frequency of occurrence is at least 60% of the  $N$ th subset. If not proceed to step 5. If so, observe subset  $N+1$  and check if it is at least 60% of the  $N$ th subset. If not proceed to step 4. If both the above conditions are satisfied keep the radii that lie in the subsets  $N$ ,  $N-1$  and  $N+1$ . Erase all objects whose radii fall in the subsets other than  $N$ ,  $N-1$  and  $N+1$ .

Step 4: Observe the frequency of occurrence of the subset  $N-2$ . If it is at least 60% of the  $N$ th subset include all the objects that have radii in subsets  $N$ ,  $N-1$  and  $N-2$ . Erase all objects whose radii fall in the subsets other than  $N$ ,  $N-1$  and  $N-2$  as outliers. If the above condition is not satisfied include radii in subsets  $N$  and  $N-1$  and erase others as outliers.

Step 5: Observe the frequency of occurrence of the subset  $N+1$ . If it is at least 60% of the  $N$ th subset proceed to step 6. If the above condition is not satisfied include radii in subset  $N$  and erase others as outliers.

Step 6: If  $N \neq 9$  observe the frequency of occurrence of the subset  $N+2$ . If it is at least 60% of the  $N$ th subset include all the objects that have radii in subsets  $N$ ,  $N+1$  and  $N+2$ . Erase all objects whose radii fall in the subsets other than  $N$ ,  $N+1$  and  $N+2$  as

outliers. If the above condition is not satisfied, include radii in subsets  $N$  and  $N+1$  and erase others as outliers. If  $N=9$  erase all objects whose radii fall in the subsets other than  $N$ ,  $N+1$  as outliers. Figure 3.8 shows the resulting ROI obtained after histogram filtering was performed upon the output of radial filtering (Figure 3.7) for the lesion shown in Figure 2.1. Thus the algorithm to converge on the ROI which was the area around the squash sign or the elevated lesion border was implemented.

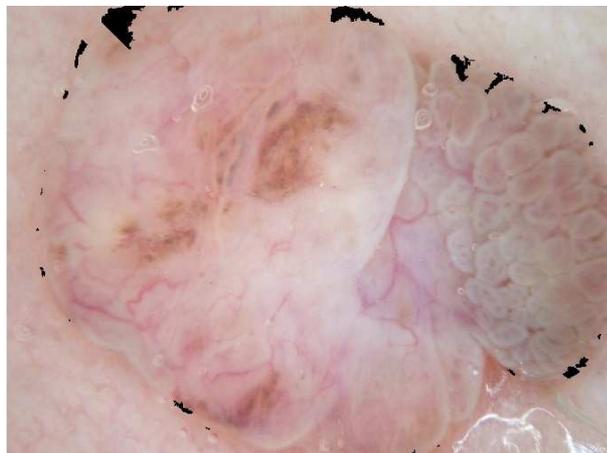


Figure 3.8 Final Mask Obtained for the ROI Shown in Black after Histogram Filtering

### **3.6 FEATURE EXTRACTION**

After the histogram filtering the ROI was obtained for the particular lesion and we proceed to the second stage of the method, i.e. to extract features. The following procedure was followed to extract features to identify potential IDNs.

This region of interest obtained after histogram filtering was superimposed upon the contrast-enhanced I-Q plane (Figure 2.3) and the bandpass-filtered image (Figure 3.3) in the case of the lesion image in Figure 2.1. The mean and standard deviation of the images within the ROI obtained after histogram filtering were calculated after superimposing the ROI on the contrast-enhanced I-Q plane and on the bandpass-filtered image. The ROI was further broadened by dilating it by a disk-shaped structuring element of radius 4 pixels. This was done to record the change in mean and standard deviation before and after dilation for the contrast-enhanced I-Q plane and the bandpass-filtered image.

**Sobel Filtering:** The region of interest obtained after histogram filtering was superimposed on the contrast-enhanced I-Q plane. At this point the precise location of the sharp edges needed to be obtained from the contrast-enhanced I-Q plane within the ROI. The area around the sharp edges was observed to be of high gradient. Sobel compass gradients as shown in Figure 3.9 were applied within the ROI obtained after histogram filtering superimposed upon the contrast-enhanced I-Q plane to find these sharp edges. The resulting four images obtained were averaged. An averaging filter of dimension 3 by 3 was applied to this result to smooth out the image. The image was divided into 66 sectors, each sector of 5.5 degrees, and the one point of highest intensity from each sector was selected from the 66 sectors. The number of points obtained on the whole image having non-zero intensity was one of the features. These points acted as control points for a second-order B-spline curve to be drawn if the given image was diagnosed as an IDN. The points obtained for the IDN lesion in Figure 2.1 is shown by a part of the lesion zoomed in, in Figure 3.10 .

-2	-1	0
-1	0	1
0	1	2

0	-1	-2
1	0	-1
2	1	0

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
-1	0	1

Figure 3.9 Sobel Gradients



Figure 3.10 B-Spline Control Points Obtained

### 3.7 SAS LOGISTIC REGRESSION FEATURES

The following 22 features were extracted from the lesions as a whole and from within the ROI mask superimposed upon the contrast-enhanced I-Q image and upon bandpass filtered image and the lesion as whole to build a mathematical model.

Feature 1: Red Deviation- The standard deviation in the red plane of the image constrained to the lesion area. This feature was chosen to reject the melanomas as the IDNs had less standard deviation in the red plane compared to the melanomas.

Feature 2: Lesion area of the lesion mask divided by square of the perimeter of lesion mask- The IDNs were observed to be piecewise circular in shape whereas the other lesions like the melanomas were observed to be irregular. So the ratio of the lesion area to the square of the lesion perimeter was added to the features list to be analyzed by stepwise logistic regression.

Feature 3: Percentage of lesion area occupied by object mask- It was expected that the IDN lesions were darker in the center of the lesion area and were bright at the borders when the contrast-enhanced Q plane was subtracted from the contrast-enhanced I plane. So the object mask obtained in Section 3.3 was expected to have greater area in the case of IDNs than other lesions. Observing this, the percentage of the lesion mask occupied by the object mask was added as a feature.

It was observed that IDNs contained less noise inside the lesion on the contrast-enhanced I-Q plane as compared to the other types of images. Also the IDN lesions were piecewise circular as compared to other lesions that were fairly irregular. Hence the number of objects erased by the object mask, the radial filtering procedure and the histogram filtering in case of non-IDNs would be more as compared to the IDNs on the contrast-enhanced I-Q plane. Thus the number of objects erased due to the object mask, radial filtering and the histogram filtering and the total number of objects erased due to these 3 types of filtering were considered. They are represented by features 4 through 7

Feature 4: Number of objects erased due to object mask- This feature was represented by the number of objects that were erased after applying the object mask on the thresholded and area filtered contrast-enhanced I-Q plane of the lesion, obtained in Section 3.3.

Feature 5: Number of objects erased due to histogram filtering- This feature was represented by the number of objects that were erased after applying the histogram filtering procedure described in Section 3.5.

Feature 6: Number of objects erased due to all three methods above- This feature is indicates the total number of objects erased, given by the number of objects erased due to the applying the object mask, histogram and radial filtering procedures described from Sections 3.3 to 3.5.

Feature 7: Standard deviation of radius of ROI objects in Figure 3.8- This feature was included as the radii of ROI objects, calculated as the distance from centroid of the object to center of the image, obtained after histogram filtering for IDNs would not be expected to fluctuate much and hence the standard deviation of the radius of ROI would be less in the case of IDNs. Thus the standard deviation of all the radii of ROI objects after histogram filtering in Section 3.5 was calculated as Feature 7.

Feature 8: Difference of radius with respect to lesion- The radius of the lesion was calculated by dividing the image into 66 sectors as described in Section 3.4 for the radial filtering. The radius of the lesion in each sector was calculated as the maximum distance of a pixel present in the lesion mask in a particular sector from the center of the image. The mean of these distances calculated is termed the radius of the lesion. The squash sign

was found to be close to the lesion border in the majority of the cases, thus the difference between the mean of the radius for ROI objects, calculated as the distance from the centroid of the object to the center of the image, obtained after histogram filtering, and the mean of the radius of the lesion would be expected to be less in the case of the IDNs.

Feature 9: Mean of the image of the lesion in the contrast-enhanced I-Q plane constrained to the lesion mask- The mean of melanomas was found to be fairly high while the mean for IDNs was found to be low.

Looking at the IDN images in the I-Q plane we observed that there are high intensity pixels surrounding the border of the elevated lesion whereas inside the lesion the pixel intensity is low. Before dilation of the ROI, the high intensity pixels in the case of IDN are not included. They are only included after the ROI is dilated. It was observed in the majority of the IDN images, that there will be a high change in mean and standard deviation in the case of the IDN images in the I-Q plane after dilation. Figure 2.3 illustrates the above observation for the IDN lesion shown in Figure 2.1. Observing Figure 3.3 we can observe that, like the contrast-enhanced I-Q plane shown in Figure 2.3, Figure 3.3 will have change in standard deviation and change in mean when the ROI is dilated. Hence features 10 to 20 were chosen based on the original ROI (after histogram filtering), the dilated ROI, the contrast-enhanced I-Q plane image (Figure 2.3) and the image after applying the bandpass filter (Figure 3.3) .

Features 10 and 11: Mean of the image on the contrast-enhanced I-Q plane, illustrated by Figure 2.3, constrained to the ROI obtained after histogram filtering before and after dilating the ROI.

Features 12 and 13: Mean of the image obtained after performing the bandpass filter operation, performed in Section 3.3 illustrated by Figure 3.3, constrained to the ROI obtained after histogram filtering before and after dilating the ROI.

Features 14 and 15: Standard deviation of the image on the contrast-enhanced I-Q plane, illustrated by Figure 2.3, constrained to the ROI obtained after histogram filtering before and after dilating the ROI.

Features 16 and 17: Standard deviation of the image obtained after performing the bandpass filter operation, performed in section 3.3 illustrated by Figure 3.3, constrained to the ROI obtained after histogram filtering before and after dilating the ROI.

Feature 18: Change in mean of the ROI on the contrast-enhanced I-Q plane before and after dilating. So this will be Feature 11 - Feature 10.

Feature 19: Change in standard deviation of the ROI on the contrast-enhanced I-Q plane before and after dilating. So this will be Feature 15 - Feature 14.

Feature 20: Change in standard deviation divided by change in mean on the contrast-enhanced I-Q plane. So this will be  $(\text{Feature 15} - \text{Feature 14}) / (\text{Feature 11} - \text{Feature 10})$ .

Feature 21: Number of points of interest obtained in an image- These points will act as control points for the B-spline curve to trace the sharp gel-like borders. Spline control points are the points of interest found after the averaging and Sobel operations described in section 3.6. Figure 3.10 shows a zoomed part of the lesion for Figure 2.1 for which points were obtained.

The above features extracted from a given image act as input data to be analyzed by stepwise logistic regression using SAS. Stepwise logistic regression is used in cases where the dependent variable is anticipated to be non-linear with respect to one or more independent variables. The stepwise logistic regression used in the given case 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. These features are run through SAS analysis software to build a stepwise logistic regression model. The SAS software runs the procedure for the stepwise logistic regression and produces a linear equation that conveys to us the important features that are used in the classification equation.

#### 4. RESULTS

A mathematical model needed to be built to form a classifier. Images in the training set for the classifier consisted of 120 melanomas, 38 IDNs and 25 other images which were benign. Thus a total of 183 images were used in the training set.

When these features were input into SAS, a logistic regression model was developed to model the IDNs with the given set of features. The given 21 features were analyzed by stepwise logistic regression and nine features were selected by the model as the most significant. The significant features are shown in Table 4.1.

Table 4.1 Logistic Regression Model

Parameter	Estimate	Std Error	Chi-Square	Pr> Chi Sq
Intercept	-20.1623	4.993	16.3062	<.0001
F1	-44.5077	12.8509	11.995	0.0005
F2	406.6	91.647	19.6832	<.0001
F4	-0.0184	0.0144	1.6317	0.2015
F5	0.1714	0.0822	4.3448	0.0371
F7	0.0238	0.0297	0.6458	0.4216
F9	-6.7969	3.3395	4.1425	0.0418
F18	-16.2352	22.3378	0.5282	0.4673
F19	80.2362	44.0322	3.3205	0.0684
F20	-3.1804	2.2089	2.073	0.1499

The Chi-Square test statistic is the squared ratio of the Estimate to the Standard Error of the respective predictor. The probability that a particular Chi-Square test statistic is as extreme as, or more so, than what has been observed under the null hypothesis is defined by Pr>ChiSq. The smaller the value of Pr>ChiSq the more important is the feature. In stepwise logistic regression, in each step a new feature is added (the one that has least value for Pr>Chi Sq), and the model and Pr>Chi Sq values are recalculated.

The stepwise regression stops when all the features to be added are greater than the  $Pr > ChiSq$  value set for slentry (a user-defined parameter for the stepwise logistic regression procedure). Also all the features added to the model must remain below the  $Pr > ChiSq$  value set as slstay (another user-defined parameter) in each step when the model is recalculated in order for those features to remain in the model. The slentry and slstay parameters were set to 0.5 for obtaining the above model. The features that did not end up in the final model all had the  $Pr > ChiSq$  values greater than 0.5.

Standard deviation in the red plane (F1), lesion area/perimeter (F2), number of objects erased due to object mask (F5), mean of image in the lesion area on the I-Q plane (F9) and the difference in the standard deviation before and after dilating the ROI on Figure 2.3 (F19) were the top five features in the model.

The Z value which further gives the probability of the given image to be IDN or not is given by the following equation. It is obtained with reference to Table 4.1 as shown.

$$Z = -20.1623 + (-44.5077)(F1) + (406.6)(F2) + (-0.0184)(F4) + (0.1714)(F5) + (0.0238)(F7) + (-6.7969)(F9) + (-16.2352)(F18) + (80.2362)(F19) + (-3.1804)(F20)$$

The negative coefficients in the Z equation indicate that the more prominent the given feature the smaller is the probability of the given image being an IDN.

The probability  $F(Z)$  was calculated as

$$F(Z) = \frac{1}{1 + e^{-Z}}$$

The receiver operating characteristic (ROC) curve, the plot of sensitivity vs. 1-specificity, is shown in Figure 4.1. The Proc Logistic c statistic = 0.98 gives an estimate for the area under the ROC curve which can also be observed in Figure 4.1. A model that is able to find many benign IDNs while misclassifying very few, if any, melanomas is more desirable than a model which identifies all IDNs at expense of misclassifying more melanomas, as it is not desirable to classify a malignant lesion like melanoma as an IDN. Thus we want the model to have a higher specificity.

Thus a specificity of 100% and a sensitivity of 68% was chosen for the training set. These were considered acceptable parameters as there were no false positives. The threshold for  $F(Z)$  corresponding to the given specificity and sensitivity was 0.78. All images with probability greater than 0.78 are classified as positive while (IDN) those lesser than 0.78 are classified as negative.

The classifier when applied on the training set produced the results as follows: Twenty-six of 38 IDNs were detected as IDNs. None of the 120 melanomas in the training set were detected as IDNs. Zero of 25 other images which were benign was detected as an IDN in the training set. Misclassifying other benign lesions as IDN does no harm since no malignant lesion is being mis-classified.

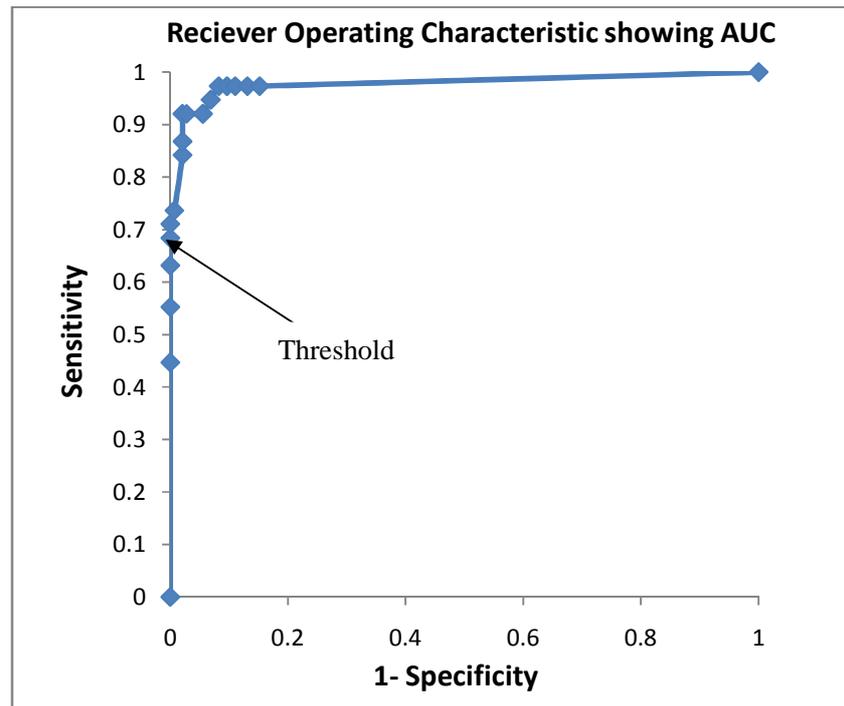


Figure 4.1 ROC Curve

The results were confirmed by applying the model on an independent test set which consisted of 75 melanomas, 47 IDNs and 604 other images which were benign. Thus a total of 726 images were used. None of the 75 melanomas was detected as IDN. Thirty of 47 (63.8%) IDNs were detected as IDNs. Seven other benign images were classified as IDNs. In this particular context classifying other benign images as IDNs does not impact the goal as the overall goal of the system is to remove benign images that were classified as malignant melanomas.

Figure 4.2 shows the portions of the squash sign in a lesion image that was detected as a true positive IDN. The detected portions of the squash sign were drawn

using a second-order B-spline curve by using the points detected as control points for the curve.

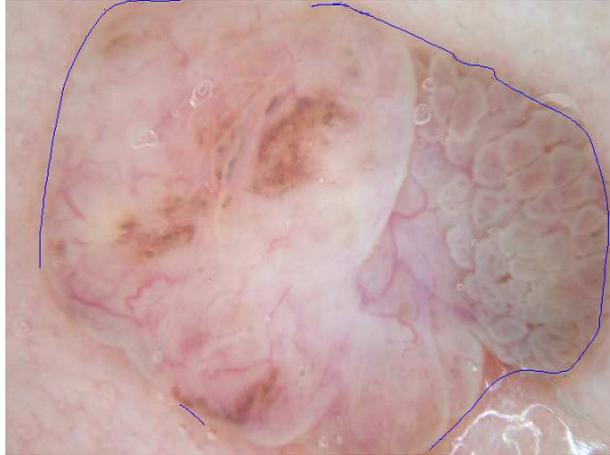


Figure 4.2 Lesion Image with Detected Border

## 5. CONCLUSION

Research was conducted to classify a lesion as intradermal nevi or cutaneous polyps based on the features discussed above. Various characteristics were studied to determine the potential images that could be IDNs. It was more critical not to classify melanomas as IDNs than vice-versa. Hence finding a smaller number of IDNs was acceptable as long as the melanomas were not misclassified. The objective as a whole was to reduce the number of benign lesions that was misclassified as melanomas. Approximately 64% of the IDNs were detected as IDNs on the test set while 68% of the IDNs were detected as IDNs on the training set. In the test set none of 75 melanomas were misclassified as IDN. Approximately 1.0% ( 7 of 604) of other benign images were classified as IDNs. The inception of the whole project was based on three images that were being classified as melanomas but that were IDNs due to their raised piecewise circular profile. These images were separately tested and the  $F(Z)$  recorded on these images was found to be the highest, greater than the images on the test and training sets, thus definitively classifying them as benign and removing them as false positives from the melanoma classification.

## 6. REFERENCES

1. Braun RP, MD, Krisher J, MD, Saurat J-H, MD."The 'Wobble Sign'" in Epiluminescence Microscopy as a Novel Clue to the Differential Diagnosis of Pigmented Skin Lesions". *Arch Dermatol.* 2000;136(7):940-942.
2. Bartels A.M, "Low Cost Computer Vision Technique for Acquisition of the 3 Dimensional Profile of an Object," M.S. thesis in Electrical Engineering, University of Missouri-Rolla, Rolla, MO, June 1991.
3. Nagane R. , "Detection Of Flash In Dermoscopy Skin Lesion Images," M.S. thesis in Electrical Engineering, University of Missouri-Rolla, Rolla, MO, 2007.
4. <http://www.mathworks.com/help/images/ref/rgb2ntsc.html>, Referred on 3/1/2013
5. 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" in Ch 5, LAG (eds), *Malignant Melanoma, National Cancer Institute*, NIH Pub. No. 06-5767. Bethesda, MD, 2006. pp 53-63.
6. Lotze MT, Dallal RM, Kirkwood JM, Flickinger JC. "*Cutaneous melanoma, Principles and Practice of Oncology*", 6e. Philadelphia: Lippincott, 2001.
7. Sedat Ozturkcan, Ali Ekber, Riza Dunder, Filiz Gulustan, Demet Etit, Huseyin Katilmis , Intradermal Nevus of the External Auditory Canal: A Case Report *Int. Adv. Otol.* 2009; 5:(3) 401-403
8. Youngs R, Hawke M, Kwok P. "Intradermal nevus of the ear canal". *J Otolaryngol.* 1988; 17:241-3.
9. Argenziano G, Longo C, Cameron A, Cavicchini S, Gourhant JY, Lallas A, et al. "Blue-black rule: a simple dermoscopic clue to recognize pigmented nodular melanoma". *Br J Dermatol.* 2011; 165(6): 1251-5.
10. Stanley RJ, Moss RH, Van Stoecker W, Aggarwal C. "A fuzzy-based histogram analysis technique for skin lesion discrimination in dermatology clinical images." *Comput Med Imaging Graph.* 2003; 27(5):387-96.
11. Cheng M-M, Zhang G-X, Mitra NJ ,Huang X, Hu S-M. "Global Contrast based Salient Region Detection " . *IEEE Computer Vision and Pattern Recognition 2011*

12. Martin DR, Fowlkes CC, and Malik J: "Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, VOL. 26, NO. 5, MAY 2000

## VITA

Saurabh Karnik was born in Indore, India on 17<sup>th</sup> of October 1988. He received his primary and secondary high school education from J.M Rathi English School and high school education from Sir Parshurambhau College,Pune. He then completed his joined an engineering college in Pune and graduated with Bachelor of Engineering with majors in electronics and telecommunication engineering from Marathwada College Of Engineering, Pune.

He enrolled as a graduate student in Missouri University of Science and Technology, Rolla, MO in August 2011, majoring in Computer Engineering. He was a Graduate Research Assistant in the DERMVIS group of cancer research from January 2012 to December 2012 and received his MS degree in Computer Engineering in May 2013.