2009

Medical Infrared Image Analysis for Detecting Skin Temperature Disparities

Ramakrishna R. Arumalla
University of Massachusetts Amherst

MEDICAL INFRARED IMAGE ANALYSIS FOR DETECTING SKIN TEMPERATURE DISPARITIES

A Thesis Presented

by

RAMAKRISHNA R ARUMALLA

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL AND COMPUTER ENGINEERING

February 2009

Electrical and Computer Engineering

MEDICAL INFRARED IMAGE ANALYSIS FOR DETECTING SKIN TEMPERATURE DISPARITIES

A Thesis Presented

by

RAMAKRISHNA R ARUMALLA

Approved as to style and content by:

Patrick A. Kelly, Chair

Dennis L Goeckel, Member

Dev Vrat Gupta, Member

Hossein Pishro-Nik, Member

C. V. Hollot, Department Chair
Electrical and Computer Engineering

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Overview

Thermal infrared imaging is increasingly being used in medicine for applications that include evaluation of allergic tests, morphea, basal cell carcinoma, chilblains, port wine stains, melanocytic naevi and melanoma extensivity, deep vein thrombosis, burn depth, diabetic foot, Raynaud's phenomenon, thyroid gland changes, pneumonia development, arthropathy and many other pathological conditions [26]. The basis for all of these applications is the fact that skin temperature (the measurement produced in Infrared (IR) thermography) is a key indication of the presence of underlying medical conditions. In this thesis we will study the feasibility of using IR thermography to diagnose two conditions: (1) testicular torsion, in which the blood flow to a testicle is severely reduced; and (2) skin infection with resulting inflammation. The major work in the thesis is to identify features to consistently diagnose these conditions as currently there are no statistical models for torsion or infection detection using IR thermography .

Thermal imaging uses the fact that inflammation causes the body to produce heat and hence any abnormal increase in the body surface temperature indicates inflammation. On the other hand, a decrease in skin temperature indicates that there is a decrease in blood flow or vasomotor tone. The different activities of tissues, organs, and vessels inside an animal's body affect skin temperature, and a disease may have a unique thermal signature depending on its effect on the body [8]. For example malignancies, inflammation and infection cause localized increases in temperature which

are seen in thermographs as hot spots. For detecting infection there are no precise distribution models. This thesis will include identifying features and finding temperature distributions for various infections. We will also use the fact that under normal conditions in the human body there is contra lateral temperature symmetry and thus a temperature difference in symmetrically located regions indicates an abnormality. If infrared thermography proves to be a reliable means for diagnosing medically significant conditions, it will be very useful in a number of clinical settings (eg., the emergency room) because it is a non-invasive, non-contact technique, and because modern infrared cameras are portable, relatively inexpensive, and easy to use [4].

1.1.1 Contributions of This Thesis

We have done the first (to our knowledge) systematic study of the features of skin temperature distributions (as measured by IR cameras) that can be used to detect and classify two significant medical conditions: testicular torsion and skin infection. We have have found that the most common first order statistic (mean temperature in a skin region) often fails to be adequate for diagnosis. Even a more comprehensive marginal statistic (Kullback-Liebler distance between suspect and control temperature distributions) often fails to be a sufficiently sensitive feature. We tried a spatial statistic (pattern spectrum) that measures shape features of the temperature distribution on the skin, and found that for both torsion and skin infection it provided a much better differentiation.

Finally, we investigated the use of hypothesis testing techniques (particularly case-based reasoning) to see if further improvement in detection is possible using multiple features (pattern spectrum and marginal features). Our conclusion is that the use of other features along with the pattern spectrum did not significantly improve performance beyond the use of the pattern spectrum statistic alone.

1.2 Torsion and Skin Infection

1.2.1 Characteristics and Current Methods of Diagnosis

In this study we focus mainly on two potential applications of infrared thermography:

(a). Testicular Torsion detection: In testicular torsion the spermatic cord that provides the blood supply to a testicle is twisted, cutting off the blood supply, often causing severe pain. Prolonged testicular torsion will result in the death of the testicle and surrounding tissues. The risk of acute scrotal pain is 1 in 160 males by the age of 25 years and the incidence of testicular torsion is 1 in 4000 [27]. Because of the risk of infarction (i.e, tissue death due to lack of oxygen) surgery is typically performed in cases of acute scrotal pain when torsion cannot be excluded. An accurate means of diagnosing torsion would help to avoid unnecessary surgery. Clinical evaluation of acute scrotal symptoms is often unreliable due to small testicular size in children, presence of a reactive hydrocele (i.e., swelling due to accumulation of fluid), and lack of patient cooperation. The test used currently to evaluate acute scrotal symptoms is Color Doppler ultrasound. The advantage of using it is that one can directly visualize the testicular blood supply and also can avoid using ionizing radiation. But its use is limited because it is difficult to detect the flow in small volume. This technique is also imperfectly available to emergency providers. In this work, we assume that the loss of blood supply to the scrotum causes substantial temperature difference detectable by an infrared camera. (This assumption has been verified in previous work) [27]. We plan to make use of this difference to detect torsion using infrared thermography.

(b) Diagnosis of soft skin infections (abscess, cellulitis): An abscess is a collection of pus that has accumulated in a cavity formed in the tissue as the result of an infectious process (usually caused by bacteria or parasites) or other foreign materials (e.g. splinters or bullet wounds). It is a defensive reaction of the tissue to prevent the spread of infectious materials to other parts of the body. The final structure of the

abscess is an abscess wall that is formed by the adjacent healthy cells in an attempt to build a barrier around the pus that limits the infected material from neighboring structures and also limits immune cells from attacking the bacteria [6]. The cardinal symptoms and signs of any kind of inflammatory process are redness, heat, swelling, pain and loss of function. Abscesses may occur in any kind of solid tissue but most frequently on skin surface (where they may be superficial pustules (boils) or deep skin abscesses), and in the lungs, brain, kidneys and tonsils. Major complications are spreading of the abscess material to adjacent or remote tissues and extensive regional tissue death (gangrene). Abscesses in most parts of the body rarely heal themselves, so prompt medical attention is indicated at the first suspicion of an abscess. Cellulitis is a bacterial infection characterized by redness, swelling, warmth, and pain or tenderness. Cellulitis frequently occurs on exposed areas of the body such as the arms, legs, and face. Other symptoms can include fever or chills and headaches. In advanced cases of cellulitis, red streaks (sometimes described as 'fingers') may be seen traveling up the affected area. The swelling can spread rapidly. Currently abscess and cellulitis are diagnosed by examining the affected area or by using techniques like computer tomography (CT) or magnetic resonance imaging (MRI) for more complicated cases. These methods are often not effective or are complicated and consume a lot of time and resources.

1.2.2 Aims and Objectives of This Thesis

Infrared thermography (IRT) cameras might offer a novel diagnostic approach to both these conditions. They are capable of non-invasive, instant estimation of temperature differences that might be associated with ischemic conditions such as testicular torsion or skin temperature increase associated with inflammation. Our primary objective is to determine whether an IR camera can reliably detect testicular torsion (using a sheep model) and skin infection in humans. A secondary objective is

to see if infrared thermography might be used to distinguish between various infection types such as abscess and cellulitis by identifying the precise temperature distributions characteristic to each of them and classifying the images based on the model developed for each of these specific infections.

The expectations underlying this study are that (1) Testicles undergoing torsion will exhibit skin temperature difference (relative to the non-torsed testicle) that are detectable with an IRT camera [28]; and (2) The inflammation associated with infection of specific tissues at particular stages of the disease should be detectable by infrared thermography. In cases of infections we know that there should be contra lateral symmetry in the majority of the cases, so diagnosis amounts to determining if the suspected region of infection has a temperature distribution, relative to the contra lateral (healthy) region, that is indicative of skin infection.

In both the cases of our interest, we can assume that diagnosis is a hypothesis testing problem. We have an image of the control case (non-torsed testicular image or non-infected contra lateral skin region image) and a test image (of torsed testicular image or skin infection region image). We consider the hypotheses as:

H0: control and test images come from same distribution

H1: control and test images come from different distributions

But for this hypothesis testing we cannot define the optimal likelihood ratio test because we do not have the complete statistical image model to define the test, so the main work of the thesis is:

(i). To determine the effective statistical features for testing the hypotheses

(ii). To develop the hypothesis test for diagnosing torsion and skin infection using infrared images.

The rest of the thesis proposal is organized as follows: Chapter 2 discusses the acquisition of the data to be analyzed and the various techniques used for analysis. It gives the details of the first order statistics that were tried to differentiate between the

torsed and non-torsed images and the results obtained using them. It then discusses the inadequacy of first order statistics and an analysis of the second order (spatial) statistics that were sufficient to differentiate between the torsed and non torsed images is done. The initial results that were obtained using first order statistics are compared with the results obtained using the spatial statistics.

Chapter 3 discusses the infection detection using thermographic images. The results obtained after trying the various statistics used for torsion study on skin infection detection are presented. Chapter 4 contains the detection strategy methods that were tried using multiple features for detection. The results obtained when Case Based Reasoning and Support Vector Machines were used are included. The final chapter is the conclusion chapter that discusses a summary of the contributions the thesis has made, and a prospect for future research.

CHAPTER 2

TORSION STUDY

2.1 Introduction

The testicular torsion study makes use of IRT image data supplied by Dr. Geoffrey Capraro of Baystate Medical Center Springfield, MA. For the study, six sheep were sedated and anesthetized. A single testicle in each sheep was subjected to experimental 720 degree torsion. Doppler Ultrasound was used to confirm the presence and absence of blood flow. The animal was placed in the prone position to allow IRT image capture from a tripod at 11 inches, and images were captured pre-procedure, and every 15 min for six hours after the procedure. After six hours the torsion was corrected to restore blood flow, and IRT images were taken every 15 min for 75 min after reduction of torsion.

The images were taken using a $FLIR\_Systems\_B2$ camera and transferred to an attached computer. In our processing, image regions corresponding to each of the testicles were manually segmented and the temperature information was extracted using MATLAB. The temperature information for each testicle was stored as a .CSV file. To test the usefulness of infrared thermography for diagnosing torsion we need to define features that distinguish torsion/non-torsion conditions in infrared images.

2.2 Technical Details

As mentioned previously the primary objective of this study is to address a two-fold problem, namely: (i). To determine the effective statistical features for testing

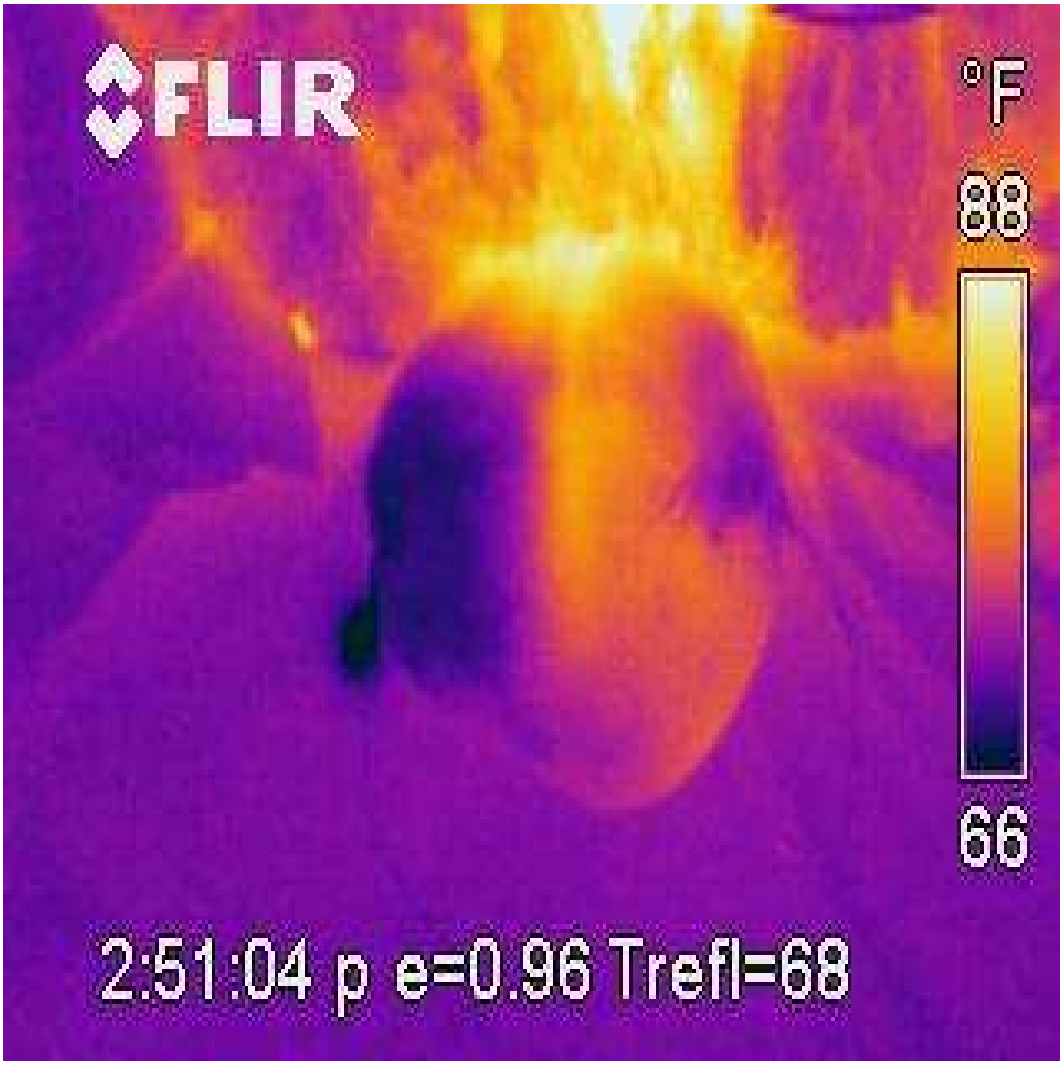
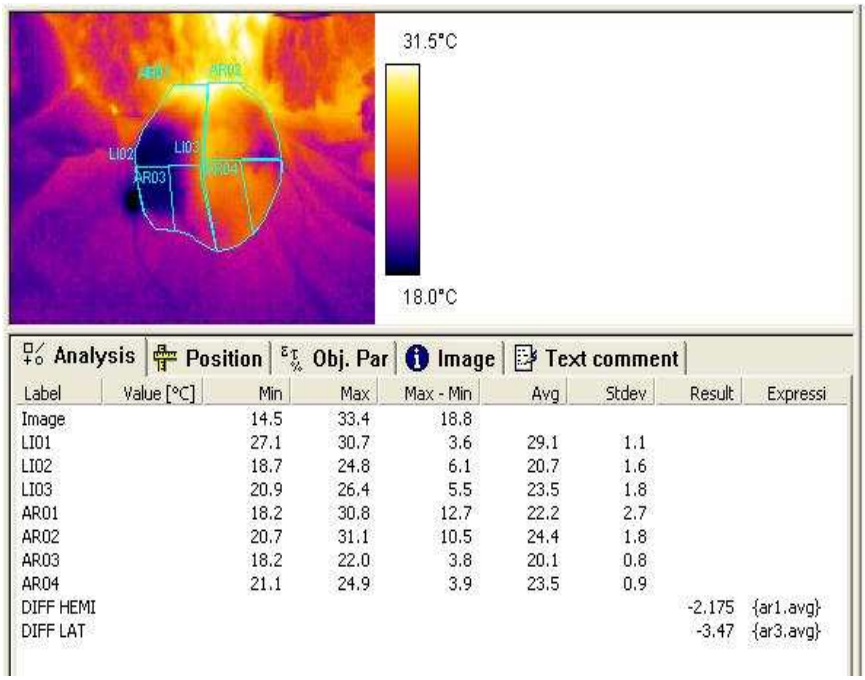


Figure 2.1. Original image from FLIR camera

the hypotheses; and (ii). To develop a hypothesis test for diagnosing torsion using infrared images.

Initially we tested a number of first order statistics to see if they could differentiate between the torsed and the non-torsed images. We used the partial least squares method [14] to determine the test statistics that are most effective (e.g., sample variance) to distinguish the torsed and non-torsed images. This led us to eliminate some statistics as not useful for discrimination. Our preliminary results showed that even



Label	Value [°C]	Min	Max	Max - Min	Avg	Stdev	Result	Expressi
Image		14.5	33.4	18.8				
LI01		27.1	30.7	3.6	29.1	1.1		
LI02		18.7	24.8	6.1	20.7	1.6		
LI03		20.9	26.4	5.5	23.5	1.8		
AR01		18.2	30.8	12.7	22.2	2.7		
AR02		20.7	31.1	10.5	24.4	1.8		
AR03		18.2	22.0	3.8	20.1	0.8		
AR04		21.1	24.9	3.9	23.5	0.9		
DIFF HEMI							-2.175	{ar1.avg}
DIFF LAT							-3.47	{ar3.avg}

Figure 2.2. Preliminary Image analysis using FLIR software.

the first order statistics that were identified as most useful were not adequate to distinguish between the torsed and the non-torsed images in several test cases. However using the first order statistics we could distinguish between the images in cases where the external noise was minimal, so the inadequacy of the first order statistics might be because of the noise (presumably due to the specific testing conditions) in the images. We next tried a spatial statistic called the pattern spectrum, which not only differentiates the temperature disparities but also makes a differentiation based on the spatial distribution of the heat. The results based on the pattern spectrum show that using that statistic we could distinguish between the torsed and non-torsed images in 28 out of 31 cases (91% of the cases) in one of the sheep, with similar performance levels in other cases.

Figure 2.3. Left Testicle.

Figure 2.4. Right testicle.

2.2.1 First Order Test Statistics

1. **Mean temperature difference.** The means of the temperature readings of the two sides were calculated and compared. The side with the torsion is expected to have a lower mean temperature. This is a very simple and commonly used test statistic but not always effective.

2. **Kullback-Liebler divergence (histogram measure).** The Kullback Leibler (K-L) divergence is a measure of the divergence between two probability distributions: from a true probability distribution to an arbitrary distribution. Typically the true distribution represents data, observations, or a precise calculated probability distribution. The arbitrary distribution represents a theory, model or an approximation of the true distribution [7] . In our case we generate the histograms of the temperature distributions of both testicles and normalize them. This creates the distributions for us. We take the normalized histogram obtained from the side with suspected torsion as the true distribution and the normalized histogram from the control side as the arbitrary distribution to calculate the Kullback Leibler divergence.

$$D_{KL}(P \parallel Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

where i images over the temperatures in the images , P(i) is the true distribution, and Q(i) is the arbitrary distribution . This test statistic measures difference in the entire histogram.

3. **Wald Wolfowitz test.** This is a nonparametric test for testing the null hypothesis that the distribution functions of two populations are the same. The observations (in our case the pixel temperature values) are taken from the respective sides and arranged in increasing order of magnitude, irrespective of the side they came from. Each value is then replaced by 0 or 1, depending on the testicle from which it was taken. The total number of runs of like elements is then counted and used as a test statistic. If the two populations differ among themselves, then elements of one type (0s or 1s) would be expected to cluster together and the total number of runs will be small. If the populations are identical then the arrangements of 0s and 1s should be random and the total number of runs should be large [27].

2.2.1.1 Preliminary Results

The various test statistics were calculated for each set of the sheep images and the results plotted.

The mean and K-L divergence were calculated for selected pairs of images for each sheep and scaled using the standard deviation. The images selected were images taken from 120 minutes to 330 minutes, after the torsion started (these represent cases where one testicle is torsed), and one pair of images before the torsion was started and 2 pairs of images that correspond to 45 and 60 minutes after the blood supply

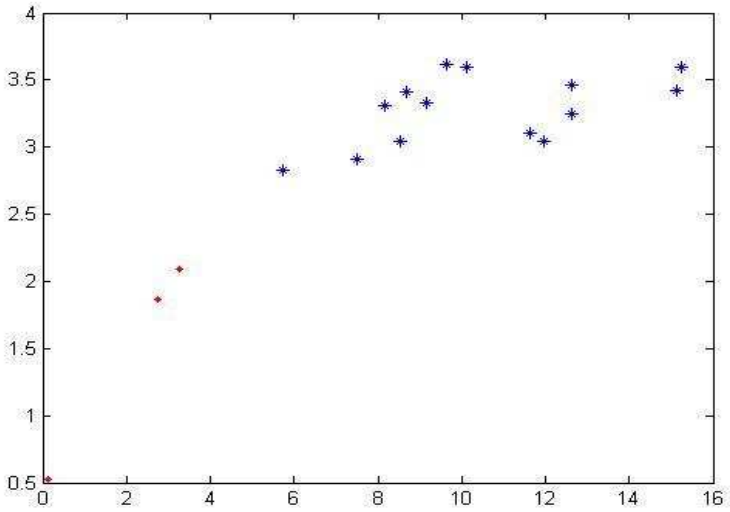


Figure 2.5. Sheep 6 KL vs MD for selected cases. Images with torsion are indicated in blue and images with no torsion in red.(The X-Axis represents the Kullback-Liebler distance and the Y-Axis represents the Mean Difference.)

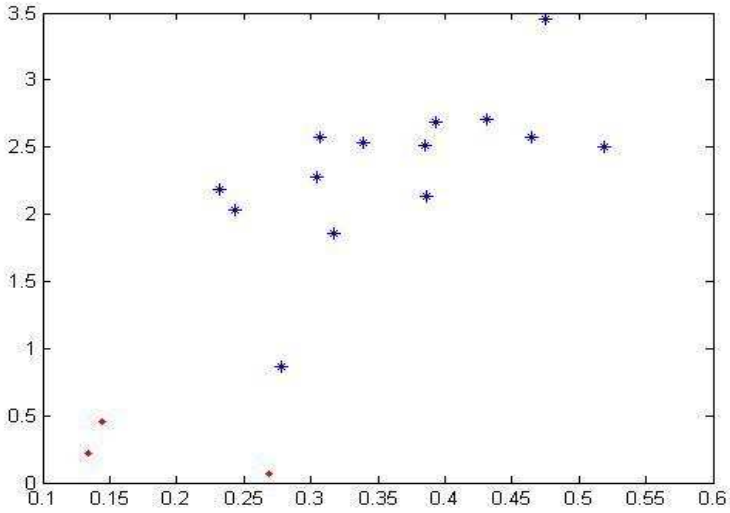


Figure 2.6. Sheep 4 KL (X-Axis) vs MD (Y-Axis) for selected cases. The blue points represent torsed image pairs and the red points represent the non-torsed image pairs.

was restored (these are used to represent cases where there actually is no torsion). From the graphs we can see that using the metrics, mean temperature difference and Kullback-Leibler divergence we can easily distinguish (i.e, separate the points corresponding to) the pairs of images with torsion (blue) and without torsion (red) in several cases.

However when all the pairs of images taken for a particular sheep were considered and plotted we found certain cases (e.g., sheep 4, see Figure 2.6) in which we cannot completely distinguish torsion/non-torsion .

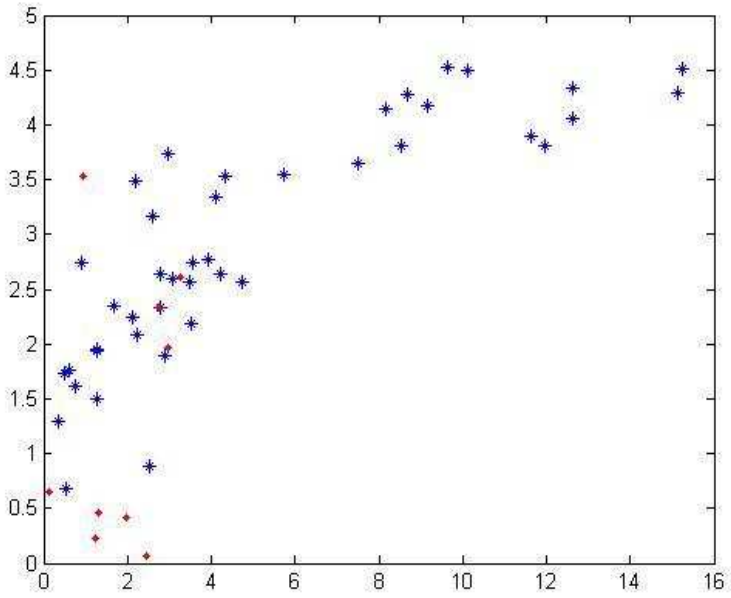


Figure 2.7. Sheep 4, 6 and 7 KL vs MD (Selected cases from sheep 4, 6 and 7). Blue Points represent images with torsion and red points images without torsion.

When we plot the metrics for images of sheep 4, 6 and 7 treated as a group, we get a plot where we can see that we cannot easily distinguish the pairs of images with torsion and images without torsion (see Figure 2.7). This problem might have arisen because of the inconsistency in taking the images for different sheep under conditions that could not be controlled.

Similar results were found using the Wald-Wolfowitz statistic - it successfully discriminated the conditions in some cases, but failed in several others. The inadequacy of these first-order statistics led us to consider a spatial statistic based on mathematical morphology, which we describe next.

2.2.2 Spatial Statistics

The preliminary results on torsion images indicate that marginal distribution statistics may not be sufficient to distinguish the torsion images from the non-torsion images in all cases. Hence, we consider a statistic that not only accounts for the differences in the marginal temperature differences but also the difference in the temperature distribution pattern. We expect that spatial pattern of temperatures may differ in torsion/non-torsion cases even when the marginal distributions are similar. In particular, we expect higher temperatures to be aligned above blood vessels in non-torsion control images and be more scattered (due to noise) in torsion images.

1. **Pattern spectrum.** Mathematical morphology is a method to quantitatively describe processing operations that are sensitive to the shape of objects in an image. It is sometimes used for modeling the processes of human recognition of visual information. Mathematical morphology describes such operations by combinations of basic set operations between an image and a small object called a structuring element. The pattern spectrum is an application of mathematical morphology. The pattern spectrum extracts the size distribution of objects contained in an image by decomposing the target image into objects of various sizes whose shapes are similar to the structuring element [9]. Mathematical morphological operations are shift-invariant image manipulations and can be decomposed into two simple basic operations - dilation and erosion. We first explain these basic operations in terms of binary structuring elements for binary images. These operations are defined as set operations. An image is assumed

to be a set of pixel positions that constitute image objects (represented as 1's in binary images). Let X denote a set representing an image. Let B be another set called a structuring element (SE). The structuring element is used to define morphological filtering operations, based on which Minkowsky set subtraction and addition are defined as follows:

$$\text{subtraction:} \quad X \ominus B = \{x | B_x \subseteq X\}$$

$$\text{addition:} \quad X \oplus B = \{x | B_x \cap X \neq \emptyset\}$$

where B_x denotes translation of B by x, defined as

$$B_x = \{z + x | z \in B\}$$

The operations of erosion and the dilation are defined as $X \oplus B^v$ and $X \ominus B^v$, respectively, where B^v is defined as $B^v = \{-x | x \in B\}$.

The effects of dilation and erosion are schematically illustrated in Figure 2.8. Clearly, dilation expands the original image while erosion shrinks the original image . These binary operations can be extended to morphological operations for gray scale images by introducing the concept of umbra. Consider a function $X(y)$, which expresses a grayscale image, where $y \in R^n$ and consider an R^{n+1} coordinate system for $(y, X(y))$. In most cases for image processing, $n = 2$. In this case y is regarded as a pixel position and $X(y)$ as the pixel value at y. The umbra of a function is defined in this R^{n+1} coordinate system as the set of all coordinate points lower than the value of the function over the extent of the function. Using this expression, the morphological operations of grayscale images are reduced to binary operations on the umbrae of the function for the image [10]. In the case that the structuring element is binary, the structuring

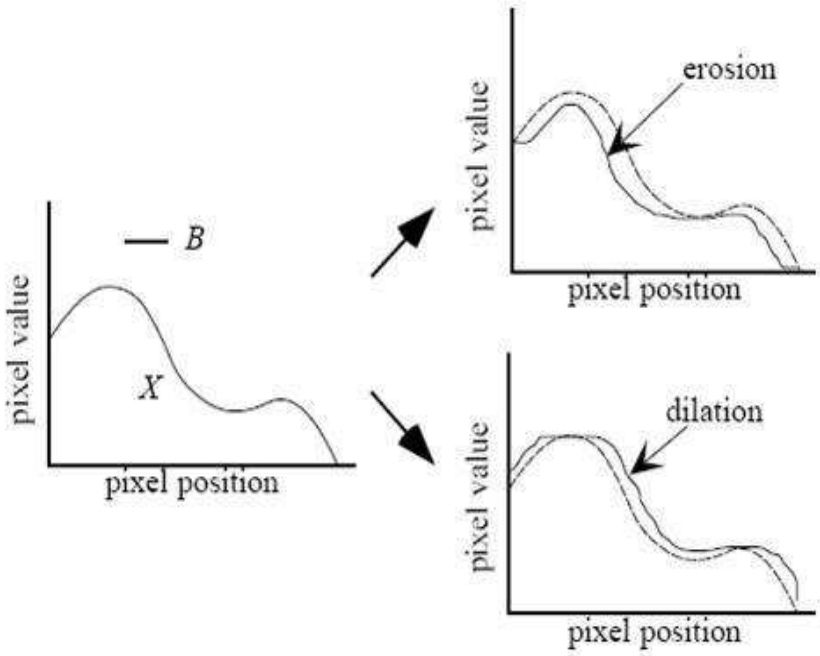


Figure 2.8. Erosion and dilation on grayscale image by binary structuring element. (One dimensional case where images X is the solid curve).

element is defined as an area above the y axis. The mathematical operations of erosion and dilution are defined similarly to the case of binary images. The results of dilation and erosion of a grayscale umbra by a binary structuring element are depicted in Figure 2.8.

Two other important and basic operations, called opening and closing, are derived from the erosion and dilation. The opening is defined as erosion followed by dilation by the mirrored SE of the original SE used for the erosion. The closing is also defined as dilation followed by erosion. More formally, opening and closing are defined as follows:

$$\text{Opening:} \quad X_B = X \ominus B_v \oplus B$$

$$\text{Closing:} \quad X^B = X \oplus B^v \ominus B$$

The opening eliminates the portions of image objects smaller than the SE while preserving the other portion of the objects. The closing is the complement: it fills up smaller spots while preserving the other portions of the background. Openings and closings can be used to quantify the size distribution of objects, which is the idea behind the definition of the pattern spectrum. The pattern spectrum of size n (defined below) by an SE is defined as the pixel-wise difference between the target image opened by a homothetic set of a SE of size n and that opened by SE of size $n+1$. Let nB be the homothetic set of a structuring element B of size n, defined as follows:

$$nB = B \oplus B \oplus ... \oplus B\{(n-1)additions\}$$

$$0B = \{0\}$$

Then the pattern spectrum of size n by the SE B for image X, denoted as $PS(X, B, n)$, is defined as follows:

$$PS(X, B, n) = \sum_{y \in \text{whole image}} \{X_{nB}(y) - X_{(n-1)B}(y)\}$$

where "-" sign denotes the pixel-wise difference. Since the opening removes the portion smaller than the SE, the difference of the images opened by the SEs of size n and size $n+1$ contains the portion whose size is exactly n. The original image is thus decomposed into sets of the SEs of various sizes. The normalized pattern spectrum, which is defined as the ratio of the original pattern spectrum to the sum of the pixel values over the whole original image, is often used. The spectral values of the normalized pattern spectrum indicate the ratio of the portions of a given size to the whole image [10].

When applied to the torsion test images, we expect that in images of non-torsed testicles the temperature pattern will exhibit relatively smooth peaks

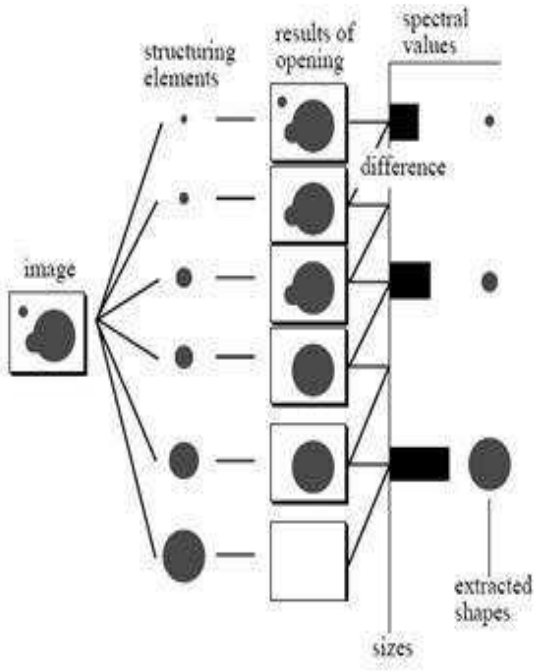


Figure 2.9. Figure illustrating the process to calculate the pattern spectrum.

corresponding to skin regions above blood vessels. Hence, when we do the opening of the images with structuring elements of growing sizes, we expect the differences calculated between subsequent openings to be relatively high compared to the testicle images with torsion. We define a test statistic which is the sum of the differences in subsequent openings normalized by the total number of pixels in the image-

That is: Test statistic = $\sum$ [(pixel wise difference in opening by n and opening by n-1)/total number of pixels in that image].

2.2.2.1 Preliminary Results

The preliminary results show that there are cases where the previous statistics could not distinguish torsion but pattern spectrum was very effective. In the case of sheep 7 where the first order statistics failed completely to distinguish the torsion

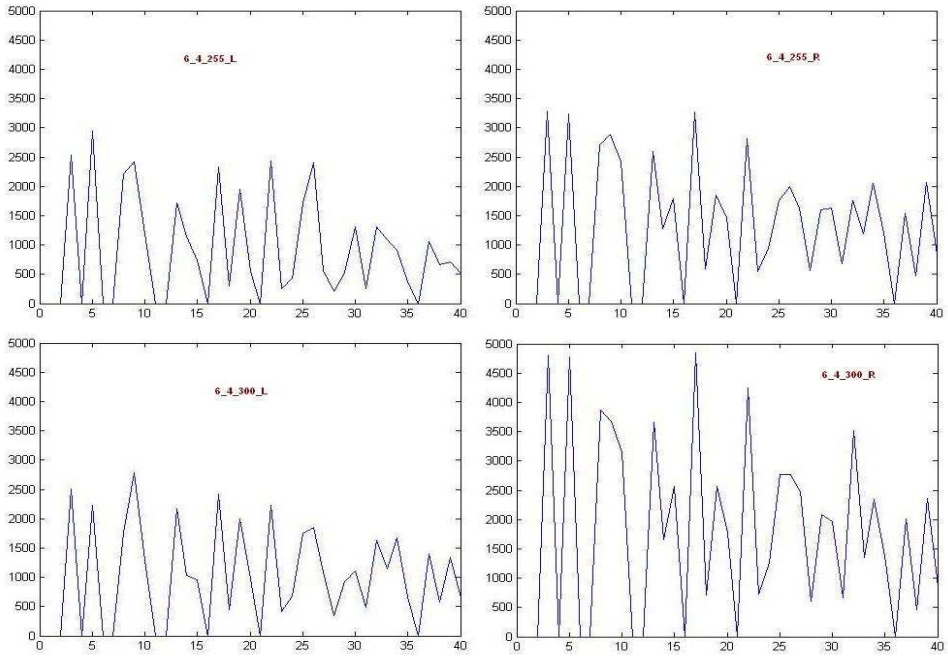


Figure 2.10. Plot of size of opening vs. pixel wise difference. The X-Axis represents the size of the opening and the Y-Axis is the pixel wise difference. We can see that there is a difference in in the pixel wise difference for various size of openings when we compare the results for left and right testicle images but less difference when we compare the results for two left testicle images or the right testicle images. We use this information to differentiate between the torsed and non-torsed images. (The images on the left are the results obtained for the left testicle images taken at 255 minutes and 300 minutes and the images on the right for the images of the corresponding right testicles).

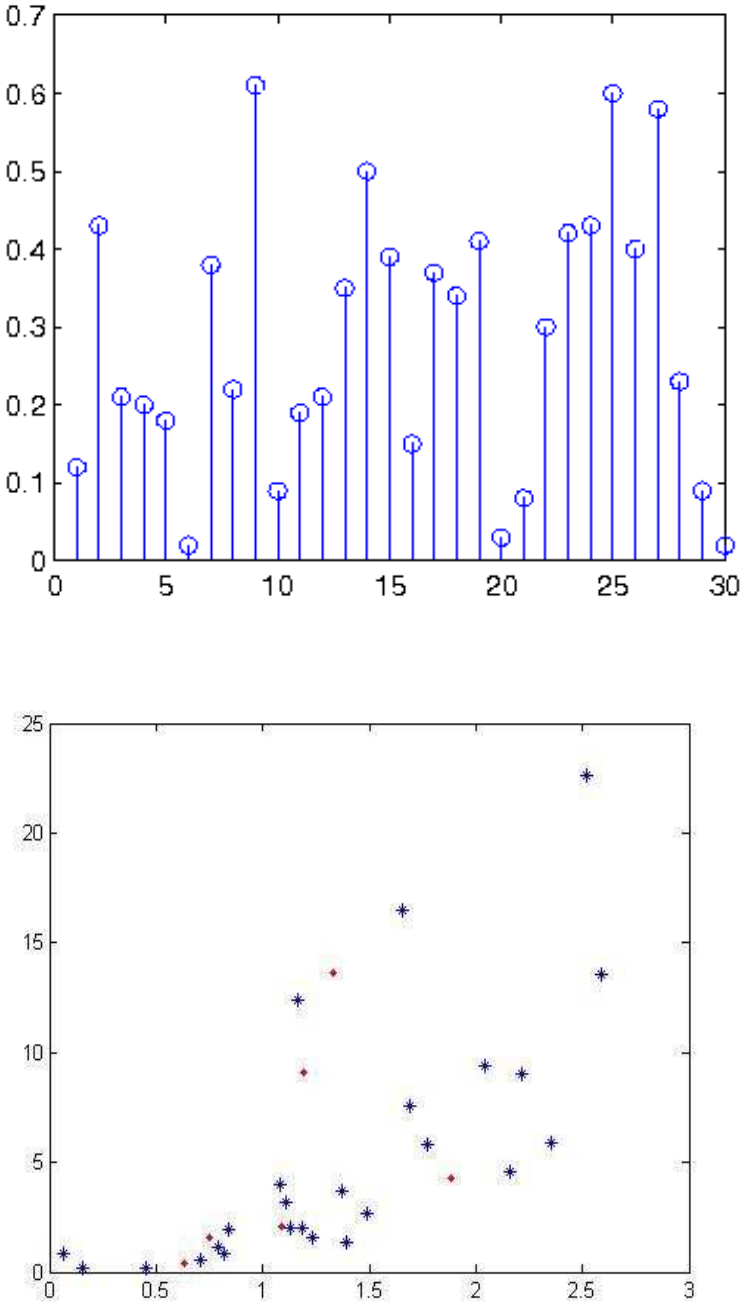


Figure 2.11. Comparing plots of Pattern Spectrum and KLvsMD for sheep 7. We can see that by thresholding the graph in the Pattern Spectrum plot one can distinguish the torsed cases from the non-torsed cases. In the pattern spectrum graph the Y-Axis represents the difference in the test statistic of the left and right testicles in each case. In the KL Vs MD graph, the X-Axis represents the KL distance and the Y-Axis represents the Mean Difference. We can see that we cannot isolate the torsed cases from the non-torsed cases.

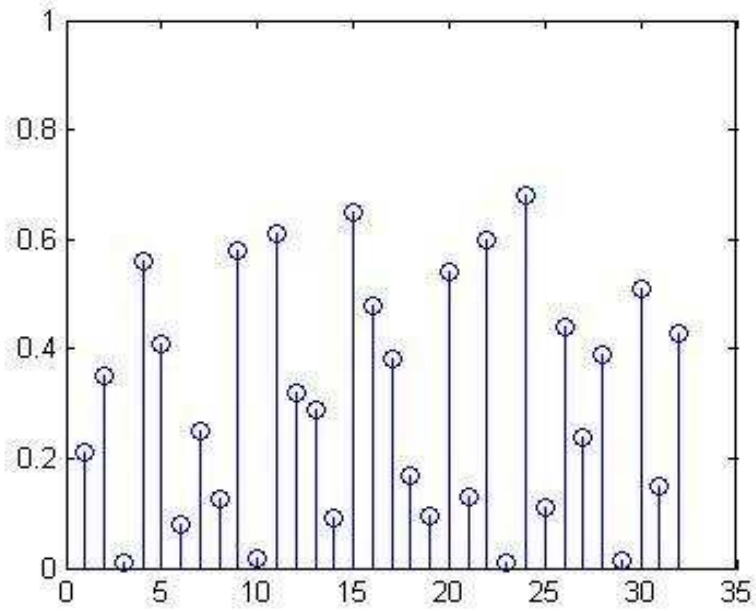


Figure 2.12. Pattern Spectrum plot in case of Sheep 6.

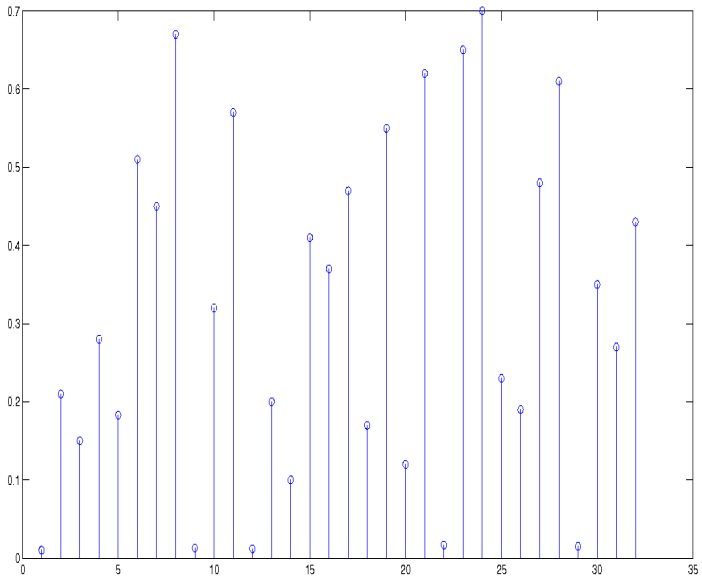


Figure 2.13. Pattern Spectrum plot in case of Sheep 4.

side from the non-torsion side pattern spectrum analysis was effective in 28 out of 31 cases by simple thresholding of the test statistic . Also pattern spectrum analysis was used in two other sets of sheep images where it was effective in 28 out of 32 cases in one and 29 out of 32 in the other set of sheep images.

In the three sheep cases we considered we can threshold the pattern spectrum statistic around 0.1 and segregate the torsion images from the non-torsion images. In sheep 6 it is approximately 0.18 and in case of sheep 4 it is approximately 0.12.

The reason why pattern spectrum analysis was effective could be that heat is distributed according to the blood flow in the blood vessels and follows a pattern. When the blood supply was cut off during torsion, there was no heat from the blood vessels and the temperature pattern could have been changed because the heat had been dissipated in the tissue but there is no heat supply from blood flow. This change in the pattern is detected and quantified in the pattern spectrum analysis enabling us to distinguish torsed images from non-torsed ones. A similar effect is expected in infection images where the temperature pattern would change due the increase in heat in the region of inflammation.

CHAPTER 3

SOFT SKIN INFECTION DIAGNOSIS

3.1 Introduction

The soft skin infection study is based on a sample of patients suspected to have either cellulitis or abscess, seeking care in a tertiary, urban ED. The study was approved by the IRB and participants and guardians provided consent and assent. Digital infrared images were captured from infected and corresponding contra-lateral regions of interest using a Flir B2 IRT camera (Flir Inc. Billerica, MA) after one minute of unclothed exposure to room temperature. Skin markers were applied at the margins of the infected region to demarcate the region of interest and facilitate post hoc image analysis. Images were stored in the camera's on-board memory and image analysis deferred until study completion.

Localized heat production is a cardinal feature of inflammation broadly and infection specifically. At present, the determination of localized heat is a clinical, bedside maneuver, and emergency providers make no attempt to quantify this sign. IRT technology is capable of quantifying heat in the infrared range of the electromagnetic spectrum and encoding this data in the pixels of a digital image. IRT cameras could allow non-invasive quantification of asymmetric temperature. This investigation was performed as a proof of concept, which if proven, might be followed by more sophisticated image analysis, and investigation of more challenging infections [28].

3.2 Data Collection

The images were taken using a $FLIR\_Systems\_B2$ camera and transferred to an attached computer. In our processing, image regions corresponding approximately to the infected area were manually segmented and the temperature information was extracted using MATLAB. The temperature information for each segmented image was stored as a .CSV file. To test the usefulness of infrared thermography for diagnosing skin infection we need to define features that distinguish infection/non-infection conditions in infrared images.

A bed mounted tripod assembly was used to capture the IRT digital images. The images were captured from a fixed distance of three feet. Images were captured from both the infected and contralateral unaffected sides. Images were taken of the dry skin, fan-cooled skin and saline- and fan-cooled skin. The margins of the tenderness of the infection were marked with metal beads.

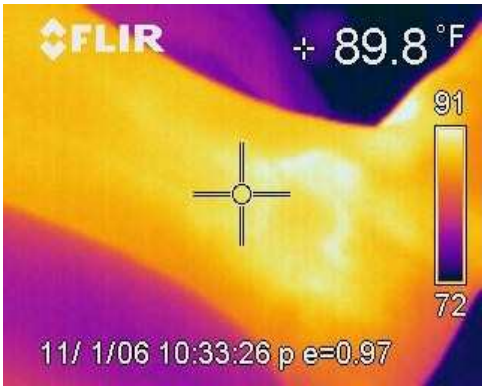


Figure 3.1. Left Ankle (Control Side).

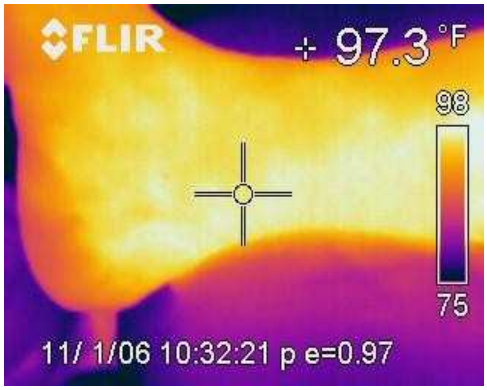


Figure 3.2. Right Ankle (Side with cellulitis infection).

3.2.1 Infection Verification

Ultrasound was performed after image capture to establish the presence or absence of an abscess, and to measure the size of the abscess. This ultrasound measurement

was used as a "gold standard" for determining presence of infection on the day of image capture [28].

3.3 Technical Details

The primary objective of this study is: (i). To determine the effective statistical features for testing the hypotheses; and (ii). To develop a hypothesis test for determining whether non-draining, untreated soft tissue infections presenting to the emergency department (ED) are associated with increased localized heat as determined by an infrared thermography (IRT) camera.

Initially we tested the first order statistics used in the torsion study to see if they could differentiate between the infected and the non-infected images. We expected a better performance with the skin infection images as these were taken from a calculated distance and angle and also care was taken to remove the external noise. Our preliminary results showed that the first order statistics were not adequate to distinguish between the infected and the non-infected images in several test cases.

However using the first order statistics we could distinguish between the images in cases where the focus of the images was good and the images not zoomed in too much. Also cases in which the region of interest was marked using beads showed better performance. Hence we can conclude that human factors like zooming and focussing on the subject of interest played an important role in the success of the first order statistics. We next tried the spatial statistic, the pattern spectrum to see if that would eliminate the role of the human factor. The results based on the pattern spectrum show that using that statistic we could distinguish between the infected and non-infected images in 78 out of 81 cases (96% of the cases).

3.3.1 First Order Test Statistics

The various test statistics that were used in the torsion study were calculated for each set of the infection images and the results plotted.

3.3.1.1 Preliminary Results

1. **Mean temperature difference.** Mean temperature within the region of interest was determined using the manufacturer's analysis software, by '"drawing" a best-fit polygon oval bounded by the skin markers around the infected site and corresponding contra-lateral region. The difference in mean temperatures, infected minus control, was calculated and analyzed. The side with the infection is expected to have a higher mean temperature because of the inflammatory nature of skin infections . When we considered the set of images in which the infection region was clearly marked with metal beads (temperature different from skin) there was a significant difference in the mean temperature of the infected image and non-infected image. Except for one case out of thirty one we can easily distinguish the infected cases from the non-infected cases.

   But when we considered all the images, including the ones without markers it was difficult to distinguish using just the mean temperature statistic.

2. **Kullback-Liebler divergence (histogram measure).** The Kullback Leibler (K-L) divergence (defined in Sec.2.2.1) was calculated for the sets of images and plotted along with the mean-difference. The K-L distance could distinguish twenty six out of the thirty one sets of images correctly.

   The mean and K-L divergence were calculated for the pairs of images for each subject and scaled using the standard deviation. When calculating the parameters we assumed both cellulitis and abscess as just infection and did not distinguish between them. From the graphs we can see that using the metrics, mean temperature difference and Kullback Leibler divergence we can easily dis-

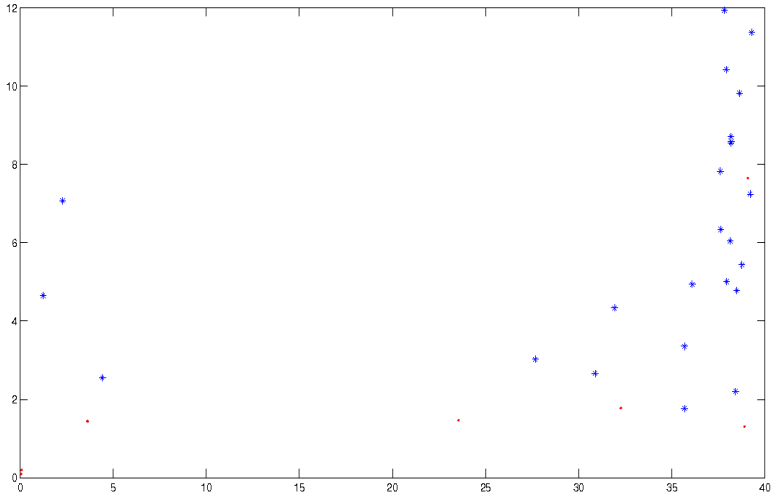


Figure 3.3. KL vs MD for selected cases where metal beads were used to mark the infected region. (The X-Axis represents the Kullback-Liebler distance and the Y-Axis represents the Mean Difference). The red points represent cases with no infection and blue points cases with infection.

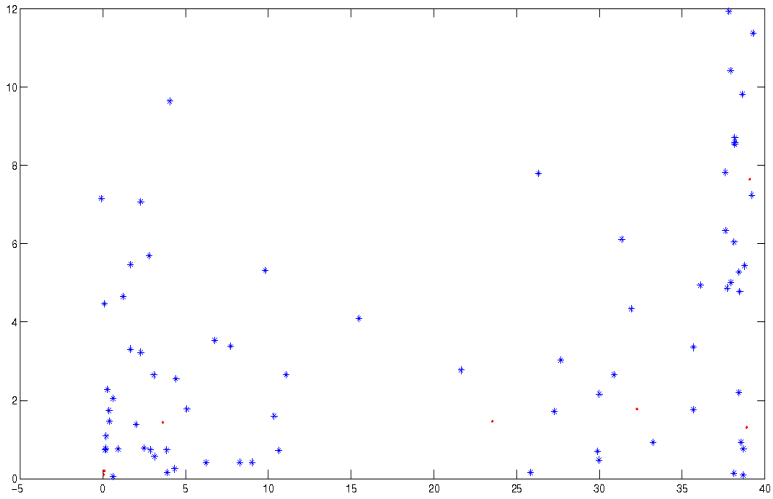


Figure 3.4. KL (X-Axis) vs MD (Y-Axis) for all cases. The red points represent cases with no infection and blue points cases with infection.

tinguish (i.e, separate the points corresponding to) the pairs of images with infection (blue) and without infection (red) in 93.5 % of the cases where the infection region was marked clearly.

3. **Wald Wolfowitz test.**

   When we tested using the Wald-Wolfowitz statistic - it successfully discriminated the conditions in some cases, but failed in several others. Hence, we deemed it not useful as a detection statistic.

3.3.2 Spatial Statistics

The results on the infected images clearly indicate that marginal distribution statistics may not be sufficient to distinguish the infected images from the non-infected images when we consider all the cases. Hence, we consider second order statistic pattern spectrum in the study. In particular, we expect high temperature at the center of the infection gradually decreasing as we move towards the non infected healthy region. We might also find a difference in the patterns of abscess and cellulitis infections.

3.3.2.1 Preliminary Results (Pattern Spectrum)

The pattern spectrum was calculated for each of the pair of images (the segmented region of the infected side and the contralateral symmetric side) and compared.

From the figure Figure 3.5 we can see that when the infection region is marked with beads in the images, the pattern spectrum of the infected side and the contralateral symmetric side differs significantly. We can attribute this to the fact that we can capture the area of interest and remove the unwanted (equivalent to noise) region from our analysis. In figure Figure 3.7 we can see that when the infected region was not marked we considered a region that included more than the infection area and

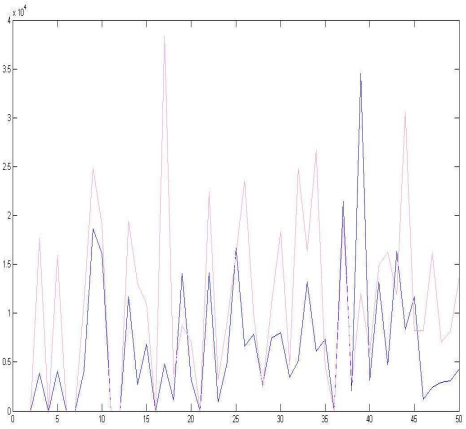


Figure 3.5. Pattern Spectrum calculated in the case where there was infection on one side and the region marked with beads.

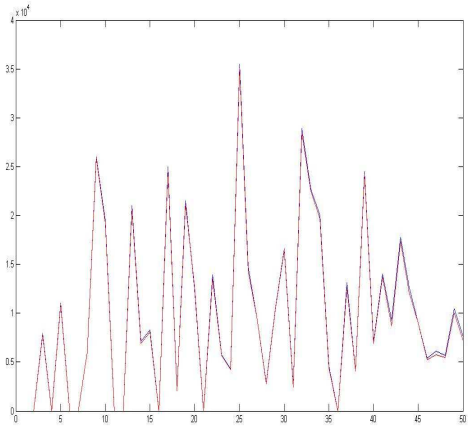


Figure 3.6. Pattern Spectrum calculated in case where there was no infection.

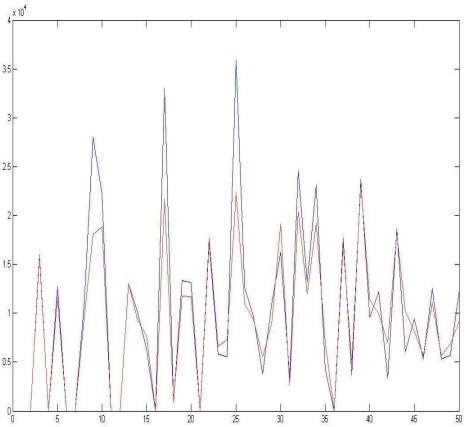


Figure 3.7. Pattern Spectrum calculated for a case where there was infection but region was not marked (had to consider a larger area when calculating) .

hence the difference in pattern spectrum is less than that of the case where the region was marked.

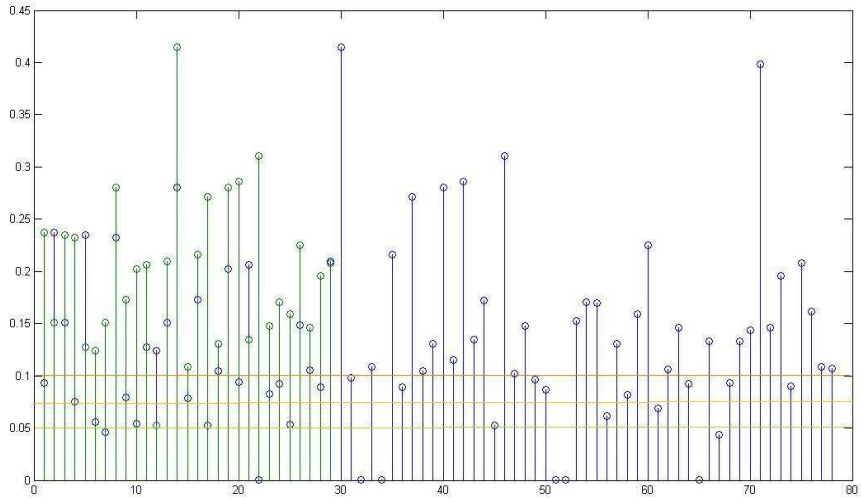


Figure 3.8. Pattern Spectrum calculated for all cases.

When we plot the K-L distance between the pattern spectrum statistics for the images, we can see (Figure 3.8 that we can distinguish the infected cases and non-infected cases at a threshold of 0.05. When we keep the threshold at 0.05 there is one case where there is infection but not detected but the value is very close to 0.05. Also there is one false positive. Instead of taking all the cases into consideration, if we take just the case of the images where the infected regions are marked using the beads then we can easily raise the threshold to a value greater than 0.1 and miss one case.See Figure 3.9. Also in this case there is no false positive. This one case is missed because the image was not focussed when taken and could easily be rectified in all future studies or real life.

The reason that pattern spectrum worked is probably due to the fact that the heat is distributed radially when there is inflammation due to infection and pattern spectrum captures this heat distribution pattern. First order statistics would capture

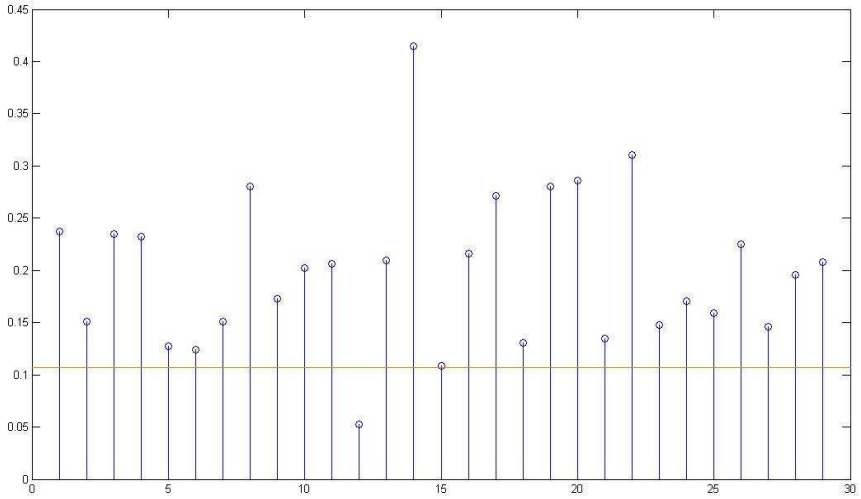


Figure 3.9. Pattern Spectrum calculated for Images with infected area marked.

the measure of heat and would miss the cases when the difference in the heat measure is too little to distinguish or if the difference is lost in external noise.

CHAPTER 4

DETECTION STRATEGY

As shown in the previous chapters, use of the pattern spectrum difference as a test statistic enables effective detection of temperature distribution disparities in a large majority of our test cases. However, even that statistic alone does not give perfect detection. In this chapter we investigate detection strategies that use multiple statistics in an attempt to improve performance.

The key for detection strategy would be to define a test with very small probability of false alarm with probability of detection almost equal to one. Since we have a very small data set we need to use a detection strategy that uses all the features available from the data and gives an effective test. Since we do not have distributions for features so the detection strategy must include training to 'learn' best division rules. We had tried two methods for detection strategy, Case Based Training and Support Vector Machines.

1. **Case based reasoning (CBR) :** The CBR classifier compares a test case to a reference (training) collection of cases, and identifies similar cases. Euclidean distance measure is used to determine the similarity between the test case and the reference cases. Suppose there are n scalar features, denoted $\phi_i i = 1, 2, ..., n$. The distance between a test case and a reference case is then

$$D_{Eucledian}(test, ref) = \sqrt{\sum_{i=1}^{n} |\phi_{i_{test}} - \phi_{i_{ref}}|^2}.$$

(Note that this weights each of the features equally.) Given this distance between the test case and a reference case, the two cases are judged to be similar

if the distance between them is less than a specified similarity threshold. In the search for the best similarity threshold, we start at 0 where few cases are similar and continue increasing the threshold. As the similarity distance increases, the CBR categorizes more cases as similar. The ''best" similarity threshold can be obtained by exhaustively examining all possible thresholds to draw the Receiver Operating Characteristic (ROC) and for example, choosing the threshold that maximizes partial ROC area (0.90AUC) (where AUC is defined as the area-under-curve. The expected performance of a classifier can be characterized by the AUC. A perfect classifier has AUC = 1).

A receiver operating characteristic (ROC), or simply ROC curve, is a graphical plot of the sensitivity vs. (1 - specificity) for a binary classifier system as its discrimination threshold is varied. The ROC can also be represented equivalently by plotting the fraction of true positives (TPR = true positive rate) vs. the fraction of false positives (FPR = false positive rate)

In our case we used both the first order statistics and pattern spectrum measure as the features, and tested the similarity between a known image of infection region and an image of the contralateral symmetric region. In our case the $\phi_i s$ are the mean difference, the K-L distance and pattern spectrum statistic. Then a threshold was determined to predict if the test case had infection or not and the probabilities of false alarm were calculated. Also we repeated the same process but used an image without infection as the test image.

In the Table sensitivity measures the proportion of actual positives which are correctly identified as such (i.e. the percentage of infection images that are correctly identified as having infection); and the specificity measures the proportion of negatives which are correctly identified (i.e. the percentage of non-infected images that are identified as not having infection).

Threshold	Sensitivity	Specificity
1	.3514	.8571
1.5	.6081	.8571
2	.7568	.8571
2.5	.8378	.8571
3	.8784	.8571
3.5	.8919	.8571
4	.9459	.1429
4.5	.9459	0
5	.9595	0
5.5	.9730	0
6	.9730	0
6.5	.9730	0
7	1	0
7.5	1	0

Table 4.1. Sensitivity and Specificity values for different thresholds.

$$Sensitivity = \frac{numberofTruePositives}{numberofTruePositives + numberofFalseNegatives}$$

$$Specificity = \frac{numberofTrueNegatives}{numberofTrueNegatives + numberofFalseNegatives}$$

$$FalsePositiveRate = 1 - Specificity$$

The ROC was plotted and AUC calculated. The exact AUC was calculated using:

$$AUC = \sum_i \frac{(y_{i+1} + y_i) * (x_{i+1} - x_i)}{2}$$

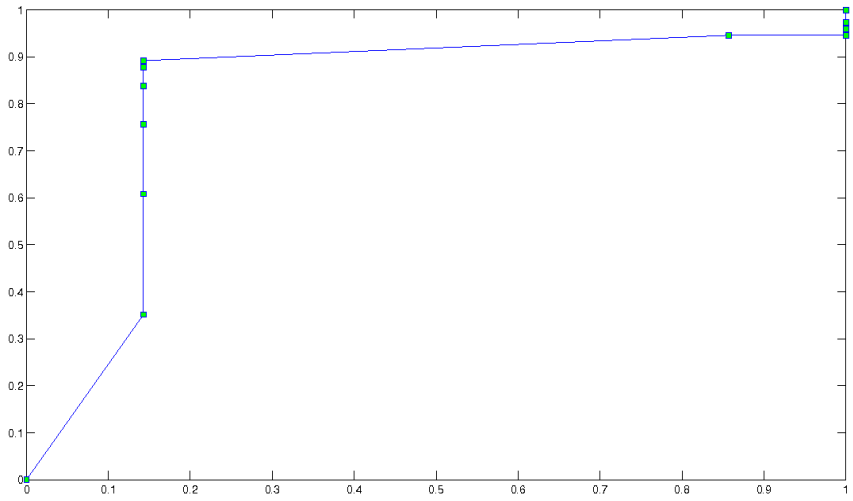


Figure 4.1. ROC when the test case had infection and all others were used as reference cases.

In the case when the test case used had infection and the rest of the images were used as reference cases the AUC was calculated to be 0.8166 which can be classified as good. In this case the detection strategy was to find out whether the test image had infection. The threshold corresponding to 0.9*AUC lay between 3.5 and 4.0, with probability of false alarm 0.1429 and detection probability 0.8919. Note from the ROC that incorporating other features in addition to the pattern spectrum provided little (if any) improvement in performance beyond the use of pattern spectrum alone.

The ROC was plotted with the detection strategy to find if the test case did not have infection. This would mean that a false positive would be an infected

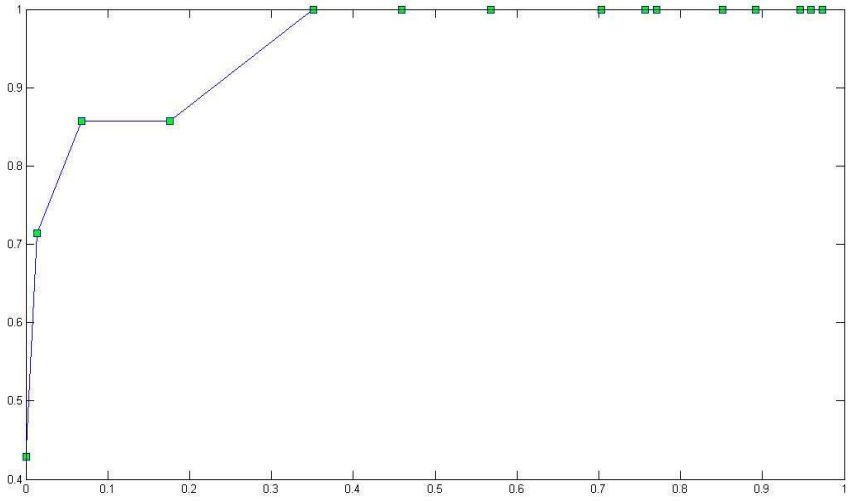


Figure 4.2. ROC when the test case had no infection and rest all were used as reference cases.

image classified an a non-infected image. In this case the AUC was calculated to be 0.9546 which can be classified as excellent and the threshold corresponding to 0.9*AUC was found to be between 6.5 and 7.0 with probability of false alarm 0. The probability of detection in this case is around 0.45 and hence might not be a useful threshold.

2. **Support Vector Machine (SVM):** Binary Support Vector Machines are based on a decision-hyper plane heuristic that attempts to impose a training instance void, or '"margin,"' around the decision hyper plane. Feature vectors are denoted by x_{ik}, where index i labels the M feature vectors ($1 \leq i \leq M$) and index k labels the N feature vector components ($1 \leq i \leq N$). For the binary SVM, labeling of training data is done using label variable $y_i = \pm 1$ (with sign according to which class the training instance is from). The optimal separating hyperplane is one that maximizes the "margin" (i.e, minimum distance)

between the data classes. Training the basic linear SVM formulation can be extended to non-linear separating curves. See [16] and [19] for more details.

Once training is complete, discrimination is based solely on position relative to the discriminating hyper plane.

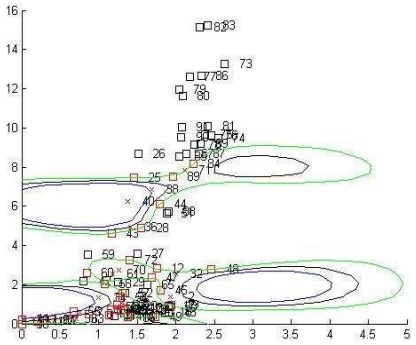


Figure 4.3. SVM classification using 3 sets of sheep.

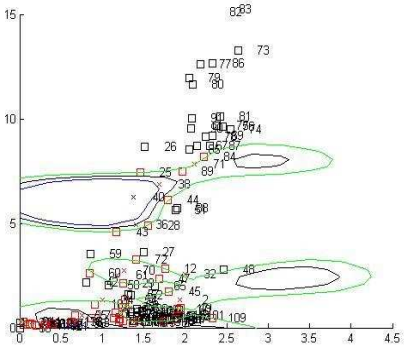


Figure 4.4. SVM classification using 4 sets of sheep.

On the testicular torsion data described in Chapter 2, we tested SVMs using the mean difference and KL Distance as features. The seperating curves were generated using the Radial Basis Kernel function.

In the graphs the areas marked in blue are the regions in which the pairs of images with no torsion fall, and the areas outside the green lines are the area in

which the pairs of images with torsion fall. We can see that, as expected the area where both the mean difference and K-L divergence is small is marked in blue. But also because of the inconsistencies and inadequate number of images there are several other regions marked in blue which should not be. From the above results we can again see the inadequacy of marginal distribution statistics. As noted in Chapter 2 , the pattern spectrum alone was successful in distinguishing torsion/non-torsion in many cases where marginal statistics failed.

The main drawback of current SVM models is their high computational complexity for large data sets. As we increase the number of parameters the number of training cases required for learning increases exponentially.

We tested SVMs again with the inclusion of pattern spectrum features along with marginal statistics to see if we can achieve nearly perfect discrimination but SVMs were not very effective. This is probably because we have a very low number of test cases when compared to the number of test cases SVMs require.

CHAPTER 5

CONCLUSION

Infrared Thermography is an excellent means to visualize soft tissue infection in humans. We developed a very fast method using pattern spectrum and MATLAB that is able to directly extract the heat pattern properties from the images. In practice this means, for example, that after exposing the infected area for a minute we can take an infrared image of the infected area and the contralateral area and compare the images. Since infection has its own heat pattern we can distinguish it from normal skin. The development of computer aided diagnosis of the skin temperature disparities is very important for the Emergency Room which can now be supported with quick diagnosis and interpretation of the inflamed region. This will lead to shorter and non-intrusive examinations. A similar study has been done for detecting torsion which is another common condition occurring in patients coming to the Emergency Room. Infrared thermography was successful in detecting torsion in sheep testicles and can be applied to humans.

The feature that best distinguished infected vs non-infected images or torsion vs non-torsion infrared images was the pattern spectrum. This is due to the fact that this second order statistic effectively captures the heat pattern distribution while the first order statistics are just a measure of the heat. For the detection strategy we considered all the features and used Case Based Reasoning (CBR). CBR was chosen as we can use most of the images we have as a reference set and use one pair of images as a test set and calculate the effectiveness of the detection strategy. The results did not show improved performance beyond that obtained using the pattern

spectrum alone as the detection statistic. We tried Support Vector Machines (SVMs) for making a detection strategy but it was not very effective as SVMs require a lot of training data which we lacked.

The Receiving Operator Curve plotted using this detection strategy gives an Area Under Curve value of 0.9546 which is a measure of the ability of the test to correctly classify images with and without infection. This test can be classified as an excellent test.

Another conclusion from our work is that, when feasible a clinician should mark the area of suspected infection (e.g., outlined by metal beads) before taking the IR picture. This substantially improved detection performance in our cases.

5.1 Future Work

While our results using Case Based Reasoning did not show gains in performance when we used marginal features in addition to the pattern spectrum, we could investigate what happens when we attempt to optimize weighing of the different features in the test statistic.

Further study can be done to distinguish between abscess and cellulitis based on the heat pattern characteristic specific to each infection. The plots of the size opening Vs pixel wise difference were plotted for cases with Abscess and Cellulitis in our data set were plotted and compared. There was significant difference between the plots. See Figure 5.1. The K-L distance between the normalized distributions was calculated to be 0.1547 which is comparable to the threshold we obtained in distinguishing the infected vs non-infected images. Then the plots were compared to the plots of the healthy contralateral regions. We can see from the plots (Figures 5.2 and 5.3) that when we compare Abscess and Cellulitis Vs Abscess and contralateral there are differences that can be used to distinguish between the infections.

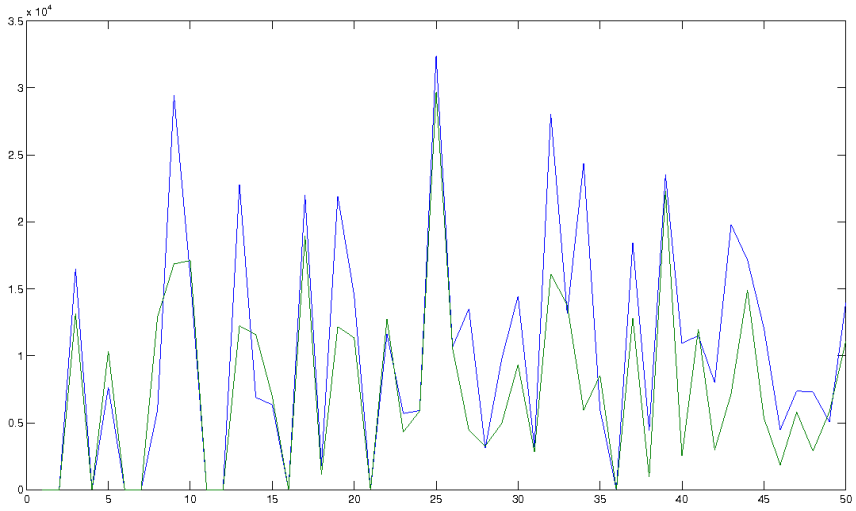


Figure 5.1. Pattern Spectrum statistic for images with Abscess and Cellulitis.

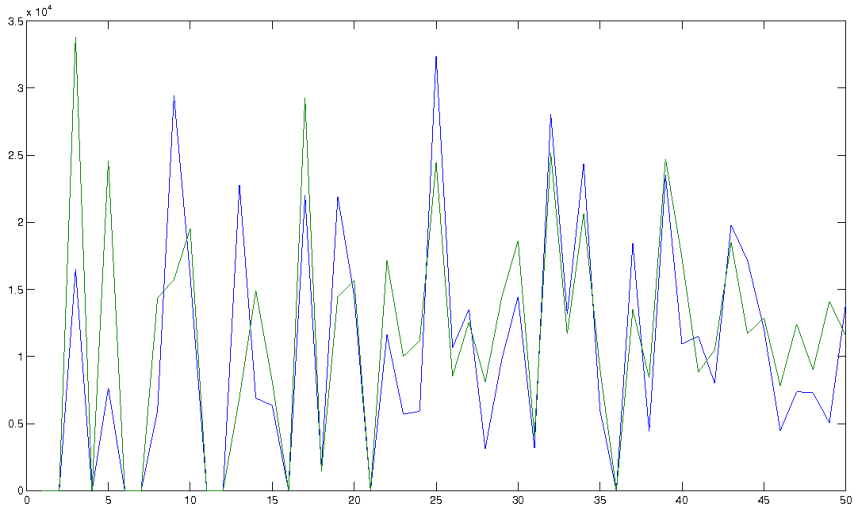


Figure 5.2. Pattern Spectrum for Abscess Vs Contralateral healthy images .

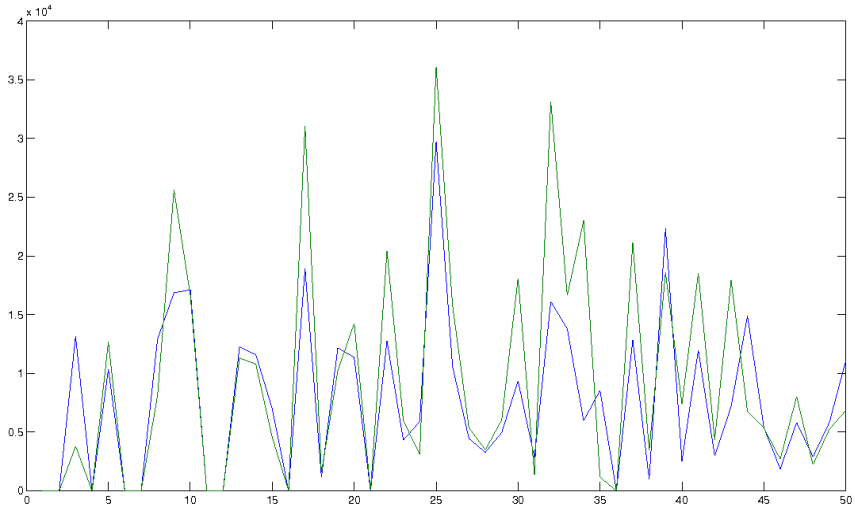


Figure 5.3. Pattern Spectrum for Cellulitis Vs Contralateral healthy images .

The differences is because each of the infections has its own heat pattern. While abscess is characterized by a cavity filled with pus with heat concentrated at the center cellulitis is characterized by layers. This difference in pattern can be used to distinguish them. Since we have limited data (only four cases of cellulitis) on the infection images we did not do a full analysis on this.

BIBLIOGRAPHY

[1] Capraro GA, Blank FS, Nathanson BH, Reiser M, Jasienowski S, Arumalla R, Kelly P, *Can the heat of localized soft tissue infections be quantified non-invasively using an infrared thermography camera?*, presented at American College of Emergency Physicians in Chicago. October 2008.

[2] Capraro GA, Blank FS, Nathanson BH, Reiser M, Jasienowski S, Arumalla R, Kelly P, *Tissue oxygenation in soft tissue infections*, presented at American College of Emergency Physicians in Chicago. October 2008.

[3] Pavlidis and J. Levine, *Thermal image analysis for polygraph testing* . IEEE Engineering in Medicine and Biology Magazine, vol. 21, no. 6, pp. 56-64, November/December 2002.

[4] A. Zalewska , B. Wiecek, A. Sysa-Jedrzejowska, G. Gralewicz, G. Owczarek. *Qualitative thermographic analysis of psoriatic skin lesions.*Proceedings of the 26th Annual International Conference of the IEEE EMBS , San Francisco, CA, USA September 1-5, 2004.

[5] B.F. Jones , G. Schaefer and S.Y. Zhu, *Content-based image retrieval for medical infrared images*, Proceedings of the 26th Annual International Conference of the IEEE EMBS , San Francisco, CA, USA September 1-5, 2004.

[6] Abraham N, Doudle M, Carson P., *Open versus closed surgical treatment of abscesses: a controlled clinical trial*, The Australian and New Zealand Journal of Surgery, 67 (4), pp.173-176, 1997.

[7] Rubner, Y., Tomasi, C., and Guibas, L. J., *The Earth Mover's distance as a metric for image retrieval*, International Journal of Computer Vision, 40(2), pp.99-121, 2000.

[8] B.F. Jones, *A Reappraisal of the use of infrared thermal image analysis in medicine*, IEEE Trans. Med. Imag., vol.17 , no. 6, pp. 1019-1027, December 1998.

[9] Akira Asano, *Texture analysis using morphological pattern spectrum and optimization of structuring elements*, Proceedings of the 10th International Conference on Image Analysis and Processing, IEEE Computer Society, p. 209, 1999.

[10] E. R. Dougherty, J. T. Newell, and J. B. Pelz, , *Morphological texture-based maximum-likelihood pixel classification based on local gramulometric moments*,Pattern Recognition, vol. 25, no. 10, pp. 1181-1198, 1992.

[11] Luc Vincent, *Fast opening functions and morphological granulometries*, Proc. SPIE Vol. 2300 Image Algebra and Morphological Image Processing V, pp. 253-267, San Diego CA, July 1994

[12] Geladi, P, and Kowalski, B., *Partial leastsquares regression: A tutorial*, Analytica Chimica Acta, 185, 1-17,1986.

[13] Alexandra Debiolles, Latifa Oukhellou, Patrice Aknin *Combined use of Partial Least Squares Regression and Neural Network for Diagnosis Tasks*, Proceedings of the 17th International Conference on Pattern Recognition,pp.1051-1076, 2004.

[14] Pabitra Mitra, C A Murthy, and Sankar K Pal., *Unsupervised feature selection using feature similarity*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(3):301- 312, 2002.

[15] Lu Wen-Cong, Chen Nian-Yi, Li Guo-Zheng, Yang Jie, *Multitask learning using Partial Least Method*, The 7th International Conference on Information Fusion, pp. 79-84,Stockholm, Sweden, 2004.

[16] Stephen Winters-Hilt, Anil Yelundur, Charlie McChesney and Matthew Landry, *Support Vector Machine implementations for classification and clustering* , BMC Bioinformatics, 7(Suppl 2):S4, 26 September 2006.

[17] Jin, R., Hauptmann, A.G. and Yan, R., *Image classification using a bigram model*,AAAI Spring Symposium Series Intelligent Multimedia Knowledge Management, pp. 117-129, Palo Alto, CA, March 24-26, 2003.

[18] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin, *A practical guide to Support Vector Classification*, National Taiwan Institute, July 2007.

[19] Xuchun Li, Lei Wang, and Eric Sung, *Multi-label SVM active learning for image classification*,IEEE International Conference on Image Processing (ICIP),pp.113-141, Singapore, 2004.

[20] Effrosyni Kokiopoulou and Pascal Frossard, *Distributed SVM applied to image classification*,Proceedings of IEEE ICME, pp.126-145, Toronto, Canada, 2006.

[21] Steve R. Gunn, *Support Vector Machines for classification and regression*

[22] Hairong Qi and Jonathan F. Head, *Asymmetry analysis using automatic segmentation and classification for breast cancer detection in thermograms* ,IEEE EMBS conference, pp.556-572 , October 2001.

[23] Hairong Qi, *Breast cancer identification through shape analysis in thermal texture maps*,Proc. of the 2nd Joint EMBS-BMES Conference, vol. 2, pages 1155-1156, Houston, 2002.

[24] Iwao Fujimasa, Hideo Nakazawa, *Structural and functional tissue analysis under skin using near infrared spectral imaging*,Proceedings of The First Joint BMES/EMBS Conference Serving Humanity, Advancing Technology, pp.1253-1275, Atlanta , GA, USA, Oct 13-16, 1999.

[25] Sergey A. Telenkov, B. Samuel Tanenbaum, Dennis M. Goodman, J. Stuart Nelson, and Thomas E. Milner, *In Vivo infrared tomographic imaging of laser-heated blood vessels*,IEEE Journal of Selected Topics In Quantum Electronics, Vol. 5, No. 4, July/August 1999.

[26] United States Animal Health Association, *committee on public health* , 2006.

[27] Harriet J.Palitel, Leslie A. Kalish, Ricardo A. Susaeta, Ferdinand Frauscher, Patrick L.O'Kane, Luiz Freitas Filho, *Pulse Inversion Imaging of testicular ischema: Quantitative and qualitative analysis in a rabbit model*, Radiaology: Volume 239, Number3 ,June 2006.

[28] Geoffrey A. Capraro, Bret F. Coughlin, Timothy J. Mader, Howard A Smithline, *Feasibility of using infrared thermography to diagnose testicular torsion: an experimental study in sheep*, presented at Society for Academic Emergency Medicine National Meeting, 2006

[29] Anna O. Bilska-Wolaka, Carey E. Floyd, Jr , Loren W. Nolte, Joseph Y. Lo, *Application of likelihood ratio to classification of mammographic masses; performance comparison to case-based reasoning*,American Association of Physicists in Medicine, Medical Physics, Vol. 30, No. 5, May 2003.